Fakultät für Maschinenwesen

Lehrstuhl für Flugsystemdynamik

Advances in Stability Analysis for Model Reference Adaptive Control Systems and Application to Unmanned Aerial Systems

Dipl.-Ing. (Univ.) Leonhard Höcht

Vollständiger Abdruck der von der Fakultät für Maschinenwesen
der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr.-Ing. Horst Baier
Prüfer der Dissertation:	1. Univ.-Prof. Dr.-Ing. Florian Holzapfel
	2. Univ.-Prof. Dr.-Ing. habil. Boris Lohmann

Die Dissertation wurde am 05.09.2013 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 27.03.2014 angenommen.

This page is accidentally left blank

Content

Figures

Tables

Lemmas

Theorems

Corollaries

Definitions

Symbols

Latin capital letters		
symbol	meaning	unit
$\mathbf{B}(\zeta,\eta)$	decoupling matrix of Byrnes-Isidori normal form	-
C_i	aerodynamic force and moment coefficients $i= L, Q, D, l, m, n$	1
$\mathbf{C}$	gain matrix of error feedback within NDI	-
$\mathbf{C}$	filtered uncertainty stack for MMQ modification	-
$\hat{\mathbf{C}}$	filtered estimated uncertainty stack for MMQ modification	-
$\tilde{\mathbf{C}}$	filtered uncertainty estimation error stack for MMQ modification	-
D	aerodynamic drag	N
D	upper bound on unmatched uncertainty	-
$\left(\vec{\mathbf{F}}_A^B\right)_C = \begin{pmatrix} \left(X_A^B\right)_C \\ \left(Y_A^B\right)_C \\ \left(Z_A^B\right)_C \end{pmatrix}$	force of type A, at point B, components written in frame C	N
$\tilde{\mathbf{F}}(\zeta,\eta,\mathbf{u})$	modeling error	-
$\mathbf{G}(\mathbf{x})$	control effectiveness of feedback linearizable system	-
$G(s)$	transfer function of an LTI filter	-
$\left(\vec{\mathbf{H}}^A\right)_B$	angular momentum relative to point A, components written in frame B	$Nm\cdot s$
$\mathbf{I}^P = \begin{pmatrix} I_x^p & I_{xy}^p & I_{xz}^p \\ I_{xy}^p & I_y^p & I_{yz}^p \\ I_{xz}^p & I_{yz}^p & I_z^p \end{pmatrix}$	inertia tensor relative to point P.	$kg\cdot m^2$
J_{Pr}	moment of inertia of propeller in shaft direction	$kg\cdot m^2$
J_S	moment of inertia of motor shaft	$kg\cdot m^2$
$\mathbf{K}$	feedback matrix, determining reference dynamics	-
$\mathbf{K}_x$	parameters for state dependent uncertainty	-
$\mathbf{K}_N$	parameters for nonaffine control effectiveness	-
L	lift force	N
$M(\phi)$	meridian curvature radius	m

Latin capital letters		
symbol	meaning	unit
$\mathbf{M}_{BA}$	transformation matrix from frame A to frame B	-
$\left(\vec{\mathbf{M}}_A^B\right)_C = \begin{pmatrix} \left(L_A^B\right)_C \\ \left(M_A^B\right)_C \\ \left(N_A^B\right)_C \end{pmatrix}$	moment of type A, at point B, components written in frame C	Nm
$N(\phi)$	normal curvature radius	m
Q	aerodynamic side force	N
$\mathbf{P}(\zeta,\eta)$	control effectiveness of internal dynamics	-
$\mathbf{P}_E$, $\mathbf{P}_R$	solution to Lyapunov equation	-
$\mathbf{Q}_E$, $\mathbf{Q}_R$	input to Lyapunov equation	-
$\mathbf{Q}_x$, $\mathbf{Q}_N$, $\mathbf{Q}_L$	filtered regressors stacks for MMQ modification	-
$\mathbf{Q}_\delta$	filtered unmatched uncertainty stack	-
S	wing reference area	m^2
T	air temperature	K
T_s	air temperature at MSL	K
$T_i(\delta_i,b_i)$	time to ultimate bound	s
U_0, $U_{0,l}$, $U_{0,r}$, $U_{0,b}$,	input voltage to aircraft engines	V
V	Lyapunov function	-
$\dot{V}$	time derivative of Lyapunov function	-
$\left(\vec{\mathbf{V}}^{AB}\right)_C^D = \begin{pmatrix} \left(u^{AB}\right)_C^D \\ \left(v^{AB}\right)_C^D \\ \left(w^{AB}\right)_C^D \end{pmatrix}$	velocity w.r.t to frame D, of point B relative to point A, components written in frame C	m/s
$\left(\dot{\vec{\mathbf{V}}}^{AB}\right)_C^{DE} = = \begin{pmatrix} \left(\dot{u}^{AB}\right)_C^{DE} \\ \left(\dot{v}^{AB}\right)_C^{DE} \\ \left(\dot{w}^{AB}\right)_C^{DE} \end{pmatrix}$	time derivative w.r.t frame E of velocity w.r.t to frame D, of point B relative to point A, components written in frame C,	m/s^2
$\mathbf{W}$	compound parameter matrix	-

Latin letters		
symbol	meaning	unit
a	semi-major axis of reference ellipsoid	m
$\mathbf{a}(\boldsymbol{\zeta},\boldsymbol{\eta})$	state dependent nonlinearity of external dynamics	-
b	semi-minor axis of reference ellipsoid	m
b_i	ultimate bounds	-
b	wing span	m
$\bar{c}$	mean aerodynamic chord	m
$\mathbf{c}$	filtered uncertainty for MMQ modification	-
$\hat{\mathbf{c}}$	filtered estimated uncertainty for MMQ modification	-
$\tilde{\mathbf{c}}$	filtered uncertainty estimation error for MMQ modification	-
$\mathbf{d}$	external disturbances	-
e	excentricity of reference ellipsoid	1
$\underline{e}$	lower bound on maximum tracking error	-
$\mathbf{e}$	tacking error, prediction-tracking error	-
$\hat{\mathbf{e}}$	prediction error	-
f	excentricity	1
$\mathbf{f}(\mathbf{x})$	state dependent nonlinear function of feedback linearizable system	-
$\mathbf{g}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$	nonaffine control map	-
h	geodetic altitude	m
$\mathbf{h}(\mathbf{x})$	nonlinear output function of feedback linearizable system	-
$\vec{\mathbf{p}}$	linear momentum	$N{\cdot}s$
$\mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta})$	state dependent nonlinearity of internal dynamics	-
$\mathbf{q}_x$, $\mathbf{q}_N$, $\mathbf{q}_L$	filtered regressors for MMQ modification	-
$\mathbf{q}_\delta$	filtered unmatched uncertainty	-
p	air pressure	N/m^2
p_s	air pressure as MSL	N/m^2
p	roll rate	rad/s
p^*	normalized roll rate	1
q	pitch rate	rad/s
q^*	normalized pitch rate	1

Latin letters		
symbol	meaning	unit
$\overline{q}$	dynamic pressure	N/m^2
$\underline{\mathbf{q}} = \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}$	unit quaternion	1
r	yaw rate	rad/s
r^*	normalized yaw rate	1
$\left(\vec{\mathbf{r}}^{AB}\right)_C = \begin{pmatrix} \left(x^{AB}\right)_C \\ \left(y^{AB}\right)_C \\ \left(z^{AB}\right)_C \end{pmatrix}$	position vector from point A to point B, components written in frame C	m
$\mathbf{u}$	input vector of a dynamic system	-
$\mathbf{u}_N$	nonaffine controls	-
$\mathbf{w}$ ($\mathbf{w}^*$)	(desired) effect of nonaffine controls	-
$\mathbf{x}$	state vector of a dynamic system	-
$\mathbf{y}$	output vector of a dynamic system	-

Greek letters		
symbol	meaning	unit
α, α_A, α_K, α_T, α_c	(aerodynamic, kinematic, total, included) angle of attack	*rad*
$\alpha_i(\|\cdot\|)$	class $\mathcal{K}$ function establishing lower bound on Lyapunov function	-
β, β_A, β_K, β_T	(aerodynamic, kinematic, total) aerodynamic angle of side-slip	*rad*
$\beta_i(\|\cdot\|)$	class $\mathcal{K}$ function establishing upper bound on Lyapunov function	-
$\mathbf{\Gamma}_x$, $\mathbf{\Gamma}_N$, $\mathbf{\Gamma}_L$	symmetric positive definite weighting matrices	-
γ_x, γ_N, γ_L	learning rates	-
$\underline{\gamma}_x$, $\underline{\gamma}_N$, $\underline{\gamma}_L$	lower bounds on learning rates	-
$\gamma_i(\|\cdot\|)$	class $\mathcal{K}$ function establishing upper bound on Lyapunov function derivative	-
δ_l, δ_r	(left, right) flap deflection	*rad*
δ_i	initial condition bound	-
ζ	rudder deflection	*rad*
ζ	relative damping of second order LTI system	1
$\boldsymbol{\zeta}$	external states of Byrnes-Isidori normal form	-
$\boldsymbol{\zeta}_R$	states of reference model	-
$\hat{\boldsymbol{\zeta}}$	states of state predictor	-
$\boldsymbol{\eta}$	internal states of Byrnes-Isidori normal form	-
η_l, η_r	(left, right) elevator deflection	*rad*
$\eta_{c,l}$, $\eta_{c,r}$	(left, right) canard deflection	*rad*
Θ	Euler pitch angle	*rad*
$\theta_{i,j}$	set of Euler angles, describing attitude of propulsion systems relative to body-fixed frame ($i=1,2,3$, $j=l,r,b$)	*rad*
θ_x, θ_N, θ_L	upper bounds on true parameters	-
$\theta_{x,max}$, $\theta_{N,max}$, $\theta_{L,max}$	bounds, imposed by projection operator	-
κ_b	back thrust vector azimuth angle	*rad*
κ	gain for MMQ modification	-
λ	eigenvalue of a matrix; $\overline{\lambda}$: greatest eigenvalue; $\underline{\lambda}$: smallest eigenvalue	-
$\mathbf{\Lambda}$	control effectiveness	-
λ	geodetic longitude	*rad*
μ, μ_A, μ_K	(aerodynamic, kinematic) bank angle	*rad*
$\boldsymbol{\nu}$	pseudo control	-
$\boldsymbol{\nu}_A$	adaptive augmentation	-
$\boldsymbol{\nu}_E$	error feedback	-
$\boldsymbol{\nu}_R$	reference pseudo control	-
$\hat{\boldsymbol{\nu}}_R$	predictor based pseudo control	-
ξ_l, ξ_r	(left, right) aileron deflection	rad
σ	singular value of a matrix $\overline{\sigma}$: greatest singular value; $\underline{\sigma}$: smallest eigenvalue	-

Greek letters		
symbol	meaning	unit
ρ	air density	kg/m^3
ρ_s	air density at MSL	kg/m^3
ρ_1	initial condition bound for ultimate boundedness	-
$\sigma_l,\ \sigma_r,\ \sigma_b,$	(left, right, back) thrust vector inclination angle	rad
$\sigma_x,\ \sigma_N,\ \sigma_L,$	gains of switching σ-modification	-
Φ	Euler bank angle	rad
$\Phi(\mathbf{x}),\ \Phi_\zeta(\mathbf{x}),\ \Phi_\eta(\mathbf{x})$	nonlinear state transformations for Byrnes-Isidori normal form	-
ϕ	geodetic latitude	rad
$\boldsymbol{\varphi}_x(\boldsymbol{\zeta},\boldsymbol{\eta})$	state dependent regressor	-
$\boldsymbol{\omega}_N(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$ $\boldsymbol{\varphi}_N(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$	nonaffine control regressor	-
Ψ	Euler azimuth	rad
$\omega_{M,l},\ \omega_{M,r},\ \omega_{M,b}$	(left, right, back) propeller rotation speed	rad/s
ω_0	natural frequency of first or second order LTI system	rad/s
$\left(\vec{\boldsymbol{\omega}}^{AB}\right)_C = \begin{pmatrix} \left(\omega_x^{AB}\right)_C \\ \left(\omega_y^{AB}\right)_C \\ \left(\omega_z^{AB}\right)_C \end{pmatrix}$	angular rate of frame B relative to frame A, components written in frame C	rad/s
$\left(\dot{\vec{\boldsymbol{\omega}}}^{AB}\right)_C^D = \begin{pmatrix} \left(\dot{\omega}_x^{AB}\right)_C^D \\ \left(\dot{\omega}_y^{AB}\right)_C^D \\ \left(\dot{\omega}_z^{AB}\right)_C^D \end{pmatrix}$	time derivative w.r.t frame D of angular rate of frame B relative to frame A, components written in frame C	rad/s^2

abbreviations	
symbol	meaning
AHRS	attitude heading reference system
AoA	angle of attack
AoS	angle of side-slip
a.s.	asymptotically stable
BLDC	brushless DC electric motor
c.g.	center of gravity
DC	direct current
EMC	electromagnetic compatibility
EOM	equation of motion
l.h.c.p.	left half of complex plane
LTI	linear time invariant
LTV	linear time variant
ISA	international standard atmosphere
MEMS	micromechanical systems
MIMO	multi input multi output
MMQ	multi model Q (modification introduced in section 4.4)
MRAC	model reference adaptive control
MSL	mean see level
NDI	nonlinear dynamic inversion
ODE	ordinary differential equation
PSD	power spectral density
RBF	radial basis function neutral network
RD	relative degree
r.h.c.p.	right half of complex plane
SISO	single input single output
SP	singular perturbation
SV	singular value
UAS	unmanned aerial system
UAV	unmanned aerial vehicle

Chapter 1

Introduction

In the 20th century, the presence of a pilot on board of an aircraft has been inevitable, however, since beginning of the millennium a rapid development of unmanned aerial systems (UAS) initiated, not at least due to the availability of computers, whose computational speed increased tremendously. Taking a look into history, military interests, unfortunately, were often the main impetus for technological progress as is also the case for UAS. The absence of a pilot opened new capabilities allowing full exploitation of physical aircraft limits, but also generated new requirements. Particularly autonomous accomplishment of complete missions became an important topic, which requires for a system that is failure tolerant and the latter requirement is largely affected by the flight control system.

Simultaneously, adaptive flight control gained more and more interest. The first adaptive flight control systems were in fact investigated since the beginning second half of the 20th century. In these days, the adaptation laws were rather designed on an empirical basis, than on exact mathematical considerations, which lead to a fatal crash of the X15A in 1967, since the control system became instable. As a result, awareness has grown that a rigorous consideration of the closed loop stability, including the adaptation, is a critical point. Since the 1980s, various adaptive control concepts have been published, but now with explicit consideration of stability ([Nar80]), mainly based on Lyapunov's methods.

Several basic adaptive control structures have emerged over the years such as adaptive backstepping, $\mathcal{L}_1$ adaptive control ([Hov10]) or model reference adaptive control (MRAC) and there are various references available, which are considered as standard work in the field of nonlinear system analysis and adaptive control such as [Ioa06], [Isi95], [Kha02], [Tao03], [Nar05], [Slo90] and [Hov10].

Adaptive control is a promising concept to account for failure tolerance as well as safe flight operation in case of uncertain environmental conditions. Nevertheless, it is a particular challenge to design an adaptive controller for an airborne system, since it needs to be certified by the official authorities. They require for guarantees, that the control system operates the aircraft safely in any flight condition, the aircraft is designed for. Robustness of classical linear controllers is analyzed by phase and gain margin, which are accepted metrics for the certification authorities. Although, it is possible to consider robustness of the inherently nonlinear adaptive control systems,

there is currently no method that is accepted, which renders this issue a main challenge for current research on adaptive fight control.

Today Lyapunov's methods are an effective tool for stability and robustness analysis of adaptive systems. However, often the results are quite conservative, meaning that guaranteed bounds are too large, lying beyond a physically relevant range, or even no bounds are computed at all, but it is merely stated that bounds exist. This is, of course, not sufficient for certification.

1.1 Contribution

Aim of the thesis is the design and implementation of a nonlinear and adaptive control system for the FSD Extreme Star, a small scale unmanned aircraft that serves as testbed for modern flight control systems. As failure tolerance is an important issue for UAS, a focus is put on concepts that account for actuator failures. Besides application, also a rigorous derivation of the algorithms and analysis of the closed loop system stability is accomplished.

Certainly, the theory on nonlinear adaptive control and the associated Lyapunov based stability considerations are quite involved and it is commonly difficult to understand derivations, if the reader is not a particular specialist on this topic. Nevertheless, the thesis at hand is intended to be such comprehensive, by providing all the necessary background that a common graduate level engineer has a chance to follow the presented derivations and arguments.

Particularly Appendix B provides mathematical background, which is important in the field of nonlinear and adaptive control, but which is commonly not known in detail. As Lyapunov's methods form the central tool for analysis of stability and boundedness of the nonlinear dynamic systems, a comprehensive introduction is given in Appendix C.

The following approaches are original contributions of this thesis.

Novel Formulation of Ultimate Boundedness Theorem

A recurrent theme of the thesis is the explicit computation of guaranteed values within the stability proofs of the control systems. In fact, there are many publications on adaptive flight control, which present innovative and sophisticated concepts, however, many of them suffer from the fact that no explicit values on stability conditions are computed. For real aircraft applications, asymptotic stability of the system cannot be proved due to unmodeled dynamics or unmatched uncertainties. At least boundedness of the system states can be proved under certain conditions. These proofs are often based on the respective theorem for ultimate boundedness in [Kha02]. A central contribution of this work is a novel formulation of the ultimate boundedness theorem, where explicit values are computed for bounds, ultimate bounds and the time, after

which the ultimate bound is reached. Moreover, the theorem is tailored to MRAC systems where the whole system state is usually partitioned into a tracking error and a parameter-estimation-error. Accounting for the special structure of MRAC systems within the theorem potentially provides less conservative bounds.

Novel Multi Model Q Modification

A known drawback of MRAC systems is the limited adaptation performance. Although high gains increase adaptation speed to a certain extent, they should not be chosen too high, since this leads to high frequency oscillations. In order to improve adaptation performance, various modifications to the update law have been proposed ([Cho102], [Cho10], [Cho101], [Cho09], [Joh04], [Ngu08], [Vol09], [Vol06] [Vol061], [Yuc09], [Yuc091]), amongst others q modification and concurrent learning. Both approaches aim to overcome the rank-1 condition of the update law, which has been identified as a reason for poor adaptation performance. Both approaches utilize the plant dynamics to gain an expression for the true uncertainty, by separating the known from the unknown part. The particular challenge thereby is that also state derivatives have to be known, which usually cannot be measured without a considerable effort.

On the one hand, concurrent learning provides a solution to that challenge by estimation of the state derivative. This is possible, since it constitutes a matrix, whose columns contain recorded uncertainties from past time instants (also referred to as “history stack”) and the state derivative is obtained by e.g. a fixed-point smoother. The additional term within the update law is derived from the gradient of a quadratic cost function involving the uncertainty estimation error. A special advantage of concurrent learning is that even exponential stability of tracking error and parameter-estimation-error can be shown, if the history stack has a sufficient number of linearly independent columns. Once the data are recorded, the columns of the history stack are fixed, if no additional algorithm exchanges the recorded data and hence concurrent learning cannot react to sudden configuration changes such as failures.

On the other hand, q modification uses a stable LTI filter, in order to gain a filtered state derivative from measured states and the additional term within the update law is derived from the gradient of a quadratic cost function of the filtered uncertainty estimation error. The use of a filter, instead of a stack of fixed recorded data provides the advantage that changes in configuration are accounted for, since old values are continuously washed out by the filter. Unfortunately, q modification does not provide the chance for exponential stability, since only a single filter is used, which adds a second direction to the update law, but never reaches the sufficient number of linearly independent columns.

Therefore, a novel multi model Q (MMQ) modification is proposed and derived in the thesis. It intends to combine the advantages of both existing approaches and is a

completely new method. Instead of sending the known part of the plant dynamics through a single filter, it is sent through multiple filters, where the results are stacked into columns of a matrix, as is done in concurrent learning. If the number of filters is sufficiently large, there is at least the change to have, at least temporarily, a number of linearly independent columns, sufficiently large for exponential stability. However, contrary to the concurrent learning approach where the matrix is constant, linear independence of a sufficient number of columns cannot be guaranteed uniformly in time for MMQ modification, since the filtered uncertainties change over time and depend on the excitation. Yet MMQ modification provides the chance for exponential convergence of the system states AND reacts to a sudden change of the aircraft configuration.

Contributions to Incorporation of Nonaffine Controls into the Control Loop

The non-adaptive part of the proposed controller is based on nonlinear dynamic inversion (NDI), a concept that transforms a nonlinear dynamic system into a linear decoupled one ([Isi95], [Kha02], [Slo90]) without using any approximation. In its basic form, NDI is developed for systems that are affine in their controls. However, the considered aircraft also comprises thrust vectoring, which enter the dynamics nonlinearly, due to trigonometric functions of the thrust vector angles. In recent years, a novel method, referred to as “nonlinear-in-control design” (NIC) has been published that allows the incorporation on nonaffine controls into the loop without explicit knowledge of the inverse of the nonlinear control map ([Lav08], [Lav07b], [Lav07a]). The algorithm does not require an explicit knowledge of the inverse but, of course, it has to exist. Within publications on NIC, sufficient conditions for existence are derived, which are based on an integral expression, involving the control map Jacobian. Theorem 4.5 in section 4.5 of the thesis summarizes the results so far, but adds another sufficient condition, which is potentially easier to verify, since it considers the map Jacobian directly without integration. In order to incorporate this concept into the NDI framework, an extension to the nonaffine case – particularly a generalized definition of relative degree for nonaffine-in-control systems – had to be defined first. Basically, NDI for nonaffine SISO systems has already been derived within the field of chemical engineering ([Hen90], [Hen96]), but an extension to the MIMO case has not been published yet, as far as is known to the author.

Novel Update of Control Effectiveness Based on Singular Value Decomposition

In MRAC schemes, adaptive compensation of state dependent uncertainties follows a quite straightforward procedure. More challenging is the compensation of uncertainties in the control channel and several solutions have been published. Narendra proposed a solution within a direct MRAC scheme by estimation of the inverted control

effectiveness in [Nar05], however, the result is only a local one. Other approaches, such as predictor based MRAC ([Bie10]) use the inverted control effectiveness estimate in the control law while the control effectiveness itself is updated. This, however, requires that the estimated control effectiveness does not become singular. The NDI based approach, which is pursued here, also uses the inverse of the estimated control effectiveness within the linearizing state feedback. In order to prevent the control effectiveness from becoming singular, it is usually constrained to diagonal matrices with positive diagonal entries. This, however, considerably restricts the cases, adaptation can compensate for. Therefore, a novel concept is introduced, which reformulates the control effectiveness update law in terms of its singular value decomposition. It allows preventing the matrix from becoming singular by limitation of the singular values from below by a positive constant, while simultaneously the diagonal constraint is released.

1.2 Outline

The thesis is organized as follows. The subsequent section 1.3 introduces nomenclature and basic mathematical results that are used throughout the thesis. Chapter 2 presents a comprehensive derivation of the aircraft equations of motions, which are used for simulation and as basis for control system design.

Chapter 3 provides the novel formulation of the ultimate boundedness theorem including explicit bounds, ultimate bounds and time after which ultimate bounds are reached.

Chapter 4 contains theoretical derivations, needed for control system design. Section 4.1 presents an introduction to NDI, starting with the established and well-known facts, but also contains extensions to the basic framework, such as incorporation of redundant affine controls, nonaffine controls and disturbances. The rest of Chapter 4 is dedicated to adaptive control. After a short introduction in section 4.2, the first variant of two adaptive flight control algorithms, including a stability analysis, is introduced in section 4.3. It does not use the nonaffine controls (respectively thrust vectoring) within the linearizing state feedback, contrary to the second variant, presented in section 4.6. Section 4.4 motivates MMQ modification and derives the necessary equations and section 4.5 provides the necessary results for setup of the NIC algorithm. The singular value decomposition based update of the control effectiveness is derived in section 4.7.

Chapter 5 presents simulation examples that show effectiveness of MMQ modification and singular value update algorithm in sections 5.1 and 5.2, while section 5.3 contains a detailed description of a concrete implementation of the control structure, derived in section 4.3 for the aircraft testbed, including a comprehensive simulation study. Section 5.3.5 should be particularly mentioned, since it contains a procedure for the

design of adaptive gains, including a discussion on certification aspects. Finally, Chapter 6 draws conclusions and provides perspectives on future research topics.

1.3 Important Mathematical Preliminaries

This section contains some notations and basic mathematical results used throughout the thesis. If z denotes a complex number, then $\bar{z}$ denotes the conjugate complex and $|z|$ denotes its magnitude. Generally, small italic letters are used for scalars (x), small bold letters are used for vectors ($\mathbf{x}$), capital bold letters are used for matrices ($\mathbf{A}$) and $\mathbf{I}$ denotes the identity matrix of appropriate dimension.

In the following, basic nomenclature, properties and definitions, related to vector spaces and matrices are stated. There are many textbooks available on this topic. For more information, refer to [Lüt96] and references therein.

The i^{th} element of a vector is denoted as x_i. A single element of a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ in row i and column j is denoted as a_{ij}, the transpose of a matrix is denoted as $\mathbf{A}^T$ and the Hermitian (conjugate complex and transpose) is denoted as $\mathbf{A}^H$. $\lambda_{i,A}$,$\sigma_{i,A}$ denote the i^{th} eigenvalue and singular value of $\mathbf{A}$ respectively and $\underline{\lambda}_A, \overline{\lambda}_A, \underline{\sigma}_A, \overline{\sigma}_A$ denote the minimum/maximum eigenvalue/singular value.

Vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$ are elements of a linear space, which is primarily a mathematical construct. For engineers it is an advantageous tool to gain conclusions that go beyond the ones obtained from an intuitive engineering view. Especially in nonlinear system theory and control as well as Lyapunov stability analysis, detailed knowledge about terms such as “Banach space”, “compact”, “closed and open” ” or the difference between a pointwise continuous and uniformly continuous functions, are often used in the proofs, which are defined in the context of linear spaces and its advanced topics. Examples therefore are the proof of existence and uniqueness of a solution to a system of nonlinear ordinary differential equations that fulfill the Lipschitz condition ([Kha02]), La Salle’s invariance principle ([Kha02]), or the Lyapunov-Krasovskii stability theorem ([GuK03], [Hal93]). Particularly the latter is concerned with abstract Banach spaces, whose elements are time domain signals on a compact interval. It provides sufficient conditions on stability of functional differential equations such as time-delay systems. By mathematical abstraction, terms that are intuitively clear in Euclidean space, such as “distance” or “convergence”, can be transferred to any abstract linear space. Those concepts are hence a useful tool to draw conclusions, which are indeed clear in Euclidean space but could never be gained without mathematical abstraction. Due to its importance for nonlinear and adaptive control and since those rather abstract mathematical topics are sometimes treated stepmotherly in engineering science, an introduction to linear spaces and some other related facts are given in Appendix B.1 and are henceforth assumed to be known.

For a subspace $S \subset \mathbb{R}^n$, the dimension, denoted as $\dim(S)$, is defined as the maximum number of linearly independent vectors in S. The kernel (or null space) of $\mathbf{A} \in \mathbb{R}^{n \times m}$ is defined as the set of all vectors $\mathbf{x}$, which satisfy $\mathbf{Ax}=\mathbf{0}$.

$$\ker(\mathbf{A}) = \left\{ \mathbf{x} \in \mathbb{R}^m \middle| \mathbf{A} \cdot \mathbf{x} = \mathbf{0} \right\}$$

As the kernel of a matrix is a linear subspace, it also can be attributed with a dimension, denoted as $\dim(\ker(\mathbf{A}))$.

The column rank of $\mathbf{A}$, $\text{rank}_C(\mathbf{A})$ is defined as the number linearly independent columns, the row rank of a matrix, $\text{rank}_R(\mathbf{A})$ is defined as the number of linearly independent rows. As shown in Theorem B.14, $\text{rank}_C(\mathbf{A}) = \text{rank}_R(\mathbf{A})$ and therefore we write

$$\text{rank}(\mathbf{A}) = \text{rank}_C(\mathbf{A}) = \text{rank}_R(\mathbf{A}).$$

The image of $\mathbf{A}$ is defined as the set of all $\mathbf{z} \in \mathbb{R}^n$ that can be generated by $\mathbf{A} \cdot \mathbf{x}$ for any $\mathbf{x} \in \mathbb{R}^m$:

$$\text{image}(\mathbf{A}) = \left\{ \mathbf{z} \in \mathbb{R}^n \middle| \mathbf{z} = \mathbf{A} \cdot \mathbf{x}, \mathbf{x} \in \mathbb{R}^m \right\}$$

$\text{image}(\mathbf{A})$ is a subspace of $\mathbb{R}^n$ that is spanned by the columns of $\mathbf{A}$. Hence it can also be attributed with a dimension and note that

$$\dim(\text{image}(\mathbf{A})) = \text{rank}(\mathbf{A}).$$

The determinant of a real valued quadratic matrix $\mathbf{P} \in \mathbb{R}^{m \times m}$ is denoted as $\det(\mathbf{P})$. For $\mathbf{Q} \in \mathbb{C}^{m \times m}$, the trace (sum of its diagonal elements) is

$$\text{tr}(\mathbf{Q}) = \sum_{i=1}^{n} q_{ii} .$$

Some important properties, pertaining to trace operator that are extensively used throughout the thesis are stated in the following. Let matrices $\mathbf{A} \in \mathbb{C}^{n \times m}, \mathbf{B} \in \mathbb{C}^{n \times p}, \mathbf{C} \in \mathbb{C}^{p \times n}, \mathbf{D} \in \mathbb{C}^{n \times m}$, vectors $\mathbf{x}, \mathbf{y} \in \mathbb{C}^m$ and scalars $\alpha, \beta \in \mathbb{C}$, then the following properties hold:

- linearity $\text{tr}[\alpha\mathbf{A} + \beta\mathbf{D}] = \alpha\,\text{tr}[\mathbf{A}] + \beta\,\text{tr}[\mathbf{D}]$
- invariance w.r.t. transposition: $\text{tr}(\mathbf{Q}) = \text{tr}(\mathbf{Q}^T)$
- cyclic property: $\text{tr}(\mathbf{ABC}) = \text{tr}(\mathbf{BCA}) = \text{tr}(\mathbf{CAB})$
- dyadic product: $\text{tr}(\mathbf{xy}^T) = \mathbf{x}^T\mathbf{y}$

The p-norm of a vector is defined as

$$\|\mathbf{x}\|_p = \left(|x_1|^p + \ldots |x_m|^p \right)^{\frac{1}{p}}$$

for $p = [1, \infty)$. Particularly $p=2$ denotes the classical Euclidean vector norm. Further, the ∞-norm is defined as:

$$\|\mathbf{x}\|_\infty = \max_{i=1,\ldots,m} \left(|x_i|\right)$$

For matrices, the induced p-norms provide a relation between p-norms of vectors that are related to each other by that matrix. Let $\mathbf{z}=\mathbf{A}\cdot\mathbf{x}$, then the matrix induced p-norm is defined as:

$$\|\mathbf{A}\|_p = \sup_{\|\mathbf{x}\|_p=1} \|\mathbf{z}\|_p = \sup_{\|\mathbf{x}\|_p=1} \|\mathbf{A}\mathbf{x}\|_p$$

The matrix induced norms are hence a least upper bound for the elongation of $\mathbf{z}$, relative to $\mathbf{x}$. For the prominent cases $p=1,2,\infty$ explicit values can be computed.

$$\|\mathbf{A}\|_1 = \max_{j=1,\ldots,m} \sum_{i=1}^{n} |a_{ij}| \ , \ \|\mathbf{A}\|_2 = \overline{\sigma}_A \ , \ \|\mathbf{A}\|_\infty = \max_{i=1,\ldots,n} \sum_{j=1}^{m} |a_{ij}| \tag{1.1}$$

An important property of matrix induced norms is given by the following inequality.

$$\|\mathbf{AB}\|_p \le \|\mathbf{A}\|_p \|\mathbf{B}\|_p \tag{1.2}$$

In order to show this, let some $\mathbf{x}\in\mathbb{C}^p$ (Caution: Do not mix it up with p of the norm!), $\mathbf{z}_1 = \mathbf{B}\mathbf{x}$, $\mathbf{z}_2 = \mathbf{A}\mathbf{z}_1$. On the one hand

$$\|\mathbf{z}_2\|_p \le \|\mathbf{A}\|_p \|\mathbf{z}_1\|_p \le \|\mathbf{A}\|_p \|\mathbf{B}\|_p \|\mathbf{x}\|_p$$

and, on the other hand

$$\|\mathbf{z}_2\|_p \le \|\mathbf{AB}\|_p \|\mathbf{x}\|_p .$$

Since $\|\mathbf{AB}\|_p$ is the least upper bound, that fulfills the latter inequality, the former inequality implies (1.2). Besides induced norms, a direct norm can be assigned to a matrix, of which we will only use the Frobenius norm.

$$\|\mathbf{A}\|_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} |a_{ij}|^2 \right)^{\frac{1}{2}}$$

It is easily verified, that the Frobenius norm can also be written in terms of trace operator.

$$\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}\left(\mathbf{A}^H \mathbf{A}\right)} = \sqrt{\mathrm{tr}\left(\mathbf{A}\mathbf{A}^H\right)}$$

Let $\sigma_{1,A},\ldots, \sigma_{r,A}$, be the singular values for $\mathbf{A}$, where $r=\min(n,m)$. The Frobenius norm evaluates to (refer e.g. to [Lüt96])

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{r} \sigma_{i,A}^2} \tag{1.3}$$

This result, together with equation (1.1), particularly implies that the matrix induced 2-norm is not bigger than the Frobenius norm.

$$\|\mathbf{A}\|_2 \le \|\mathbf{A}\|_F \tag{1.4}$$

If the columns of $\mathbf{A}$ are all stacked into one column vector, denoted by $\mathrm{vec}(\mathbf{A})$, the Frobenius norm could also be written as

$$\|\mathbf{A}\|_F = \sqrt{\mathrm{vec}(\mathbf{A})^H \, \mathrm{vec}(\mathbf{A})}. \tag{1.5}$$

It is hence a natural extension of Euclidean vector norm to matrices and obviously, if the matrix collapses to a vector $\mathbf{x}\in\mathbb{C}^m$, its Frobenius norm is equal to the Euclidean vector norm.

$$\|\mathbf{x}\|_F = \|\mathbf{x}\|_2 \tag{1.6}$$

There is also a matrix analog for scalar product of vectors.

$$\mathrm{tr}\left(\mathbf{A}^H\mathbf{D}\right) = \mathrm{vec}(\mathbf{A})^H \, \mathrm{vec}(\mathbf{D})$$

In this context, it can be shown that the following inequalities hold

$$\left|\mathrm{tr}\left(\mathbf{A}^H\mathbf{D}\right)\right| \le \|\mathbf{A}\|_F \|\mathbf{D}\|_F \tag{1.7}$$

$$\left\|\mathbf{A}^H\mathbf{D}\right\|_F \le \|\mathbf{A}\|_F \|\mathbf{D}\|_F \tag{1.8}$$

which is a generalization of the Cauchy-Schwartz inequality for vectors.

$$\left|\mathbf{x}^H\mathbf{y}\right| \le \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

Cauchy-Schwartz is in turn a special case of Hölder's inequality. Let some $p,q \ge 1$ such that $1/p + 1/q = 1$, then

$$\left|\mathbf{x}^T\mathbf{y}\right| \le \|\mathbf{x}\|_p \|\mathbf{y}\|_q .$$

Vector p-norms as well as matrix induced p-norms are equivalent, meaning that, for any $\mathbf{x}$, $\mathbf{y}$ of the same vector space S, there are constants $0<c_1\le c_2$ such that

$$c_1\|\mathbf{x}\|_a \le \|\mathbf{x}\|_b \le c_2\|\mathbf{x}\|_a .$$

Therefore the norm index is dropped, if some vector or matrix induced p-norm is meant and it is no matter, which p-norm to take.

Frequently used throughout the thesis are expressions of the form $\mathbf{v}^T\boldsymbol{\Gamma}\mathbf{v}$, also referred to as quadratic forms, where $\mathbf{v}\in\mathbb{R}^n$, $\boldsymbol{\Gamma}\in\mathbb{R}^{n\times n}$ symmetric and positive definite. An important inequality pertaining to quadratic forms is given in Theorem B.20.

$$\underline{\lambda}_\Gamma \|\mathbf{v}\|_2^2 \le \mathbf{v}^T\boldsymbol{\Gamma}\mathbf{v} \le \overline{\lambda}_\Gamma \|\mathbf{v}\|_2^2 \tag{1.9}$$

A generalization of quadratic forms to matrices is given by $\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}\boldsymbol{\Theta}\right)$, where $\boldsymbol{\Theta}\in\mathbb{R}^{n\times m}$. Theorem B.21 shows that the following inequality holds.

$$\underline{\lambda}_\Gamma \|\boldsymbol{\Theta}\|_F^2 \le \mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}\boldsymbol{\Theta}\right) \le \overline{\lambda}_\Gamma \|\boldsymbol{\Theta}\|_F^2 \tag{1.10}$$

Quadratic forms such as (1.9) fulfill the axioms of a norm, assigned to a linear space (refer to section B.1.1). It is therefore appropriate to define a weighted Euclidean norm on $\mathbb{R}^m$

$$\|\mathbf{v}\|_\Gamma = \mathbf{v}^T\boldsymbol{\Gamma}^{-1}\mathbf{v} \tag{1.11}$$

where the weighting matrix $\boldsymbol{\Gamma}\in\mathbb{R}^{n\times n}$ is symmetric and positive definite. Let $\mathbf{w}\in\mathbb{R}^n$, then a weighted analogue of the Cauchy-Schwartz inequality is

$$\left|\mathbf{v}^T\boldsymbol{\Gamma}^{-1}\mathbf{w}\right| \le \|\mathbf{v}\|_\Gamma\|\mathbf{w}\|_\Gamma \tag{1.12}$$

which can be shown, using the Cauchy-Schwartz inequality in its original form. Since $\boldsymbol{\Gamma}^{-1}$ is also symmetric and positive definite, it has a Cholescy factorization ([Gol85]), i.e. there is a lower triangular matrix $\mathbf{S}\in\mathbb{R}^{n\times n}$ with positive diagonal entries, such that $\mathbf{S}\mathbf{S}^T=\boldsymbol{\Gamma}^{-1}$. Let $\mathbf{v}_s=\mathbf{S}^T\mathbf{v}$, $\mathbf{w}_s=\mathbf{S}^T\mathbf{w}$, then

$$\left|\mathbf{v}^T\boldsymbol{\Gamma}^{-1}\mathbf{w}\right| = \left|\mathbf{v}_s^T\mathbf{w}_s\right| \le \|\mathbf{v}_s\|_2\|\mathbf{w}_s\|_2 = \sqrt{\mathbf{v}^T\mathbf{S}\mathbf{S}^T\mathbf{v}}\sqrt{\mathbf{w}^T\mathbf{S}\mathbf{S}^T\mathbf{w}} = \|\mathbf{v}\|_\Gamma\|\mathbf{w}\|_\Gamma .$$

This concept can also be generalized to matrices $\boldsymbol{\Theta}\in\mathbb{R}^{n\times m}$, by defining the weighted Frobenius norm.

$$\|\boldsymbol{\Theta}\|_\Gamma = \sqrt{\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}\right)} \tag{1.13}$$

Notice that the weighting matrix is a square matrix whose size equals the number rows of $\boldsymbol{\Theta}$ and

$$\left\|\boldsymbol{\Theta}^T\right\|_\Gamma = \sqrt{\mathrm{tr}\left(\boldsymbol{\Theta}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}^T\right)} \tag{1.14}$$

is different from (1.13) while the size of $\boldsymbol{\Gamma}$ equals the column dimension of $\boldsymbol{\Theta}$. From Theorem B.17, Theorem B.21 in Appendix B, one can conclude the following relationship between Frobenius norm and weighted Frobenius norm.

$$\overline{\lambda}_\Gamma^{-1}\|\boldsymbol{\Theta}\|_F^2 \le \|\boldsymbol{\Theta}\|_\Gamma \le \underline{\lambda}_\Gamma^{-1}\|\boldsymbol{\Theta}\|_F^2 \tag{1.15}$$

There is also a generalization of the Cauchy-Schwartz inequality to weighed Frobenius norms.

$$\left|\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}^*\right)\right| \le \|\boldsymbol{\Theta}\|_\Gamma\left\|\boldsymbol{\Theta}^*\right\|_\Gamma \tag{1.16}$$

for $\boldsymbol{\Theta}^*\in\mathbb{R}^{n\times m}$. In order to show this, let $\boldsymbol{\theta}_i, \boldsymbol{\theta}_i^*$ be the columns of $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}^*$ respectively. The left hand side of (1.16) evaluates to

$$\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}^*\right) = \boldsymbol{\theta}_1^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\theta}_1^* + \ldots + \boldsymbol{\theta}_m^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\theta}_m^* . \tag{1.17}$$

With $\boldsymbol{\theta}_{is}=\mathbf{S}^T\boldsymbol{\theta}_i$ and $\boldsymbol{\theta}_{is}^*=\mathbf{S}^T\boldsymbol{\theta}_i^*$, where $\mathbf{S}$ is the Cholesky factorization of $\boldsymbol{\Gamma}^{-1}$, we get

$$\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}^*\right) = \boldsymbol{\theta}_{1s}^T\boldsymbol{\theta}_{1s}^* + \ldots + \boldsymbol{\theta}_{ms}^T\boldsymbol{\theta}_{ms}^*$$

and if all $\boldsymbol{\theta}_{is}$ and $\boldsymbol{\theta}_{is}^{*}$ are stacked into a column vector, such that $\boldsymbol{\theta}_{s}^{T}=\left(\boldsymbol{\theta}_{1s}^{T} \quad \cdots \quad \boldsymbol{\theta}_{ms}^{T}\right)$, $\boldsymbol{\theta}_{s}^{*T}=\left(\boldsymbol{\theta}_{1s}^{*T} \quad \cdots \quad \boldsymbol{\theta}_{ms}^{*T}\right)$, we obtain

$$\operatorname{tr}\left(\boldsymbol{\Theta}^{T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}^{*}\right)=\boldsymbol{\theta}_{s}^{T}\boldsymbol{\theta}_{s}^{*}.$$

Now Cauchy-Schwartz inequality can be applied.

$$\left|\operatorname{tr}\left(\boldsymbol{\Theta}^{T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}^{*}\right)\right| \leq \sqrt{\boldsymbol{\theta}_{s}^{T}\boldsymbol{\theta}_{s}}\sqrt{\boldsymbol{\theta}_{s}^{*T}\boldsymbol{\theta}_{s}^{*}}=\sqrt{\boldsymbol{\theta}_{1s}^{T}\boldsymbol{\theta}_{1s}+\ldots+\boldsymbol{\theta}_{ms}^{T}\boldsymbol{\theta}_{ms}}\sqrt{\boldsymbol{\theta}_{1s}^{*T}\boldsymbol{\theta}_{1s}^{*}+\ldots+\boldsymbol{\theta}_{ms}^{*T}\boldsymbol{\theta}_{ms}^{*}}$$

Using $\boldsymbol{\Gamma}^{-1}=\mathbf{S}\mathbf{S}^{T}$ and transforming the expression back to trace notation, analogous to (1.17) yields (1.16).

$$\begin{aligned}\left|\operatorname{tr}\left(\boldsymbol{\Theta}^{T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}^{*}\right)\right| &\leq \sqrt{\boldsymbol{\theta}_{1}^{T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\theta}_{1}+\ldots+\boldsymbol{\theta}_{m}^{T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\theta}_{m}}\sqrt{\boldsymbol{\theta}_{1}^{*T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\theta}_{1}^{*}+\ldots+\boldsymbol{\theta}_{m}^{*T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\theta}_{m}^{*}}\\ &=\sqrt{\operatorname{tr}\left(\boldsymbol{\Theta}^{T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}\right)}\sqrt{\operatorname{tr}\left(\boldsymbol{\Theta}^{*T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}^{*}\right)}=\|\boldsymbol{\Theta}\|_{\Gamma}\|\boldsymbol{\Theta}^{*}\|_{\Gamma}\end{aligned}$$

Weighted Euclidean norm can also be interpreted graphically. While, for classical Euclidean norm, the contour lines of constant norm are balls, in the weighted case, they are ellipsoids as depicted in Figure 1.1.

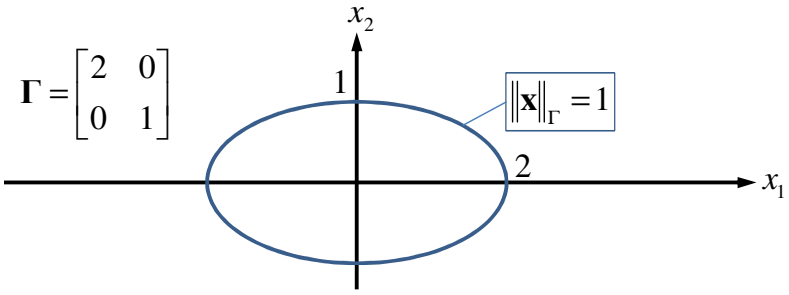


Figure 1.1 Weighted Euclidean Norm

For a real valued vector field $\mathbf{f}(\mathbf{x})$, $\mathbf{x}^{T}=(x_1,\ldots,x_n)$, $\mathbf{f}^{T}(\mathbf{x})=(f_1(\mathbf{x}),\ldots,f_m(\mathbf{x}))$, the Jacobian is defined as

$$\mathbf{J}_{f}(\mathbf{x})=\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}}=\begin{bmatrix}\frac{\partial f_1(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n}\\ \vdots & & \vdots\\ \frac{\partial f_m(\mathbf{x})}{\partial x_1} & \cdots & \frac{\partial f_m(\mathbf{x})}{\partial x_n}\end{bmatrix}$$

For scalar fields $f:\mathbb{R}^{n}\rightarrow\mathbb{R}$, the Jacobian is a row vector. However, the gradient is defined as a column vector, which is obtained by taking the transpose of the Jacobian.

$$\operatorname{grad}\left(f(\mathbf{x})\right)=\nabla f(\mathbf{x})=\left(\frac{\partial f(\mathbf{x})}{\partial\mathbf{x}}\right)^{T}=\begin{pmatrix}\frac{\partial f(\mathbf{x})}{\partial x_1}\\ \vdots\\ \frac{\partial f(\mathbf{x})}{\partial x_n}\end{pmatrix}$$

Chapter 2

Aircraft Modeling

2.1 Flying Testbed – The FSD Extreme Star

The FSD Extreme Star is a small UAV and has been developed at the Institute of Flight System Dynamics (FSD) in cooperation with AkaModell, a student group that constructs and operates remotely piloted aircraft, but also dedicate themselves to fundamental research on low Reynolds numbers ([Aka12]). Both institutions are affiliated to Technische Universität München (TUM).

Technical Data

Figure 2.1 shows the aircraft which is in fact a modified version of the commercial "Twin Star" developed and produced by Multiplex ([Mul12]). It is intended to serve as an institute's testbed for classical linear, nonlinear and adaptive flight controllers as well as control allocation algorithms. Particularly for the latter, it is desirable to have as many control surfaces as possible, as control allocation is concerned with full exploitation of redundancies in the control channel in order to optimize secondary objectives or to provide fault-tolerant systems in case of actuator failure.

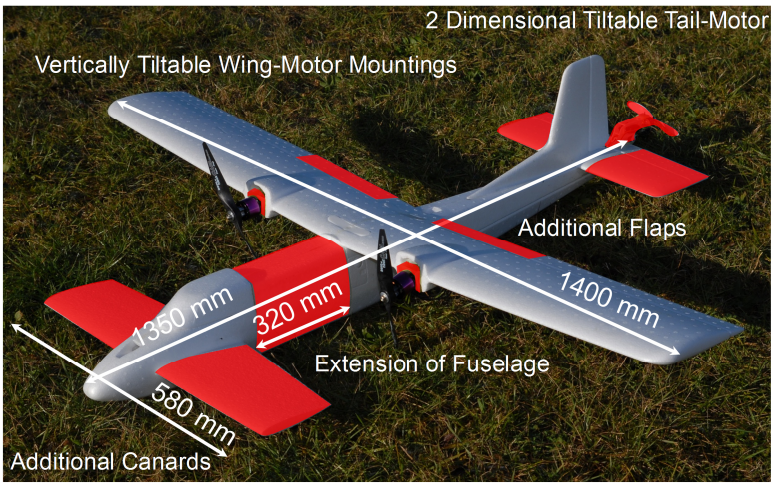


Figure 2.1 FSD Extreme Star with Modifications, [Bau10]

The following modifications were applied to the FSD Extreme Star

- additional separately controlled canards

- elongation of fuselage to provide space for payload (particularly flight control computer, sensors, receiver)
- additional separately controlled flaps at the inner part of the wing
- 2 main engines, fixed on the left and right hand wing, including pitch-axis thrust vectoring
- 1 auxiliary engine, mounted at the tail, including pitch-yaw-axis thrust vectoring
- split elevator control surfaces, where the both parts are controlled separately

Overall, there are 16 controls available, which render the actuation highly redundant. Table 2.1 presents an overview of all control surfaces and Table 2.2 lists the main technical data.

Table 2.1 Extreme Star Controls

control	symbol	unit
left/right canard deflection	$\eta_{c,l}$ / $\eta_{c,r}$	°
left/right aileron deflection	ξ_l / ξ_r	°
left/right flap deflection	δ_l / δ_r	°
left/right elevator deflection	η_l / η_r	°
rudder deflection	ζ	°
left/right main propeller rotation speed	$\omega_{M,l}$ / $\omega_{M,r}$	°/s
back auxiliary propeller rotation speed	$\omega_{M,b}$	°/s
left/right main engine thrust vector elevation angle	σ_l / σ_r	°
back auxiliary engine thrust vector elevation angle	σ_b	°
back auxiliary engine thrust vector azimuth angle	κ_b	°

Table 2.2 FSD Extreme Star: Main Technical Data

parameter		value
total mass		3,2 kg
wing	span	1400 mm
	mean aerodynamic chord	223 mm
	reference area	0.3014 m^2
	airfoil	CLARK Y
canard	span	580 mm
	mean aerodynamic chord	170 mm
	reference area	0.0747 m^2
	airfoil	NACA 0013
horizontal tail	span	420 mm
	mean aerodynamic chord	170 mm
	reference area	0.0693 m^2
fuselage length		1350 mm
propulsion	power left/right main engine	500 W
	power back auxiliary engine	75 W

Numeric Simulation

For development of flight control systems, it is inevitable to have a mathematical model, representing the dynamics. It serves as a basis for design of the control algorithm and tuning of the associated parameters. It is thereby desirable for the model to be as accurate as possible. However, also the modeling effort increases with the level of detail and hence a trade-off between accuracy and modeling effort has to be found.

In [Bau10] a simulation model has been developed and implemented in MATLAB/SIMULINK comprising a high level of detail. It includes the rigid body dynamics, which are derived from the principles of linear, and angular momentum and kinematic considerations that propagate aircraft attitude and position. It additionally comprises dynamics of subsystems such as actuator and sensor. Commonly such subsystem dynamics are much faster than the rigid body dynamics, but the incorporation of the former in the design of flight control systems plays an important role, since neglect of them usually results in too high gains for the flight control system and even might lead to instability, when the controller is applied to the real system.

Aerodynamics as well as propulsive forces and moments, acting onto the vehicle, depend on model parameters, which can be obtained by various methods. One option is the identification of the model parameters by flight experiments using methods of flight system identification ([Jat06], [Kle06]). Another possibility is given by wind tunnel tests. In our case, however, the aerodynamic and propulsive parameters were obtained by a MATLAB tool, developed in cooperation of FSD and Bauhaus Luftfahrt ([Bau12]). It allows for the computation of aerodynamic coefficients and derivatives of a geometric composition of lifting surfaces and rotors, accounting for interaction between the elements. It uses methods based on potential flow theory for lifting surfaces and momentum and vortex theory for computation of the induced flow of propellers. Inputs to the tool are geometric data as well as 2D-profile data of the lifting surfaces. Clearly, the parameters thus obtained are subjected to uncertainties. Nevertheless the parameters should be sufficiently accurate, which is particularly legitimate since adaptive controllers have the inherent property to compensate for uncertainties.

2.2 Rigid Body Equations of Motion

In this section, the rigid body differential equations, describing the aircraft motion in Euclidean space, are derived. The rigid body states are reasonably divided into four groups for description of translational and rotational motions, attitude as well as position.

The translation and rotation equations of motion (EOM) are derived from the linear and angular momentum principle according to Newton’s second law. In a first step all

forces and moments, acting onto the aircraft, are collected and applied to the momentum principles as a whole and a further specification of aerodynamic, propulsion and gravitational forces and moments, which depend on the rigid body states, is presented in section 2.3. While translation and rotation dynamics highly dependent on model parameters, the differential equations for attitude and position are obtained by pure kinematic considerations.

As the aircraft motion is quite complex, it involves various coordinate frames which move relative to each other translational and rotational. It is well-known that Newton's second law only holds w.r.t. frames that are not accelerated (inertial). For aircraft applications, the "earth centered inertial" frame (ECI), which has its origin at the earth center and does not rotate with the earth, is considered as inertial frame. Other frames, import for the derivations are the "earth centered earth fixed" frame (ECEF) which, contrary to ECI frame, rotates with the earth. The "north-east-down" frame (NED) is a local frame, with its origin at an aircraft-fixed reference point and whose x- and y-axes are aligned with north and east direction, while the z-axis points downwards, vertically to the local tangent plane of the earth surface. The body-fixed frame, has its origin and axis fixed with the aircraft. Descriptions of the utilized frames as well as angles that define the rotation between different frames are contained in Appendix A.

The subsequent equations naturally involve Euclidean vectors in various frames. In order not to lose track, a very well-arranged and comprehensive notation has been developed at the FSD, which will be introduced briefly.

A position vector $\vec{\mathbf{r}}$ from point G to R, where the components are written in a frame B, is written as

$$\left(\vec{\mathbf{r}}^{GR}\right)_B = \begin{pmatrix} \left(x^{GR}\right)_B \\ \left(y^{GR}\right)_B \\ \left(z^{GR}\right)_B \end{pmatrix}$$

In presence of different frames that rotate relative to each other, we have angular rates that specify that relative motions. Let us assume that frame B rotates relative to some frame O. The angular rate of B relative to O, whose components are given in B frame, is written as

$$\left(\vec{\boldsymbol{\omega}}^{OB}\right)_B = \begin{pmatrix} \left(\omega_x^{OB}\right)_B \\ \left(\omega_y^{OB}\right)_B \\ \left(\omega_z^{OB}\right)_B \end{pmatrix}.$$

If a position vector is derived w.r.t. time, the frame of reference has to be specified. Hence, if the position vector above is derived w.r.t to B frame, we write

$$\left(\dot{\vec{\mathbf{r}}}^{GR}\right)_B^B$$

and the same holds for the time derivative of angular rates.

$$\left(\dot{\vec{\boldsymbol{\omega}}}^{OB}\right)_B^B$$

If the vector is to be derived w.r.t. to the O frame, this cannot be done directly since its components are written in B, but by help Euler's differentiation rule.

$$\left(\dot{\vec{\mathbf{r}}}^{GR}\right)_B^O = \left(\dot{\vec{\mathbf{r}}}^{GR}\right)_B^B + \left(\vec{\boldsymbol{\omega}}^{OB}\right)_B \times \left(\vec{\mathbf{r}}^{GR}\right)_B .$$

Further, as EOMs intensively use velocities, a short notation for the first time derivative of the position vector is given by

$$\left(\vec{\mathbf{V}}^{GR}\right)_B^O = \left(\dot{\vec{\mathbf{r}}}^{GR}\right)_B^O = \begin{pmatrix} \left(u^{GR}\right)_B^O \\ \left(v^{GR}\right)_B^O \\ \left(w^{GR}\right)_B^O \end{pmatrix}.$$

The time derivative of the velocity vector can be taken w.r.t. different reference frames, too and hence, if we derive the velocity once more w.r.t B frame, we obtain

$$\left(\dot{\vec{\mathbf{V}}}^{GR}\right)_B^{OB} .$$

As wind plays an important role in aircraft dynamics, aerodynamic and kinematic velocities have to be distinguished.

kinematic: $\left(\vec{\mathbf{V}}_K^{GR}\right)_B^B$, aerodynamic: $\left(\vec{\mathbf{V}}_A^{GR}\right)_B^B$, wind: $\left(\vec{\mathbf{V}}_W^{GR}\right)_B^B$

A similar notation is introduced for forces and moments, e.g.

$$\left(\vec{\mathbf{F}}_{\mathrm{Pr}}^R\right)_B = \begin{pmatrix} \left(X_{\mathrm{Pr}}^R\right)_B \\ \left(Y_{\mathrm{Pr}}^R\right)_B \\ \left(Z_{\mathrm{Pr}}^R\right)_B \end{pmatrix}, \left(\vec{\mathbf{M}}_{\mathrm{Pr}}^R\right)_B = \begin{pmatrix} \left(L_{\mathrm{Pr}}^R\right)_B \\ \left(M_{\mathrm{Pr}}^R\right)_B \\ \left(N_{\mathrm{Pr}}^R\right)_B \end{pmatrix}$$

denote propulsive forces and moments relative to aircraft reference point R with components written in aircraft-fixed frame B frame. Also of great importance in this framework are transformation matrices between different frames, e.g. a transformation matrix from O to B is written as

$$\mathbf{M}_{BO}$$

and, since transformation matrices are orthonormal, the inverse operation is simply obtained by taking the transpose.

$$\mathbf{M}_{OB} = \mathbf{M}_{BO}^T$$

Moreover, the following simplifications are declared.

- Since point 0, the origin of ECI frame, is the inertial reference point, it is dropped.

- If the frame of notation does not play a role, it is dropped.

As the ECI frame is considered inertial, the equations incorporate effects due to earth rotation such as centrifugal and Coriolis forces. Nevertheless, the following simplifications are assumed:

- The earth rotation rate is considered constant: $\left(\dot{\vec{\boldsymbol{\omega}}}^{IE}\right)^{E} = \mathbf{0}$
- The vehicle is considered as a rigid body, i.e. motion of mass elements relative to each other is excluded: $\left(\dot{\vec{\mathbf{r}}}^{RP}\right)^{B} = 0$, where R is the aircraft reference point and P is some arbitrary aircraft-fixed point.

The aircraft reference point (index R) is defined geometrically and therefore aircraft-fixed. Generally, it deviates from the center of gravity (index G) which G could move due to different payloads and hence the EOMs are derived w.r.t R. Moreover, it has to be mentioned that the derivation of EOMs subsequently is accomplished independently of a frame of notation.

Generally, the choice of the rigid body states is not unique and hence the states that will be used are introduced in the following.

Translation

Translation is appropriately described by the kinematic velocity of some aircraft-fixed reference point (index R) relative to the earth surface, which complies with ECEF frame (index E), where the components are given in the aircraft-fixed frame (index B):

$$\left(\vec{\mathbf{V}}_K^R\right)_B^E = \begin{pmatrix} \left(u_K^R\right)_B^E \\ \left(v_K^R\right)_B^E \\ \left(w_K^R\right)_B^E \end{pmatrix} \tag{2.1}$$

Equivalently, the translation states can also be described in terms of

absolute kinematic velocity: $\left(V_K^R\right)_B^E$

kinematic angle of attack: $\left(\alpha_K^R\right)_B^E$

kinematic angle of side-slip: $\left(\beta_K^R\right)_B^E$.

Notice that α_K is defined in the range $[-\pi, \pi]$ and β_K is defined in the range $[-0.5\pi, 0.5\pi]$.

Rotation

The aircraft's rotation is described by the angular velocity of the aircraft-fixed frame relative to the ECI frame with components written in aircraft-fixed frame.

$$\left(\vec{\boldsymbol{\omega}}^{IB}\right)_B = \begin{pmatrix} p \\ q \\ r \end{pmatrix} \tag{2.2}$$

Attitude

The attitude is described relative to the local tangent plane onto the earth surface, in fact by the angles between NED and body-fixed frame. In aviation the Euler angles, that describe the rotation of the aircraft-fixed frame relative to the NED frame, the used as attitude parameterization (Appendix A).

$$\boldsymbol{\Psi} = \begin{pmatrix} \Psi \\ \Theta \\ \Phi \end{pmatrix}$$

Alternatively, the rotation from NED to aircraft-fixed frame is described by quaternions which avoid the singularity of the Euler angle parameterization at $\Theta = \pm 0.5\pi$ (refer to B.6).

$$\underline{\mathbf{q}}^T = \begin{pmatrix} q_0 & q_1 & q_2 & q_3 \end{pmatrix}$$

Position

On a global scale, the aircraft position is commonly described by WGS-84 coordinates ([EUR98]).

geodetic longitude: λ
geodetic latitude: ϕ
altitude above reference ellipsoid: h

Alternatively, also Cartesian ECEF coordinates could be used.

$$\left(\vec{\mathbf{r}}^R\right)_E = \begin{pmatrix} \left(x^R\right)_E \\ \left(y^R\right)_E \\ \left(z^R\right)_E \end{pmatrix}$$

2.2.1 Translation and Rotation Dynamics

Translation Dynamics

The ECI frame is considered as inertial frame and consequently its origin (index *0*) is an inertial point and therefore qualifies for formulation of the principles of linear and angular momentum. With R, some aircraft-fixed reference point, G the center of gravity and P, some arbitrary point of the aircraft (Figure 2.2), the position of P relative to earth center is

$$\left(\vec{\mathbf{r}}^{\,P}\right) = \left(\vec{\mathbf{r}}^{\,R}\right) + \left(\vec{\mathbf{r}}^{\,RG}\right) + \left(\vec{\mathbf{r}}^{\,GP}\right) \tag{2.3}$$

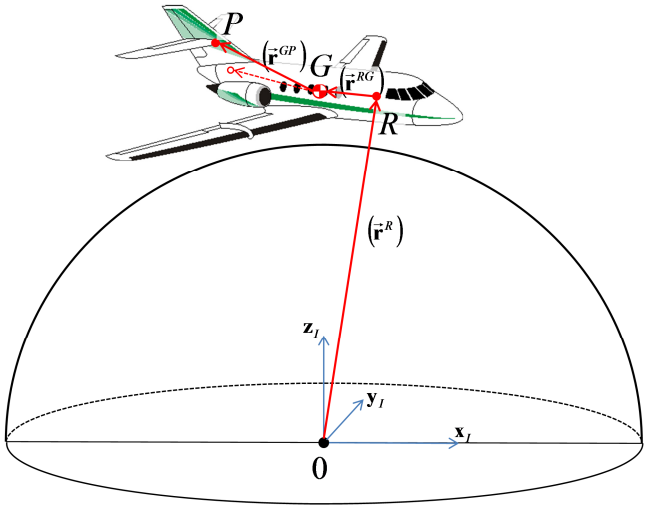


Figure 2.2 Aircraft Reference Point

Using Euler's differentiation rule, the velocity is derived as

$$\begin{aligned}\left(\vec{\mathbf{V}}_K^{\,P}\right)^I &= \left(\dot{\vec{\mathbf{r}}}^{\,R}\right)^I + \left(\dot{\vec{\mathbf{r}}}^{\,RP}\right)^I = \left(\vec{\mathbf{V}}_K^{\,R}\right)^E + \left(\vec{\boldsymbol{\omega}}^{\,IE}\right)\times\left(\vec{\mathbf{r}}^{\,R}\right) + \underbrace{\left(\dot{\vec{\mathbf{r}}}^{\,RP}\right)^B}_{=0} + \left(\vec{\boldsymbol{\omega}}^{\,IB}\right)\times\left[\left(\vec{\mathbf{r}}^{\,RG}\right) + \left(\vec{\mathbf{r}}^{\,GP}\right)\right] \\ &= \underbrace{\left(\vec{\mathbf{V}}_K^{\,R}\right)^E + \left(\vec{\boldsymbol{\omega}}^{\,IE}\right)\times\left(\vec{\mathbf{r}}^{\,R}\right)}_{\left(\vec{\mathbf{V}}_K^{\,R}\right)^I} + \underbrace{\left(\vec{\boldsymbol{\omega}}^{\,IB}\right)\times\left(\vec{\mathbf{r}}^{\,RG}\right)}_{+\left(\vec{\mathbf{V}}_K^{\,RG}\right)^I} + \underbrace{\left(\vec{\boldsymbol{\omega}}^{\,IB}\right)\times\left(\vec{\mathbf{r}}^{\,GP}\right)}_{+\left(\vec{\mathbf{V}}_K^{\,GP}\right)^I}\end{aligned} \tag{2.4}$$

The center of gravity (c.g.) of a rigid body is defined such that

$$\int_m \left(\vec{\mathbf{r}}^{\,GP}\right)\cdot dm = \mathbf{0} \tag{2.5}$$

Further the linear momentum of a mass element of mass dm at some point P is

$$\begin{aligned}d\vec{\mathbf{p}} &= \left(\vec{\mathbf{V}}_K^{\,P}\right)^I \cdot dm \\ &= \left(\vec{\mathbf{V}}_K^{\,R}\right)^I \cdot dm + \left(\vec{\boldsymbol{\omega}}^{\,IB}\right)\times\left[\left(\vec{\mathbf{r}}^{\,RG}\right) + \left(\vec{\mathbf{r}}^{\,GP}\right)\right]\cdot dm\end{aligned} \tag{2.6}$$

where (2.4) has been utilized. The aircraft linear momentum is obtained by integrating (2.6) over all mass elements

$$\vec{\mathbf{p}} = \int_m d\mathbf{p} = \int_m \left(\vec{\mathbf{V}}_K^R\right)^I \cdot dm + \left(\vec{\boldsymbol{\omega}}^{IB}\right) \times \int_m \left(\vec{\mathbf{r}}^{RG}\right) \cdot dm + \left(\vec{\boldsymbol{\omega}}^{IB}\right) \times \underbrace{\int_m \left(\vec{\mathbf{r}}^{GP}\right) \cdot dm}_{=\mathbf{0}}$$

$$\vec{\mathbf{p}} = \left(\vec{\mathbf{V}}^R\right)^I \cdot m + \left(\vec{\boldsymbol{\omega}}^{IB}\right) \times \left(\vec{\mathbf{r}}^{RG}\right) \cdot m \tag{2.7}$$

Thereby, the c.g. condition (2.5) has been used and expressions that are constant over the aircraft body are written outside of the respective integrals. The principle of linear momentum says that the inertial time derivative of linear momentum equals the sum of all external forces. Therefore, derive (2.7) once w.r.t. the ECI frame.

$$\sum\left(\vec{\mathbf{F}}^R\right) = \left(\dot{\vec{\mathbf{p}}}\right)^I = \left(\left(\dot{\vec{\mathbf{V}}}_K^R\right)^{II} + \left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^B \times \left(\vec{\mathbf{r}}^{RG}\right) + \left(\vec{\boldsymbol{\omega}}^{IB}\right) \times \left(\mathbf{V}^{RG}\right)^I \right) \cdot m \tag{2.8}$$

Note that we have applied the Euler differentiation rule to $\left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^I$:

$$\left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^I = \left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^B + \overbrace{\left(\vec{\boldsymbol{\omega}}^{IB}\right) \times \left(\vec{\boldsymbol{\omega}}^{IB}\right)}^{=0} = \left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^B \tag{2.9}$$

It is left to derive an expression for $\left(\dot{\vec{\mathbf{V}}}_K^R\right)^{II}$. Note that we are seeking for an expression in terms of the velocity relative to the earth surface $\left(\vec{\mathbf{V}}_K^R\right)^E$ since the rigid body states are defined accordingly. Deriving once w.r.t. ECI frame we obtain, using (2.4):

$$\begin{aligned}
\left(\vec{\mathbf{V}}_K^R\right)^{II} &= \left(\dot{\vec{\mathbf{V}}}_K^R\right)^{IE} + \left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left(\vec{\mathbf{V}}_K^R\right)^I \\
&= \left(\dot{\vec{\mathbf{V}}}_K^R\right)^{EE} + \left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left(\vec{\mathbf{V}}_K^R\right)^E + \underbrace{\left(\dot{\vec{\boldsymbol{\omega}}}^{IE}\right)^E}_{=0} \times \left(\vec{\mathbf{r}}^R\right) + \left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left[\left(\vec{\mathbf{V}}_K^R\right)^E + \left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left(\vec{\mathbf{r}}^R\right)\right] \\
\left(\dot{\vec{\mathbf{V}}}_K^R\right)^{II} &= \left(\dot{\vec{\mathbf{V}}}_K^R\right)^{EE} + 2 \cdot \left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left(\vec{\mathbf{V}}_K^R\right)^E + \left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left[\left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left(\vec{\mathbf{r}}^R\right)\right]
\end{aligned}$$

Note that some terms appear to be prominent in mechanics:

- $2 \cdot \left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left(\vec{\mathbf{V}}_K^R\right)^E$: Coriolis acceleration of reference point due to earth rotation
- $\left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left[\left(\vec{\boldsymbol{\omega}}^{IE}\right) \times \left(\vec{\mathbf{V}}_K^R\right)^E\right]$: centripetal acceleration of reference point due to earth rotation

Since the components of our translational states are written in body-fixed frame, the state derivative w.r.t. the body fixed frame is preferred in the formulation and thus we further derive

$$\begin{aligned}
\left(\dot{\vec{\mathbf{V}}}_K^R\right)^{EE} &= \left(\dot{\vec{\mathbf{V}}}_K^R\right)^{EB} + \left(\vec{\boldsymbol{\omega}}^{EB}\right) \times \left(\vec{\mathbf{V}}_K^R\right)^E \\
&= \left(\dot{\vec{\mathbf{V}}}_K^R\right)^{EB} + \left[\left(\vec{\boldsymbol{\omega}}^{EO}\right) + \left(\vec{\boldsymbol{\omega}}^{OB}\right)\right] \times \left(\vec{\mathbf{V}}_K^R\right)^E
\end{aligned}$$

where $(\vec{\boldsymbol{\omega}}^{EO})$ is denoted as transport rate and $(\vec{\boldsymbol{\omega}}^{OB})$ describes the rotation of the aircraft w.r.t. the local earth surface. We obtain

$$\left(\dot{\vec{\mathbf{V}}}_K^R\right)^I = \left(\dot{\vec{\mathbf{V}}}_K^R\right)^{EB} + \left(\vec{\boldsymbol{\omega}}^{EB}\right)\times\left(\vec{\mathbf{V}}_K^R\right)^E + 2\left(\vec{\boldsymbol{\omega}}^{IE}\right)\times\left(\vec{\mathbf{V}}_K^R\right)^E + \left(\vec{\boldsymbol{\omega}}^{IE}\right)\times\left[\left(\vec{\boldsymbol{\omega}}^{IE}\right)\times\left(\vec{\mathbf{r}}^R\right)\right] \tag{2.10}$$

The derivations so far have been done without specification of any frame of notation. But in order to obtain differential equations of the Cartesian translation states as defined above, we choose the aircraft-fixed frame. Applying the expressions, derived in (2.4), (2.10) to the principle of linear momentum (2.8) and solving for the state derivatives yields.

$$\begin{aligned}\left(\dot{\vec{\mathbf{V}}}_K^R\right)_B^{EB} m + \left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)_B^B \times \left(\vec{\mathbf{r}}^{RG}\right)_B m = \quad & \sum\left(\vec{\mathbf{F}}^R\right)_B && -\left(\vec{\boldsymbol{\omega}}^{EB}\right)_B \times \left(\vec{\mathbf{V}}_K^R\right)_B^E m \\ & -2\left(\vec{\boldsymbol{\omega}}^{IE}\right)_B \times \left(\vec{\mathbf{V}}_K^R\right)_B^E m && -\left(\vec{\boldsymbol{\omega}}^{IE}\right)_B \times \left[\left(\vec{\boldsymbol{\omega}}^{IE}\right)_B \times \left(\vec{\mathbf{r}}^R\right)_B\right] m \\ & && -\left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \times \left[\left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \times \left(\vec{\mathbf{r}}^{RG}\right)_B\right] m\end{aligned} \tag{2.11}$$

where $\left(\vec{\mathbf{V}}_K^R\right)_B^E$ is defined in (2.1).

Rotation Dynamics

The angular momentum of a mass element relative to the inertial earth center is defined as

$$d\left(\vec{\mathbf{H}}^0\right) = \left(\vec{\mathbf{r}}^P\right)\times\left(\vec{\mathbf{V}}_K^P\right)^I dm \tag{2.12}$$

using (2.3) and (2.4), we obtain

$$\begin{aligned} d\left(\vec{\mathbf{H}}^0\right) &= \left[\left(\vec{\mathbf{r}}^R\right) + \left(\vec{\mathbf{r}}^{RP}\right)\right]\times\left[\left(\vec{\mathbf{V}}_K^R\right)^I + \left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\left(\vec{\mathbf{r}}^{RP}\right)\right]dm \\ &= \left(\vec{\mathbf{r}}^R\right)\times\left[\left(\vec{\mathbf{V}}_K^R\right)^I + \left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\left(\left(\vec{\mathbf{r}}^{RG}\right) + \left(\vec{\mathbf{r}}^{GP}\right)\right)\right]dm + \left(\vec{\mathbf{r}}^{RP}\right)\times\left[\left(\vec{\mathbf{V}}_K^R\right)^I + \left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\left(\vec{\mathbf{r}}^{RP}\right)\right]dm \end{aligned}$$

Integration over the body yields the aircraft angular momentum.

$$\left(\vec{\mathbf{H}}^0\right) = \left(\vec{\mathbf{r}}^R\right)\times\left\{\left(\vec{\mathbf{V}}_K^R\right)^I m + \left[\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\left(\vec{\mathbf{r}}^{RG}\right)\right]\right\}m + \left(\vec{\mathbf{r}}^{RG}\right)\times\left(\vec{\mathbf{V}}_K^R\right)^I m + \int_m \left(\vec{\mathbf{r}}^{RP}\right)\times\left[\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\left(\vec{\mathbf{r}}^{RP}\right)\right]dm \tag{2.13}$$

Note that the term containing $(\vec{\mathbf{r}}^{GP})$ has canceled out due to the c.g. condition (2.5). The last term in (2.13) has to be resolved further. Therefore, define the vector $(\vec{\mathbf{r}}^{RP})$ component wise as

$$\left(\vec{\mathbf{r}}^{RP}\right) = \begin{pmatrix} x^{RP} \\ y^{RP} \\ z^{RP} \end{pmatrix}$$

and use the fact that the cross product can also be written as matrix-vector product involving a skew-symmetric matrix

$$\left(\vec{\mathbf{r}}^{RB}\right)\times\vec{\mathbf{x}}=\tilde{\mathbf{R}}^{RB}\cdot\vec{\mathbf{x}} \tag{2.14}$$

for some $\mathbf{x}\in\mathbb{R}^3$, where

$$\tilde{\mathbf{R}}^{RB}=\begin{bmatrix}0 & -z^{RB} & y^{RB}\\ z^{RB} & 0 & -x^{RB}\\ -y^{RB} & z^{RB} & 0\end{bmatrix}$$

Employing the anti commutativity property of the cross product, the integral term is

$$\begin{aligned}\int_m\left(\vec{\mathbf{r}}^{RP}\right)\times\left[\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\left(\vec{\mathbf{r}}^{RP}\right)\right]dm&=-\int_m\left(\vec{\mathbf{r}}^{RP}\right)\times\left[\left(\vec{\mathbf{r}}^{RP}\right)\times\left(\vec{\boldsymbol{\omega}}^{IB}\right)\right]dm\\&=-\int_m\tilde{\mathbf{R}}^{RP}\cdot\tilde{\mathbf{R}}^{RP}dm\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)\end{aligned}$$

The matrix product within the integral can be further expanded.

$$-\int_m\tilde{\mathbf{R}}^{RP}\cdot\tilde{\mathbf{R}}^{RP}dm=\begin{bmatrix}\int_m\left(y^{RP}\right)^2+\left(z^{RP}\right)^2dm & -\int_m\left(x^{RP}\right)\left(y^{RP}\right)dm & -\int_m\left(x^{RP}\right)\left(z^{RP}\right)dm\\ -\int_m\left(x^{RP}\right)\left(y^{RP}\right)dm & \int_m\left(x^{RP}\right)^2+\left(z^{RP}\right)^2dm & -\int_m\left(y^{RP}\right)\left(z^{RP}\right)dm\\ -\int_m\left(x^{RP}\right)\left(z^{RP}\right)dm & -\int_m\left(y^{RP}\right)\left(z^{RP}\right)dm & \int_m\left(x^{RP}\right)^2+\left(y^{RP}\right)^2dm\end{bmatrix}=:\mathbf{I}^R=:\begin{bmatrix}I_x^R & I_{xy}^R & I_{xz}^R\\ I_{xy}^R & I_y^R & I_{yz}^R\\ I_{xz}^R & I_{yz}^R & I_z^R\end{bmatrix} \tag{2.15}$$

$\mathbf{I}^R$ is the mass moment of inertia w.r.t. R. Further, note that the expression in curly brackets in (2.13) equals the linear momentum defined in (2.7). With this insight and the definition of moment of inertia, the angular momentum reads quite compact

$$\left(\vec{\mathbf{H}}^0\right)=\left(\vec{\mathbf{r}}^R\right)\times\left(\vec{\mathbf{p}}\right)+\left(\vec{\mathbf{r}}^{RG}\right)\times\left(\vec{\mathbf{V}}_K^R\right)^I m+\mathbf{I}^R\left(\vec{\boldsymbol{\omega}}^{IB}\right) \tag{2.16}$$

The principal of angular momentum says that the inertial time derivative of the angular momentum equals the sum of all moments, relative to the inertial earth center

$$\left(\dot{\vec{\mathbf{H}}}^0\right)^I=\sum\left(\vec{\mathbf{M}}^0\right)_B \tag{2.17}$$

Hence (2.16) has to be derived once w.r.t. the ECI frame.

$$\left(\dot{\vec{\mathbf{H}}}^0\right)_B^I=\left(\vec{\mathbf{V}}_K^R\right)^I\times\left(\vec{\mathbf{p}}\right)+\left(\vec{\mathbf{r}}^R\right)\times\left(\dot{\vec{\mathbf{p}}}\right)^I+\left[\left(\vec{\mathbf{V}}_K^{RG}\right)^I\times\left(\vec{\mathbf{V}}_K^R\right)^I+\left(\vec{\mathbf{r}}^{RG}\right)\times\left(\dot{\vec{\mathbf{V}}}_K^R\right)^{II}\right]\cdot m+\left(\frac{d}{dt}\right)^I\left(\mathbf{I}^R\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)\right) \tag{2.18}$$

The last term is further specified using Euler's differentiation rule

$$\left(\frac{d}{dt}\right)^I\left(\mathbf{I}^R\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)\right)=\left(\frac{d}{dt}\right)^B\left(\mathbf{I}^R\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)\right)+\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\left(\mathbf{I}_R\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)\right)$$

Then, since the aircraft is assumed to be a rigid body, the inertia tensor is constant w.r.t. the body-fixed frame and hence

$$\left(\frac{d}{dt}\right)^I\left(\mathbf{I}^R\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)\right)=\mathbf{I}^R\cdot\left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^B+\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\mathbf{I}^R\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right) \tag{2.19}$$

By a shift of reference point, the moment w.r.t. to earth center is equivalently written as

$$\begin{aligned}\sum\left(\vec{\mathbf{M}}^{0}\right) &= \sum\left(\vec{\mathbf{M}}^{R}\right)+\left(\vec{\mathbf{r}}^{R}\right)\times\sum\left(\vec{\mathbf{F}}^{R}\right)\\ &= \sum\left(\vec{\mathbf{M}}^{R}\right)+\left(\vec{\mathbf{r}}^{R}\right)\times\left(\dot{\vec{\mathbf{p}}}\right)^{I}\end{aligned} \tag{2.20}$$

where the principle of linear momentum in (2.8) has been used. On the other hand (2.18) contains the same lever arm correction term and hence it can be canceled out in the principal of angular momentum (2.17).

$$\begin{aligned}\sum\left(\vec{\mathbf{M}}^{R}\right) &= \left(\vec{\mathbf{V}}_{K}^{R}\right)^{I}\times\left\{\left(\vec{\mathbf{V}}_{K}^{R}\right)^{I}+\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\left(\vec{\mathbf{r}}^{RG}\right)\right\}\cdot m\\ &+\left[\left(\vec{\mathbf{V}}_{K}^{RG}\right)^{I}\times\left(\vec{\mathbf{V}}_{K}^{R}\right)^{I}+\left(\vec{\mathbf{r}}^{RG}\right)\times\left(\dot{\vec{\mathbf{V}}}_{K}^{R}\right)^{II}\right]\cdot m+\mathbf{I}^{R}\cdot\left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^{B}+\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\mathbf{I}^{R}\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)\end{aligned}$$

Note that also (2.19) and the definition of linear momentum (2.8) have been inserted. It reveals that the term $\left(\vec{\mathbf{V}}_{K}^{R}\right)^{I}\times\left(\vec{\mathbf{V}}_{K}^{R}\right)^{I}$ results to 0. Using the definition of $\left(\vec{\mathbf{V}}_{k}^{RG}\right)^{I}$ in (2.4) we further get

$$\sum\left(\vec{\mathbf{M}}^{R}\right)=\left[\left(\vec{\mathbf{V}}_{K}^{R}\right)^{I}\times\left(\vec{\mathbf{V}}_{K}^{RG}\right)^{I}+\left(\vec{\mathbf{V}}_{K}^{RG}\right)^{I}\times\left(\vec{\mathbf{V}}_{K}^{R}\right)^{I}+\left(\vec{\mathbf{r}}^{RG}\right)\times\left(\dot{\vec{\mathbf{V}}}_{K}^{R}\right)^{II}\right]\cdot m+\mathbf{I}^{R}\cdot\left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^{B}+\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\mathbf{I}^{R}\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)$$

and, due to anti commutativity of cross product, also the term

$$\left(\vec{\mathbf{V}}_{K}^{R}\right)^{I}\times\left(\vec{\mathbf{V}}_{K}^{RG}\right)^{I}$$

cancels out and we have left

$$\sum\left(\vec{\mathbf{M}}^{R}\right)=\left(\vec{\mathbf{r}}^{RG}\right)\times\left(\dot{\vec{\mathbf{V}}}_{K}^{R}\right)^{II}\cdot m+\mathbf{I}^{R}\cdot\left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^{B}+\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\mathbf{I}^{R}\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right).$$

With (2.7), we get

$$\begin{aligned}\sum\left(\vec{\mathbf{M}}^{R}\right) &= \left(\vec{\mathbf{r}}^{RG}\right)\times\left\{\left(\dot{\vec{\mathbf{V}}}_{K}^{R}\right)^{EB}+\left(\vec{\boldsymbol{\omega}}^{EB}\right)\times\left(\vec{\mathbf{V}}_{K}^{R}\right)^{E}+2\cdot\left(\vec{\boldsymbol{\omega}}^{IE}\right)\times\left(\vec{\mathbf{V}}_{K}^{R}\right)^{E}+\left(\vec{\boldsymbol{\omega}}^{IE}\right)\times\left[\left(\vec{\boldsymbol{\omega}}^{IE}\right)\times\left(\vec{\mathbf{r}}^{R}\right)\right]\right\}\cdot m\\ &+\mathbf{I}^{R}\cdot\left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)^{B}+\left(\vec{\boldsymbol{\omega}}^{IB}\right)\times\left[\mathbf{I}^{R}\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)\right].\end{aligned}$$

The equations have been derived without specification a frame of notation but due to the chosen rotation states, the equation are appropriately written in body-fixed axes. Finally, solving the latest equation for the state derivatives yields the rotation dynamics

$$\begin{aligned}m\cdot\left(\vec{\mathbf{r}}^{RG}\right)_{B}\times\left(\dot{\vec{\mathbf{V}}}_{K}^{R}\right)_{B}^{EB}+\left(\mathbf{I}^{R}\right)_{B}\cdot\left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)_{B}^{B} &= \sum\left(\vec{\mathbf{M}}^{R}\right)_{B}-\left(\vec{\boldsymbol{\omega}}^{IB}\right)_{B}\times\left[\left(\mathbf{I}^{R}\right)_{B}\cdot\left(\vec{\boldsymbol{\omega}}^{IB}\right)_{B}\right]\\ &-\left(\vec{\mathbf{r}}^{RG}\right)_{B}\times\left\{\begin{array}{l}\left(\vec{\boldsymbol{\omega}}^{EB}\right)_{B}\times\left(\vec{\mathbf{V}}_{K}^{R}\right)_{B}^{E}\\ +2\cdot\left(\vec{\boldsymbol{\omega}}^{IE}\right)_{B}\times\left(\vec{\mathbf{V}}^{R}\right)_{B}^{E}+\left(\vec{\boldsymbol{\omega}}^{IE}\right)_{B}\times\left[\left(\vec{\boldsymbol{\omega}}^{IE}\right)_{B}\times\left(\vec{\mathbf{r}}^{R}\right)_{B}\right]\end{array}\right\}\cdot m\end{aligned} \tag{2.21}$$

where $\left(\vec{\boldsymbol{\omega}}^{IB}\right)_{B}$ is defined in (2.2).

Coupled Solution of Translation and Rotation Dynamics

Taking a look to (2.11) and (2.21), it reveals that translation and rotation dynamics cannot be solved separately, since each equation contains both state derivatives. However using the alternative notation of cross product of (2.14), the coupled translation and rotation dynamics are:

$$\begin{bmatrix} m\mathbf{I} & -m\left(\tilde{\mathbf{R}}^{RG}\right)_B \\ m\left(\tilde{\mathbf{R}}^{RG}\right)_B & \left(\mathbf{I}^R\right)_B \end{bmatrix} \cdot \begin{pmatrix} \left(\dot{\vec{\mathbf{V}}}_K^R\right)_B^{EB} \\ \left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)_B^B \end{pmatrix} = \begin{pmatrix} \sum\left(\vec{\mathbf{F}}^R\right)_B - \mathbf{h}_1 \\ \sum\left(\vec{\mathbf{M}}^R\right)_B - \mathbf{h}_2 \end{pmatrix}$$

where

$$\mathbf{h}_1 = \left\{\left(\vec{\boldsymbol{\omega}}^{EB}\right)_B \times \left(\vec{\mathbf{V}}_K^R\right)_B^E + 2\cdot\left(\vec{\boldsymbol{\omega}}^{IE}\right)_B \times \left(\vec{\mathbf{V}}_K^R\right)_B^E + \left(\vec{\boldsymbol{\omega}}^{IE}\right)_B \times \left[\left(\vec{\boldsymbol{\omega}}^{IE}\right)_B \times \left(\vec{\mathbf{r}}^R\right)_B\right] + \left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \times \left[\left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \times \left(\vec{\mathbf{r}}^{RG}\right)_B\right]\right\}\cdot m$$

$$\mathbf{h}_2 = \left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \times \left[\left(\mathbf{I}^R\right)_B \cdot \left(\vec{\boldsymbol{\omega}}^{IB}\right)_B\right] + \left(\vec{\mathbf{r}}^{RG}\right)_B \times \left\{\left(\vec{\boldsymbol{\omega}}^{EB}\right)_B \times \left(\vec{\mathbf{V}}_K^R\right)_B^E + 2\cdot\left(\vec{\boldsymbol{\omega}}^{IE}\right)_B \times \left(\vec{\mathbf{V}}_K^R\right)_B^E + \left(\vec{\boldsymbol{\omega}}^{IE}\right)_B \times \left[\left(\vec{\boldsymbol{\omega}}^{IE}\right)_B \times \left(\vec{\mathbf{r}}^R\right)_B\right]\right\}\cdot m$$

and $\mathbf{I}$ denotes the identity matrix. Note further that the generalized mass matrix

$$\mathbf{M} = \begin{bmatrix} m\mathbf{I} & -m\left(\tilde{\mathbf{R}}^{RG}\right)_B \\ m\left(\tilde{\mathbf{R}}^{RG}\right)_B & \left(\mathbf{I}^R\right)_B \end{bmatrix}$$

is regular since, using the determinant formula for *2x2* block matrices (see e.g. [Lüt96]), its determinant is,

$$\det(\mathbf{M}) = m\cdot\det\left(\mathbf{I}^R + \tilde{\mathbf{R}}^{RG}\tilde{\mathbf{R}}^{RG}\right) = m\cdot\det\left(\mathbf{I}^G\right)$$

where $\mathbf{I}^G$ is the moment of inertia w.r.t the center of gravity and Steiner's Theorem has been used. Finally $\det\left(\mathbf{I}^G\right) = I_x^G\cdot I_y^G\cdot I_z^G > 0$ and hence, $\mathbf{M}$ is invertible and the coupled translation/rotation dynamics can be solved for the state derivatives:

$$\begin{pmatrix} \left(\dot{\vec{\mathbf{V}}}_K^R\right)^{EB} \\ \left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right) \end{pmatrix} = \mathbf{M}^{-1}\cdot\begin{pmatrix} \sum\left(\vec{\mathbf{F}}^R\right) - \mathbf{h}_1 \\ \sum\left(\vec{\mathbf{M}}^R\right) - \mathbf{h}_2 \end{pmatrix} \qquad (2.22)$$

2.2.2 Attitude Dynamics

Euler Angles

In aviation, the aircraft attitude is usually described by the Euler angles. These are Ψ (azimuth), Θ (pitch) and Φ (bank) (refer to Appendix A). The first time derivative of these quantities is related to the aircraft's angular rate relative to the NED frame by

$$\begin{pmatrix} \dot{\Phi} \\ \dot{\Theta} \\ \dot{\Psi} \end{pmatrix} = \begin{bmatrix} 1 & \tan\Theta\sin\Phi & \tan\Theta\cos\Phi \\ 0 & \cos\Phi & -\sin\Phi \\ 0 & \sin\Phi/\cos\Theta & \cos\Phi/\cos\Theta \end{bmatrix}\cdot\begin{pmatrix} p^{OB} \\ q^{OB} \\ r^{OB} \end{pmatrix} \qquad (2.23)$$

which is derived from kinematic considerations ([Hol12]). Thereby

$$\left(\vec{\boldsymbol{\omega}}^{OB}\right)_B = \begin{pmatrix} p^{OB} \\ q^{OB} \\ r^{OB} \end{pmatrix} \tag{2.24}$$

is the angular rate of the aircraft-fixed frame relative to the NED frame. The dynamics (2.24) obviously become singular for $\Theta = \pm\pi/2$, i.e. if the aircraft nose is perpendicularly pointing upwards or downwards. This is also physically meaningful, since no well-defined azimuth can be assigned to this attitude and hence also its first time derivative is not defined. This is a well-known drawback of Euler angles. A possible remedy is the use of quaternions instead.

Quaternions

With the definitions and derivations, given in B.6, let $\underline{\mathbf{q}}_{BO}$ be the quaternion of the rotation of the NED frame to the body-fixed frame. The kinematic quaternion differential equation, according to (B.58) in B.6, is

$$\dot{\underline{\mathbf{q}}}_{BO} = \frac{1}{2} \cdot \bar{\mathbf{E}}_q^T \cdot \left(\vec{\boldsymbol{\omega}}^{OB}\right)_B \tag{2.25}$$

The exact solution of this differential equation preserves unity length property of $\underline{\mathbf{q}}_{BO}$. However due to numerical errors, the unity length will be corrupted as time evolves. Therefore a numerical correction term has to be introduced, which “pulls” the quaternion back to unity length in case of deviations.

$$\dot{\underline{\mathbf{q}}}_{BO} = \frac{1}{2} \cdot \bar{\mathbf{E}}_q^T \cdot \left(\vec{\boldsymbol{\omega}}^{OB}\right)_B - \lambda \cdot \underline{\mathbf{q}}_{BO} \cdot \left(\left\|\underline{\mathbf{q}}_{BO}\right\| - 1\right) \tag{2.26}$$

where λ is some positive constant.

2.2.3 Position Dynamics

The aircraft position is described by use of the WGS-84 model ([EUR98]), where the earth is modeled as a rotationally symmetric ellipsoid, which is oblate in north-south direction. The position on the earth is described by geodetic longitude λ, geodetic latitude ϕ and altitude h, which are also referred to as WGS-84 coordinates. Table 2.3 lists the parameters associated with the ellipsoidal shape of the WGS-84 model. For further information, the reader is referred to [Wen07].

Table 2.3 Parameters WGS-84

description	symbol	value
semi-major axis	a	$6378137.0m$
semi-minor axis	b	$a(1-f)$
flattening	f	$\frac{a-b}{a}=\frac{1}{298.257223563}$
excentricity	e	$\sqrt{f(2-f)}$
normal curvature radius	$N(\phi)$	$\frac{a}{\sqrt{1-e^2\sin^2(\phi)}}$
meridian curvature radius	$M(\phi)$	$\frac{a(1-e^2)}{\sqrt{(1-e^2\sin^2(\phi))^3}}$

The differential equation for the WGS-84 is obtained from the strap down equations using the kinematic velocity w.r.t. to the earth surface, where the components are written in NED frame. These are obtained by a coordinate transformation of the translational states.

$$\left(\vec{\mathbf{V}}_K^R\right)_O^E=\begin{pmatrix}V_N\\V_E\\V_D\end{pmatrix}=\mathbf{M}_{OB}\cdot\left(\vec{\mathbf{V}}_K^R\right)_B^E \tag{2.27}$$

where we abbreviated $V_N=\left(u_K^R\right)_O^E$, $V_E=\left(v_K^R\right)_O^E$, $V_D=\left(w_K^R\right)_O^E$. The position differential equations are ([Wen07])

$$\begin{aligned}\dot{\lambda}&=\frac{V_E}{(N(\phi)+h)\cdot\cos\phi}\\\dot{\phi}&=\frac{V_N}{M(\phi)+h}\\\dot{h}&=-V_D.\end{aligned} \tag{2.28}$$

Alternatively, the position is given in Cartesian ECEF coordinates. In this case, the position differential equations read as

$$\left(\dot{\vec{\mathbf{r}}}^R\right)_E^E=\mathbf{M}_{EO}(\lambda,\phi)\cdot\mathbf{M}_{OB}(\Phi,\Theta,\Psi)\cdot\left(\vec{\mathbf{V}}_K^R\right)_B^E. \tag{2.29}$$

2.2.4 Kinematics

The EOMs derived so far require the definition of forces and moments consisting of aerodynamic, propulsion and gravity. These quantities generally are not specified directly dependent on the chosen rigid body states, but on quantities that can be computed from the latter, which is described in the following.

Kinematic Flow Angles

The kinematic flow angles describe the direction of the kinematic velocity w.r.t. the body-fixed frame (Appendix A). They are related to the Cartesian components by

$$\alpha_K = \operatorname{atan}_2\left(\frac{w_K}{u_K}\right) \;,\; \beta_K = \operatorname{asin}\left(\frac{v_K}{\sqrt{u_K^2 + w_K^2}}\right) \;,\; V_K = \sqrt{u_K^2 + v_K^2 + w_K^2}$$

and the inverse transformation is

$$u_K = V_K \cos\beta_k \cos\alpha_K \;,\; v_K = V_K \sin\beta_K \;,\; w_K = V_K \cos\beta_k \sin\alpha_K \,. \tag{2.30}$$

Thereby for readability, we have abbreviated:

$$u_K = \left(u_K^R\right)_B^E \quad , \quad v_K = \left(v_K^R\right)_B^E \quad , \quad w_K = \left(w_K^R\right)_B^E$$

Aerodynamic Flow Angles and Wind

Wind is modeled by a constant wind velocity defined relative to the earth surface. Wind fields that vary with position and turbulence are not modeled.

$$\left(\vec{\mathbf{V}}_W\right)_O^E = \begin{pmatrix} u_W \\ v_W \\ w_W \end{pmatrix}$$

Here the following abbreviations are used: $u_W = \left(u_W\right)_O^E$, $v_W = \left(v_W\right)_O^E$, $w_W = \left(w_W\right)_O^E$. The aerodynamic velocity is

$$\left(\vec{\mathbf{V}}_A^{\,E}\right)_B^E = \begin{pmatrix} \left(u_A^R\right)_B^E \\ \left(v_A^R\right)_B^E \\ \left(w_A^R\right)_B^E \end{pmatrix} := \left(\vec{\mathbf{V}}_K^{\,E}\right)_B^E - \mathbf{M}_{BO} \cdot \left(\vec{\mathbf{V}}_W\right)_O^E .$$

Analogous to the kinematic case, the aerodynamic velocity is equivalently represented by absolute value V_A, aerodynamic angle of attack α_A and aerodynamic angle of side-slip β_A, where the transformation from and to Cartesian coordinates is computed accordingly.

Angular Rates

The equations of motions only deliver the rotation of the aircraft w.r.t. the ECI frame. In the computations however various angular rates w.r.t. to other reference frames are necessary. The translation and rotation equations of motion require the earth rotation rate as input, which is assumed constant and approximately 360° per 24h, hence

$$\left(\vec{\boldsymbol{\omega}}^{\,IE}\right)_E = \begin{pmatrix} 0 \\ 0 \\ 7.292115 \cdot 10^{-5} \end{pmatrix} \frac{rad}{s}$$

Further, the transport rate $\left(\vec{\boldsymbol{\omega}}^{EO}\right)$ is the rotation that is induced by constraint that the NED axes remain aligned with north-east-down direction, when the aircraft moves along the earth surface. It is needed for the computation of the body-fixed angular rates $\left(\vec{\boldsymbol{\omega}}^{OB}\right)$, which are in turn necessary for the attitude dynamics (2.23)

$$\left(\vec{\boldsymbol{\omega}}^{OB}\right) = \left(\vec{\boldsymbol{\omega}}^{IB}\right) - \left(\vec{\boldsymbol{\omega}}^{IE}\right) - \left(\vec{\boldsymbol{\omega}}^{EO}\right)$$

The transport rate is computed with help of the position differential equations ([Wen07]).

$$\left(\vec{\boldsymbol{\omega}}^{EO}\right)_O = \begin{pmatrix} \dot{\lambda}\cos\phi \\ -\dot{\phi} \\ -\dot{\lambda}\sin\phi \end{pmatrix}$$

Alternative Descriptions

In some cases, it is convenient to describe the aircraft translation by means of absolute velocity, flight path and course angle, which are obtained from the Cartesian velocity components in NED frame (2.27).

$$\gamma_K = \operatorname{asin}\left(\frac{-V_D}{\sqrt{V_N^2 + V_E^2}}\right) \tag{2.31}$$

$$\chi_K = \operatorname{atan}_2\left(\frac{V_E}{V_N}\right) \tag{2.32}$$

where atan_2 uses the information of numerator and denominator separately in order to obtain a unique course angle in the range $-180°$ to $+180°$.

2.3 External Forces and Moments

The equations of motions have been derived, based on the principle of linear and angular momentum. The forces and moments acting onto the vehicle have not been specified but simply were applied to the equations as a general expression. The following sections therefore specify forces and moments further. These consist of an aerodynamic, a propulsive and a gravitational part.

$$\sum\left(\vec{\mathbf{F}}^R\right)_B = \left(\vec{\mathbf{F}}_A^R\right)_B + \left(\vec{\mathbf{F}}_P^R\right)_B + \left(\vec{\mathbf{F}}_G^R\right)_B \tag{2.33}$$

$$\sum\left(\vec{\mathbf{M}}^R\right)_B = \left(\vec{\mathbf{M}}_A^R\right)_B + \left(\vec{\mathbf{M}}_P^R\right)_B + \left(\vec{\mathbf{M}}_G^R\right)_B \tag{2.34}$$

2.3.1 Aerodynamic Configuration

The aerodynamic forces and moments are stored in multidimensional data tables, which were computed numerically from geometry and profile data, based on potential flow, blade element and vortex theory using the institute's own tool ([Ste]).

The aerodynamic forces and moments are computed w.r.t. to an aerodynamic reference point (index A), which in general differs from the aircraft reference point (index R). The force components are given in aerodynamic frame (index A) and moment components are given in body-fixed frame (index B).

$$\left(\vec{\mathbf{F}}_A^A\right)_A = \bar{q} \cdot S \cdot \begin{pmatrix} -C_D \\ C_Q \\ -C_L \end{pmatrix} \tag{2.35}$$

$$\left(\vec{\mathbf{M}}_A^A\right)_B = \bar{q} \cdot S \cdot \begin{pmatrix} b/2 \cdot C_l \\ \bar{c} \cdot C_m \\ b/2 \cdot C_n \end{pmatrix} \tag{2.36}$$

Thereby

- S: wing reference area,
- $\bar{q} = \frac{1}{2}\rho V_A^2$: dynamics pressure, ρ: air density, V_A: aerodyn. velocity
- b: wing span
- $\bar{c}$: mean aerodynamic chord
- C_D, C_Q, C_L: drag-, side force-, lift coefficient
- C_l, C_m, C_n: roll-, pitch-, yaw moment coefficient

The force and moment coefficients are dimensionless and contain all modeling information. The structural dependencies of the respective coefficients are given by the so-called application rule. In [Bau10] the main dependencies of the control surfaces onto each other were identified and minor influence were neglected in order to reduce the amount of data. The dependent variables, used for the aerodynamic table data are given in Table 2.4. It came out, that the main engines have an essential influence on the aerodynamic, as one would intuitively suspect. The back engine however only has a minor influence, which is owed to the comparatively small power and the fact that there is no lifting surface in its slipstream. Moreover, many influences are allowed to be modeled symmetrical w.r.t.to the body-fixed xz-plane. Therefore, the single summands are split up into a left (index l) and a right hand side (index r). The summands incorporate contributions of the aircraft itself, depending on velocity, angle of attack, angle of side-slip, propeller rotation speed and thrust vector inclination angle, which are labeled with an index 0, increments Δ that stem from the control surface deflections and a damping part. In detail the coefficients are given by

$$\begin{aligned} C_i = {} & C_{i,0,l}\left(V_A,\alpha_A,\beta_A,\omega_{M,l},\sigma_l\right)+C_{i,0,r}\left(V_A,\alpha_A,\beta_A,\omega_{M,r},\sigma_r\right) \\ & +\Delta C_{i,\xi,l}\left(\alpha_A,\zeta_l\right)+\Delta C_{i,\xi,r}\left(\alpha_A,\zeta_r\right) \\ & +\Delta C_{i,\eta,l}\left(V_A,\alpha_A,\omega_{M,l},\sigma_l,\eta_l\right)+\Delta C_{i,\eta,r}\left(V_A,\alpha_A,\omega_{M,r},\sigma_r,\eta_r\right) \\ & +\Delta C_{i,\eta_c,l}\left(\alpha_A,\eta_{c,l}\right)+\Delta C_{i,\eta_c,r}\left(\alpha_A,\eta_{c,r}\right) \\ & +\Delta C_{i,\zeta}\left(V_A,\beta_A,\zeta\right) \\ & +\Delta C_{i,\delta,l}\left(V_A,\alpha_A,\omega_{M,l},\delta_l\right)+\Delta C_{i,\delta,r}\left(V_A,\alpha_A,\omega_{M,r},\delta_r\right) \\ & +\Delta C_{i,Damp} \end{aligned} \tag{2.37}$$

where i is replaced by D, Q, L, l, m or n. The damping part enters the equation linearly by means of camping derivatives.

$$\begin{aligned} \Delta C_{i,Damp} = {} & \left[C_{i,p,l}\left(V_A,\alpha_A,\beta_A,\omega_{M,l},\sigma_l\right)+C_{i,p,r}\left(V_A,\alpha_A,\beta_A,\omega_{M,r},\sigma_r\right)\right]\cdot p^* \\ & +\left[C_{i,q,l}\left(V_A,\alpha_A,\beta_A,\omega_{M,l},\sigma_l\right)+C_{i,q,r}\left(V_A,\alpha_A,\beta_A,\omega_{M,r},\sigma_r\right)\right]\cdot q^* \\ & +\left[C_{i,r,l}\left(V_A,\alpha_A,\beta_A,\omega_{M,l},\sigma_l\right)+C_{i,r,r}\left(V_A,\alpha_A,\beta_A,\omega_{M,r},\sigma_r\right)\right]\cdot r^* \end{aligned} \tag{2.38}$$

where

$$p^* = \frac{p_A^{OB}\cdot b}{2\cdot V_A} \quad , \quad q^* = \frac{q_A^{OB}\cdot \bar{c}}{2\cdot V_A} \quad , \quad r^* = \frac{r_A^{OB}\cdot b}{2\cdot V_A} \tag{2.39}$$

are the dimensionless body-fixed angular rates w.r.t. NED frame.

Table 2.4 Dependent Variables in Aerodynamic Dataset

dependent variable	symbol	unit	min	max	number of breakpoints
angle of attack	α_A	°	-27,5	27,5	14
angle of side-slip	β_A	°	-15	15	7
aerodynamic velocity	V_A	m/s	5	20	3
aileron deflection	ξ_l / ξ_r	°	-20	25	10
elevator deflection	η_l / η_r	°	-20	20	9
canard deflection	$\eta_{c,l} / \eta_{c,r}$	°	-25	25	11
rudder deflection	ζ	°	-30	30	13
flap deflection	δ_l / δ_r	°	-20	25	10
main engine thrust vector inclination angle	σ_l / σ_r	°	-25	25	9
main engine rotation speed	$\omega_{M,l} / \omega_{M,r}$	RPM	-13500	13500	7

In order to fit into the dynamic equations (2.22) the aerodynamic forces and moments have to be transformed into the body-fixed frame and the moments have to be shifted to the aircraft reference point.

$$\left(\vec{\mathbf{F}}_A^R\right)_B = \mathbf{M}_{BA} \cdot \left(\vec{\mathbf{F}}_A^A\right)_A \tag{2.40}$$

$$\left(\vec{\mathbf{M}}_A^R\right)_B = \mathbf{M}_{BA} \cdot \left(\vec{\mathbf{M}}_A^A\right)_A + \left(\vec{\mathbf{r}}^{RA}\right)_B \times \mathbf{M}_{BA} \cdot \left(\vec{\mathbf{F}}_A^A\right)_A \tag{2.41}$$

2.3.2 Propulsion System

The description of forces and moments, produced by the propellers require the definition of coordinate frames that are fixed at the respective engine shaft (frame/origin index Pr_i , l: left, r: right, b: back). Thereby, the attitude of the propulsion systems is described relative to the body-fixed frame and is divided into 2 parts. The first part describes the attitude due to the installation of the respective engines w.r.t. the aircraft and is described by a set of Euler angles $\theta_{1,i}$ to $\theta_{3,i}$, beginning with the z-axis when rotating from body-fixed to propulsion frame. Table 2.5 lists the installation angles for all 3 engines.

Table 2.5 Installation Angle Propulsion System

angle	COSY axis	left engine	right engine	back engine
$\theta_{1,i}$	x	3°	-3°	0°
$\theta_{2,i}$	y	0°	0	0°
$\theta_{3,i}$	z	0°	0°	0°

The second part describes the additional rotation, stemming from thrust vectoring and is parameterized by 2 angles, an angle σ about the y-axis and an angle κ about the z-axis, as illustrated in Figure 2.3. The two main engines are only tiltable in the y-axis while the back engine is tiltable in both.

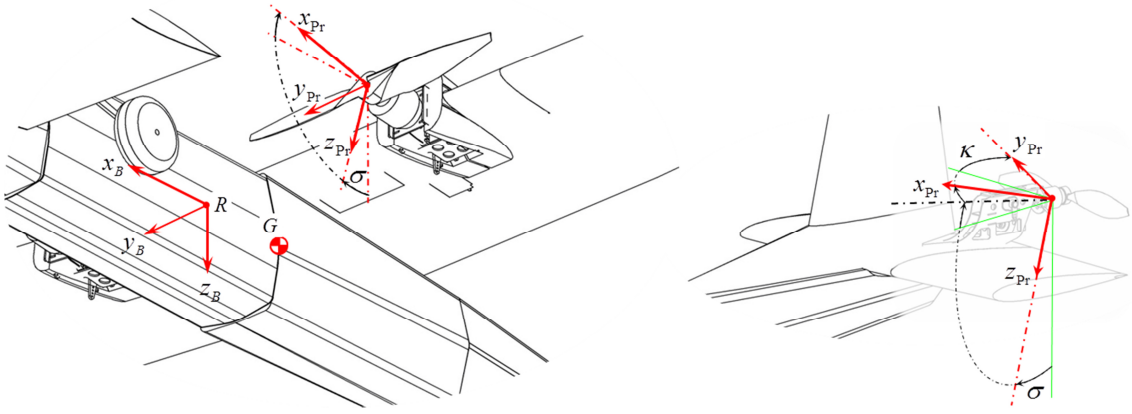


Figure 2.3 Propulsion Coordinate Frames

Table 2.6 lists the thrust vectoring angles including their range.

Table 2.6 Thrust Vectoring Angles

dependent variable	symbol	unit	min	max
main engine thrust vector inclination angle	σ_l / σ_r	°	-25	25
back engine thrust vector inclination angle	σ_b	°	-15	15
back engine thrust vector azimuth angle	κ_b	°	-25	25

The transformation matrix from body-fixed to propulsion system is given by

$$\mathbf{M}_{\mathrm{Pr}_i B} = \mathbf{M}^{(2)}_{\mathrm{Pr}_i B} \cdot \mathbf{M}^{(1)}_{\mathrm{Pr}_i B} \tag{2.42}$$

where

$$\mathbf{M}^{(2)}_{\mathrm{Pr}_i B} = \begin{bmatrix} \cos\kappa_i \cos\sigma_i & \sin\kappa_i & -\cos\kappa_i \sin\sigma_i \\ -\cos\sigma_i \sin\kappa_i & \cos\kappa_i & \sin\kappa_i \sin\sigma_i \\ \sin\sigma_i & 0 & \cos\sigma_i \end{bmatrix}$$

$$\mathbf{M}^{(1)}_{\mathrm{Pr}_i B} = \begin{bmatrix} \cos\theta_{2,i}\cos\theta_{3,i} & \cos\theta_{1,i}\sin\theta_{3,i} + \cos\theta_{3,i}\sin\theta_{1,i}\sin\theta_{2,i} & \sin\theta_{1,i}\sin\theta_{3,i} - \cos\theta_{1,i}\cos\theta_{3,i}\sin\theta_{2,i} \\ -\cos\theta_{2,i}\sin\theta_{3,i} & \cos\theta_{1,i}\cos\theta_{3,i} - \sin\theta_{1,i}\sin\theta_{2,i}\sin\theta_{3,i} & \cos\theta_{3,i}\sin\theta_{1,i} + \cos\theta_{1,i}\sin\theta_{2,i}\sin\theta_{3,i} \\ \sin\theta_{2,i} & -\cos\theta_{2,i}\sin\theta_{1,i} & \cos\theta_{1,i}\cos\theta_{2,i} \end{bmatrix}$$

and i is replaced by l (left), r (right) or b (back). The forces and moments, produced by the propellers, are stored in multidimensional look-up tables and depend on

- total angle of attack $\alpha_{T,i}$
- total angle of side-slip $\beta_{T,i}$
- aerodynamic velocity V_A
- propeller rotation speed $\omega_{M,i}$

$$\left(\vec{\mathbf{F}}^{\mathrm{Pr}_i}_{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} = \begin{pmatrix} \left(X^{\mathrm{Pr}_i}_{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} \\ \left(Y^{\mathrm{Pr}_i}_{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} \\ \left(Z^{\mathrm{Pr}_i}_{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} \end{pmatrix} = \mathbf{f}\left(\alpha_{T,i}, \beta_{T,i}, V_A, \omega_{M,i}\right) \tag{2.43}$$

$$\left(\vec{\mathbf{M}}^{\mathrm{Pr}_i}_{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} = \begin{pmatrix} \left(L^{\mathrm{Pr}_i}_{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} \\ \left(M^{\mathrm{Pr}_i}_{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} \\ \left(N^{\mathrm{Pr}_i}_{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} \end{pmatrix} = \mathbf{f}\left(\alpha_{T,i}, \beta_{T,i}, V_A, \omega_{M,i}\right) \tag{2.44}$$

Total angle of attack and angle of side-slip are obtained from the direction of the aerodynamic velocity w.r.t. the respective propulsion frame that is in fact the first column of $\mathbf{M}_{\mathrm{Pr}_i A}$, the transformation matrix from aerodynamic to the respective propulsive frame.

$$\mathbf{m}_{\mathrm{Pr}_i A}^{(1)} = \begin{pmatrix} \cos\alpha_{T,i} \cdot \cos\beta_{T,i} \\ \sin\beta_{T,i} \\ \sin\alpha_{T,i} \cdot \cos\beta_{T,i} \end{pmatrix}$$

With this insight, the total aerodynamic angles w.r.t. to the propulsion system are

$$\alpha_{T,i} = \mathrm{atan}_2\left(\frac{\left[\mathbf{m}_{\mathrm{Pr}_i A}^{(1)}\right]_3}{\left[\mathbf{m}_{\mathrm{Pr}_i A}^{(1)}\right]_1} \right) \tag{2.45}$$

$$\beta_{T,i} = \mathrm{asin}\left(\left[\mathbf{m}_{\mathrm{Pr}_i A}^{(1)}\right]_2\right) \tag{2.46}$$

where $[\,\cdot\,]_j$ denotes the j^{th} vector element and the atan_2 function computes a unique angle in the range *-180°* to *+180°* due to the separate evaluation of numerator and denominator. Accounting for angles that vanish according to Table 2.5 we obtain

$$\alpha_{T,l/r} = \mathrm{atan}_2\left(\frac{\cos\alpha_A \cos\beta_A \sin\sigma_{l/r} - \cos\sigma_{l/r}\left(\sin\beta_A \sin\theta_{1,l/r} - \cos\beta_A \sin\alpha_A \cos\theta_{1,l/r}\right)}{\sin\sigma_{l/r}\left(\sin\beta_A \sin\theta_{1,l/r} - \cos\beta_A \sin\alpha_A \cos\theta_{1,l/r}\right) + \cos\alpha_A \cos\beta_A \cos\sigma_{l/r}} \right) \tag{2.47}$$

$$\beta_{T,l/r} = \mathrm{asin}\left(\sin\beta_A \cos\theta_1 + \cos\beta_A \sin\alpha_A \sin\theta_1\right) \tag{2.48}$$

$$\alpha_{T,b} = \mathrm{atan}_2\left(\frac{\sin(\alpha_A + \sigma_b)\cos\beta_A}{\sin\beta_A \sin\kappa_b + \cos\beta_A \cos\kappa_b \cos(\alpha_A + \sigma_b)} \right) \tag{2.49}$$

$$\beta_{T,b} = \mathrm{asin}\left(\cos\kappa_b \sin\beta_A - \cos\beta_A \sin\kappa_b \cos(\alpha_A + \sigma_b)\right). \tag{2.50}$$

In order to be applied to the EOMs, the respective reference points have to be shifted to R and forces and moments have to be transformed to the body-fixed frame.

$$\left(\vec{\mathbf{F}}_P^R\right)_B = \mathbf{M}_{B\,\mathrm{Pr}_l} \cdot \left(\vec{\mathbf{F}}_{\mathrm{Pr}_l}^{\mathrm{Pr}_l}\right)_{\mathrm{Pr}_l} + \mathbf{M}_{B\,\mathrm{Pr}_r} \cdot \left(\vec{\mathbf{F}}_{\mathrm{Pr}_r}^{\mathrm{Pr}_r}\right)_{\mathrm{Pr}_r} + \mathbf{M}_{B\,\mathrm{Pr}_b} \cdot \left(\vec{\mathbf{F}}_{\mathrm{Pr}_b}^{\mathrm{Pr}_b}\right)_{\mathrm{Pr}_b} \tag{2.51}$$

$$\left(\vec{\mathbf{M}}_P^R\right)_B = \mathbf{M}_{B\mathrm{Pr}_l} \cdot \left(\vec{\mathbf{M}}_{\mathrm{Pr}_l}^{\mathrm{Pr}_l}\right)_{\mathrm{Pr}_l} + \mathbf{M}_{B\mathrm{Pr}_r} \cdot \left(\vec{\mathbf{M}}_{\mathrm{Pr}_r}^{\mathrm{Pr}_r}\right)_{\mathrm{Pr}_r} + \mathbf{M}_{B\mathrm{Pr}_b} \cdot \left(\vec{\mathbf{M}}_{B\mathrm{Pr}_b}^{B\mathrm{Pr}_b}\right)_{B\mathrm{Pr}_b}$$
$$+ \left(\vec{\mathbf{r}}^{R\mathrm{Pr}_l}\right)_B \times \left[\mathbf{M}_{B\mathrm{Pr}_l} \cdot \left(\vec{\mathbf{F}}_{\mathrm{Pr}_l}^{\mathrm{Pr}_l}\right)_{\mathrm{Pr}_l}\right] + \left(\vec{\mathbf{r}}^{R\mathrm{Pr}_r}\right)_B \times \left[\mathbf{M}_{B\mathrm{Pr}_r} \cdot \left(\vec{\mathbf{F}}_{\mathrm{Pr}_r}^{\mathrm{Pr}_r}\right)_{\mathrm{Pr}_r}\right] + \left(\vec{\mathbf{r}}^{R\mathrm{Pr}_b}\right)_B \times \left[\mathbf{M}_{BB\mathrm{Pr}_b} \cdot \left(\vec{\mathbf{F}}_{B\mathrm{Pr}_b}^{B\mathrm{Pr}_b}\right)_{B\mathrm{Pr}_b}\right] \tag{2.52}$$

BLDC Motor

The FSD Extreme Star is powered by three brushless DC electric motors (BLDC) with a nominal voltage of *12V*. Nevertheless BLDCs have similar characteristics as brushed DC electric permanent magnet motors ([Röß12]), described by the following well-known equations (e.g. [Dör08]).

The whole motor circuit is modeled as shown in Figure 2.4. The outer circuit consists of a motor controller and cables, modeled as ohmic resistances R_{Ctrl}, R_{Cable} and the motor. The motor itself consists of the armature circuit with resistance R_A, inductivity L_A

and an induced electromagnetic voltage U_{EM}, induced by the exciting magnetic field and the rotation speed of the motor ω_M.

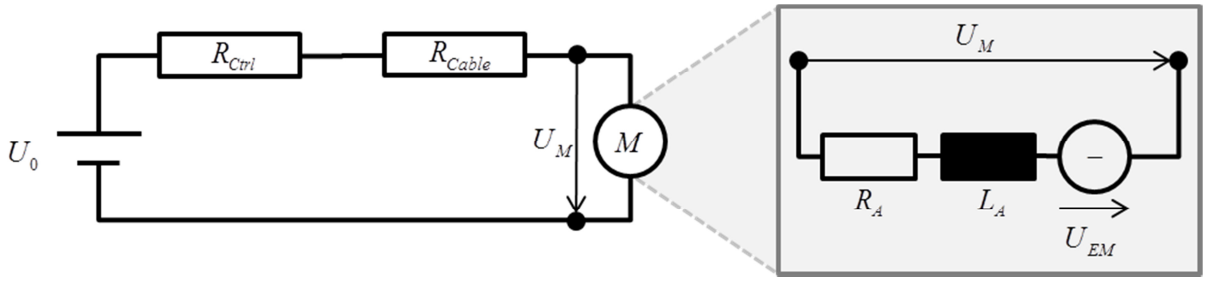


Figure 2.4 Modeling of DC Motor

Using Kirchhoff's mesh rule, one obtains for the outer motor circuit

$$U_0 = (R_{Ctrl} + R_{Cable}) \cdot I + U_M \tag{2.53}$$

and similarly for the armature circuit

$$U_M = R_A \cdot I + U_{EM} + L_A \cdot \dot{I} \tag{2.54}$$

where $U_{EM}=K_E \cdot \omega_M$ with the machine specific constant K_E. The moment M_{EM} , induced to the shaft by the magnetic field is further given by $M_{EM}=K_E I$ and the loss of moment M_L due to friction and other dissipative effects is assumed to be proportional to the rotation speed with a factor K_L: $M_L=K_L \cdot \omega$. The total moment onto the shaft is $M_{tot}=M_S-M_L-M_{Load}$, where M_{load} is the load moment, in our case the moment, induced by the propeller drag. With the moment of inertia of shaft and propeller about the shaft axis J_S and J_{Pr}, the equation of motion for the rotation speed becomes

$$(J_S + J_{\mathrm{Pr}})\dot{\omega}_M = K_E I - K_L \omega_M - M_{Load} \tag{2.55}$$

Combining equations (2.53) to (2.56) and writing in LTI form yields

$$\begin{pmatrix} \dot{I} \\ \dot{\omega}_M \end{pmatrix} = \begin{bmatrix} \frac{R_T}{L_A} & \frac{K_E}{L_A} \\ \frac{K_E}{J_T} & -\frac{K_L}{J_T} \end{bmatrix} \cdot \begin{pmatrix} I \\ \omega_M \end{pmatrix} + \begin{bmatrix} \frac{1}{L_A} & 0 \\ 0 & \frac{1}{J_T} \end{bmatrix} \cdot \begin{pmatrix} U_0 \\ M_{Load} \end{pmatrix} \tag{2.56}$$

where $J_T = J_S + J_{\mathrm{Pr}}$ is the total moment of inertia and $R_T = R_{Ctrl} + R_{Cable} + R_A$ is the total resistance.

Steady State Conditions

For flight performance analysis, steady state conditions of the motor system are of interest. Therefore, the time derivatives in (2.56) are set to zero, which yields

$$\omega_s = \frac{K_E}{K_E^2 + R_T K_L} \cdot U_0 - \frac{R_T}{K_E^2 + R_T K_L} \cdot M_{Load} \tag{2.57}$$

$$I_s = \frac{K_L}{K_E^2 + R_T K_L} \cdot U_0 + \frac{K_E}{K_E^2 + R_T K_L} \cdot M_{Load} \tag{2.58}$$

Two interesting quantities, which are usually found in data sheets, can be extracted from (2.57) and (2.58) by setting the load moment to zero:

1. Idle rotation speed: $\omega_{Idle} = K_E \left(K_E^2 + R_T K_L\right)^{-1} U_0$
2. No-load current: $I_{Idle} = K_L \left(K_E^2 + R_T K_L\right)^{-1} U_0$

Effective Power

The effective power of the motor is defined as $P_{Mot}=M_{Load}\omega_s$. Solving (2.57) for M_{load} and inserting into the power definition yields:

$$P_{Mot}(\omega_s) = \frac{K_E}{R_T} \cdot U_0 \cdot \omega_s - \frac{K_E^2 + R_T \cdot K_L}{R_T} \omega_s^2 \tag{2.59}$$

Motor Controller

The BLDC motors are actuated by motor controllers, which receive pulse width modulated signals as commands. In the context of the motor modeling, described recently, the pulse width is translated into an input voltage U_0 to the respective motor. Commonly actuator commands are modeled as normed values, in this case ranging from 0 to 1 which are linearly assigned to the input voltage, where 0 represents $U_0=0V$ and 1 represents $U_0=12V$.

Motor Propeller Interaction

A change in the input voltage effects a change in the shaft moment and which in turn effects a rotational acceleration until the steady state conditions (2.57) are met again. Thereby, the load moment is the x-component of the propeller moment in (2.44). Thus, the resulting propeller rotational speed is obtained by intersection of propeller characteristic (2.44) and motor characteristic (2.57)

$$\omega_{s,i} = \frac{K_E}{K_E^2 + R_T K_L} \cdot U_0 - \frac{R_T}{K_E^2 + R_T K_L} \cdot \left(L^{\mathrm{Pr}_i}_{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} \tag{2.60}$$

2.3.3 Gravity

The third group of forces, acting onto the aircraft is the gravity. Since the considered aircraft is rather small, operating locally and in low altitudes, the gravity is modeled as a constant force. Decreasing of gravity with altitude and variations dependent on global position, as e.g. provided by the Earth Gravitational Model 2008 (EGM2008) published

by the Geospatial Intelligence Agency, are not taken into account. The gravitational force is assumed to act perpendicular to the local tangential plane of the WGS-84 reference ellipsoid, i.e. it is aligned with the z-axis of the NED frame. As it is well-known, the gravitational force, which is actually a volume force, acting onto every mass element of a body, can be lumped into the center of gravity, thus:

$$\left(\vec{\mathbf{F}}_G^G\right)_O = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \cdot m \cdot g$$

where $g = 9.80665\ {}^{m}/_{s^2}$ is the standard gravity and m is the total aircraft mass. Transformed to the body-fixed frame, the gravity is

$$\left(\vec{\mathbf{F}}_G^G\right)_B = \begin{pmatrix} -\sin\Theta \\ \cos\Theta \cdot \sin\Phi \\ \cos\Theta \cdot \cos\Phi \end{pmatrix} \cdot m \cdot g \tag{2.61}$$

Since the equations of motion are formulated w.r.t. to R, which is generally different from the c.g., the gravitational force also induces a moment.

$$\left(\vec{\mathbf{M}}_G^R\right)_B = \left(\vec{\mathbf{r}}^{RG}\right)_B \times \left(\vec{\mathbf{F}}_G^G\right)_B \tag{2.62}$$

2.4 Subsystems

The rigid body equations of motions derived so far represent the main part of the aircraft dynamic. However also further dynamic elements, such as actuators and sensors that are located within the control loop effect the closed-loop behavior. These subsystem dynamics are relatively fast compared to the rigid body dynamics and for low gain control systems it should be sufficient to neglect these fast dynamics. However, if it is an ambition to fully exploit the aircraft capabilities, leading typically to high gain control systems, the actuator and sensor dynamics have to be taken into account very well.

2.4.1 Actuators

All control surfaces as well as the thrust vectoring angles are actuated by off-the-shelf electrical servos as they are typically utilized for model airplanes. The servos are controlled by pulse-width-modulated signals and are operated at a nominal voltage between $4V$ and $6V$.

The servos consist of an electric motor and a gear box, which is loaded by the actuator hinge moments. In section 2.3.2 the electric motor is modeled as a second order system. Usually the rotation dynamic (2.55) is dominating while the current dynamic, given by (2.54) is comparatively fast and hence it is neglected here. Thus, we have a first order dynamics from input voltage to rotation speed. Since the quantity of interest

is the actuator position, the preceding discussion motivates a linear second order dynamics for the actuators given by the transfer function.

$$G_A(s) = \frac{U(s)}{U_C(s)} = \frac{\omega_0^2}{s^2 + 2 \cdot \zeta \cdot \omega_0 \cdot s + \omega_0^2} \tag{2.63}$$

where ω_0 is the natural frequency and ζ is the relative damping. Figure 2.5 shows the actuator model block diagram.

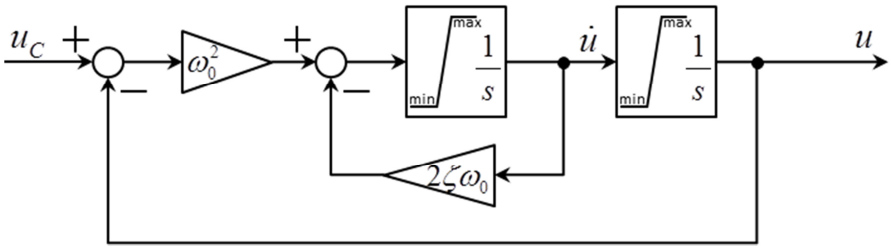


Figure 2.5 Actuator Model

While the output of the right integrator is the actuator position itself, the output of the left integrator is the actuator velocity. The following parameters are chosen for the actuators.

$$\omega_0 = 30 \frac{rad}{s} \quad , \quad \zeta = 0.7 \tag{2.64}$$

which could be considered as a worst case as the real natural frequency is expected to be higher. However, these parameters form a reasonable basis for control design. Moreover, the outputs of the integrators of the actuator model are limited. The right integrator limits the actuator position, which is necessary due to the subcritical damping, since even if the actuator command remains within its limits a transient overshoot might occur due oscillations. The limit of the left integrator represents the maximum actuator velocity, which complies with reality since the rotation speed of the electric motor of the servo is subjected to constraints as well. Rate limit of the actuators also have a distinctive effect on the closed loop dynamic, particularly in case of high amplitudes. In these cases, the rate limit could lead to instability at high amplitude maneuvers, while small amplitude maneuvers can be stabilized very well. It has to be mentioned that such a behavior is a nonlinear and the transfer function, except for small amplitudes, does not fully represent the actuator model depicted in Figure 2.5. The actuator parameters, position and rate limits have been determined in a series of measurements in [Bau10] and the main results are presented in Table 2.7.

Table 2.7 Actuator Data

		ω_0	ζ	symbols			rate Limit		position limit	
				u_C	$\dot{u}$	u	$\dot{u}_{\max}$	$\dot{u}_{\min}$	$u_{\max}$	$u_{\min}$
							$\frac{\deg}{s}$		deg	
canard	left	$30^{rad}/_{s}$	0.7	$\eta_{c,l,C}$	$\dot{\eta}_{c,l}$	$\eta_{c,l}$	545	-545	25	-24
	right			$\eta_{c,r,C}$	$\dot{\eta}_{c,r}$	$\eta_{c,r}$			19	-30
flaps	left	$30^{rad}/_{s}$	0.7	$\delta_{l,C}$	$\dot{\delta}_l$	δ_l	500	-500	25	-22
	right			$\delta_{r,C}$	$\dot{\delta}_r$	δ_r			28	-20
elevator	left	$30^{rad}/_{s}$	0.7	$\eta_{l,C}$	$\dot{\eta}_l$	η_l	545	-545	20	-19
	right			$\eta_{r,C}$	$\dot{\eta}_r$	η_r			15	-19
rudder		$30^{rad}/_{s}$	0.7	ζ_C	$\dot{\zeta}$	ζ	545	-545	35	-31
aileron	left	$30^{rad}/_{s}$	0.7	$\xi_{l,C}$	$\dot{\xi}_l$	ξ_l	500	-500	23	-19
	right			$\xi_{r,C}$	$\dot{\xi}_r$	ξ_r			30	-20
main motor	left	$30^{rad}/_{s}$	0.7	$\sigma_{l,C}$	$\dot{\sigma}_l$	σ_l	222	-222	30	-26
	right			$\sigma_{r,C}$	$\dot{\sigma}_r$	σ_r			25	-26
back motor	σ	$30^{rad}/_{s}$	0.7	$\sigma_{b,C}$	$\dot{\sigma}_b$	σ_b	545	-545	15	-15
	κ			$\kappa_{b,C}$	$\dot{\kappa}_b$	κ_b	545	545	25	-25

2.4.2 Sensors

The FSD Extreme Star is equipped with an Xsens MTi-G unit, which is an integrated GPS and inertial measurement unit ([Xse09]). It comprises MEMS inertial sensors, measuring specific external forces and angular rates and a miniature GPS receiver as well as additional sensors such as 3D magnetometer and static pressure sensor. The Xsens offers a complete navigation solution, computed by an internal Kalman Filter. The navigation solution consists of position in WGS-84 coordinates, kinematic velocity in NED components, aircraft attitude as a set of Euler angles, angular rate, inertial acceleration as well as static pressure. The attitude is also alternatively specified as unit quaternion in order to avoid the known Euler angle singularity.

The specific forces and rotational rate as well as static pressure are delivered as calibrated raw data, that are partially compensated for various sensor inherent errors such as temperature drift, scaling errors and bias, however the position and velocity data as well as attitude are processed data from the Kalman Filter. Although different sensors run at different sample rates – e.g. the GPS receiver delivers position and velocity information at a rate of $4Hz$ – the data are processed internally such that all sensor information are delivered at a higher update that can be chosen by the user.

Generally, the sensor data are subjected to various inaccuracies even though the Kalman Filter tries to minimize the stochastic errors (noise) as much as possible. Subsequently, dynamics equations used for flight control design are formulated w.r.t. the c.g. In reality, the placement of the sensor is subjected to constraints that avoid the exact placement of the sensor unit at the c.g. such as EMC issues, visibility of satellites for the GPS antenna low-vibration mounting etc... Moreover, the c.g. position is not

known exactly. In the subsequent equations, the vector from aircraft c.g. to the sensor point is employed for this purpose.

$$\left(\vec{\mathbf{r}}^{GS}\right)_B \tag{2.65}$$

Another source of inaccuracy is the installation of the sensors; thus the sensor unit does not measure the diverse quantities in body-fixed components, but in a deviated sensor system, which will be denoted with an index S for the remainder. The transformation matrix is parameterized by a set of Euler angles $\theta_x, \theta_y, \theta_z$ with sequence z-y-x when rotating from body-fixed to sensor system.

$$\mathbf{M}_{SB} = \begin{bmatrix} \cos\theta_y \cos\theta_z & \cos\theta_y \sin\theta_z & -\sin\theta_y \\ \sin\theta_x \sin\theta_y \cos\theta_z - \cos\theta_x \sin\theta_z & \sin\theta_x \sin\theta_y \sin\theta_z + \cos\theta_x \cos\theta_z & \sin\theta_x \cos\theta_y \\ \cos\theta_x \sin\theta_y \cos\theta_z + \sin\theta_x \sin\theta_z & \cos\theta_x \sin\theta_y \sin\theta_z - \sin\theta_x \cos\theta_z & \cos\theta_x \cos\theta_y \end{bmatrix} \tag{2.66}$$

Additionally the raw data, delivered by the MEMS sensors, are corrupted by scaling and bias errors. As the various sensor errors have an influence on the performance of the control systems, the significant deviations, which are incorporated within the simulation model, are described separately for each sensor group.

Accelerometer

The accelerometer measures all accelerations induced by surface forces. This implies that, for aircraft applications, only aerodynamic and propulsion forces are measured but not gravitational forces since it is a volume force. This fact as well as the positioning of the accelerometer at some aircraft-fixed point deviated from the c.g. is accounted for in the sensor model.

At first, the acceleration at the sensor point is computed by means of a lever arm correction term. From the simulation model, we know the total forces acting onto the aircraft.

$$\sum\left(\vec{\mathbf{F}}^{G}\right)_B$$

Applying successively Euler's differentiation rule, we get the acceleration at the sensor point.

$$\begin{aligned} \left(\dot{\vec{\mathbf{V}}}_K^S\right)_B^{II} &= \left(\vec{\mathbf{V}}_K^G\right)^{II} + \left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)_B^B \times \left(\vec{\mathbf{r}}^{GS}\right)_B + \left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \times \left[\left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \times \left(\vec{\mathbf{r}}^{GS}\right)_B\right] \\ &= \frac{\sum\left(\vec{\mathbf{F}}^{G}\right)_B}{m} + \left(\dot{\vec{\boldsymbol{\omega}}}^{IB}\right)_B^B \times \left(\vec{\mathbf{r}}^{GS}\right)_B + \left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \times \left[\left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \times \left(\vec{\mathbf{r}}^{GS}\right)_B\right] \end{aligned} \tag{2.67}$$

with the vector from the aircraft c.g. to the sensor point $\left(\vec{\mathbf{r}}^{GS}\right)_B$ and the total aircraft mass m. This is the total acceleration at S, which consists of a part stemming from surface forces and another part from gravity, which is not measured by the sensor. Hence the acceleration, measured by the sensor, is

$$\left(\vec{\mathbf{f}}^{S}\right)_{S} = \mathbf{M}_{SB}\left[\left(\vec{\mathbf{V}}_{K}^{S}\right)_{B}^{II} - \left(\vec{\mathbf{g}}\right)_{B}\right] \tag{2.68}$$

where

$$\left(\vec{\mathbf{g}}\right)_{B} = g_{0} \cdot \begin{pmatrix} -\sin\Theta \\ \cos\Theta\sin\Phi \\ \cos\Theta\cos\Phi \end{pmatrix} \tag{2.69}$$

denotes the gravitational acceleration in body-fixed components with standard gravity constant $g_0 = 9{,}81665\,{}^{m}/_{s^2}$. So far, the clean measurements without sensor inherent corruptions have been computed. In order to simulate a realistic sensor behavior, the clean measurements are charged with scaling and orientation errors, bias as well as noise disturbance. The raw sensor measurements are

$$\left(\vec{\mathbf{f}}_{MEAS}^{S}\right)_{S} = \left(\mathbf{I} + \mathbf{M}_{sc} + \mathbf{M}_{cc}\right) \cdot \left(\vec{\mathbf{f}}^{S}\right)_{S} + \vec{\mathbf{b}}_{f} + \vec{\mathbf{n}}_{f} \tag{2.70}$$

where

$$\mathbf{M}_{sc} = \begin{bmatrix} x_{sc} & 0 & 0 \\ 0 & y_{sc} & 0 \\ 0 & 0 & z_{sc} \end{bmatrix} \tag{2.71}$$

represents the scaling error and

$$\mathbf{M}_{cc} = \begin{bmatrix} 0 & \delta_{xy} & \delta_{xz} \\ \delta_{yx} & 0 & \delta_{yz} \\ \delta_{zx} & \delta_{zy} & 0 \end{bmatrix} \tag{2.72}$$

represents the cross coupling of the axes due to non-orthogonality of the sensor axes, $\vec{\mathbf{b}}_f$ is some 3-dimensional bias vector and $\vec{\mathbf{n}}_f$ is some Gaussian distributed sequence with standard deviation σ_x, σ_y, σ_z, representing sensor noise. Table 2.8 presents the data, used for the acceleration sensor model, which are compliant with the data sheet of the sensor unit. The noise standard deviation is dependent on the sample time of the sensor T_s. This is because noise is assumed to be a continuous time white noise with a constant power spectral density in the data sheet, which is approximated as discrete time Gaussian distributed zero mean sequence. The standard deviation of the discrete sequence decreases with the square root of the sample time, roughly spoken, due to averaging effects. For further information on this topic, refer to literature about statistic processes, e.g. [Ben10].

Table 2.8 Sensor Errors for Accelerometer

		x-axis : $(f_x^S)_S$	y-axis : $(f_y^S)_S$	z-axis : $(f_z^S)_S$
scaling i_{sc}		0.0003	0.0003	0.0003
cross coupling				
	δ_{ix}	-	$0.1\frac{\pi}{180}$	$0.1\frac{\pi}{180}$
	δ_{iy}	$0.1\frac{\pi}{180}$	-	$0.1\frac{\pi}{180}$
	δ_{iz}	$0.1\frac{\pi}{180}$	$0.1\frac{\pi}{180}$	-
noise std. dev. $\sigma_i[m/s^2]$		$\frac{0.02}{\sqrt{T_s \cdot s^{-1}}}$	$\frac{0.02}{\sqrt{T_s \cdot s^{-1}}}$	$\frac{0.02}{\sqrt{T_s \cdot s^{-1}}}$

Gyroscope

Since the aircraft is assumed to be a rigid body, the rotation rate is invariant with changes in reference point. Therefore, only the deviated orientation of the sensors axes relative to the aircraft-fixed frame has to be taken into account.

$$\left(\vec{\boldsymbol{\omega}}^{IS}\right)_S = \mathbf{M}_{SB} \cdot \left(\vec{\boldsymbol{\omega}}^{IB}\right)_B \tag{2.73}$$

Additional scaling, cross coupling, bias and noise errors are added according to

$$\left(\vec{\boldsymbol{\omega}}^{IS}_{MEAS}\right)_S = \left(\mathbf{I} + \mathbf{M}_{sc} + \mathbf{M}_{cc}\right) \cdot \left(\vec{\boldsymbol{\omega}}^{IB}\right)_S + \vec{\mathbf{b}}_{\omega} + \vec{\mathbf{n}}_{\omega} \tag{2.74}$$

where $\mathbf{M}_{sc}$ and $\mathbf{M}_{cc}$ are defined in (2.71), (2.72). Table 2.9 lists the specific values for the gyroscope inherent errors.

Table 2.9 Sensor Errors for Gyroscope

		x-axis $(p^{IS})_S$	y-axis : $(q^{IS})_S$	z-axis : $(r^{IS})_S$
scaling i_{sc}		0	0	0
cross coupling				
	δ_{ix}	-	$0.1\frac{\pi}{180}$	$0.1\frac{\pi}{180}$
	δ_{iy}	$0.1\frac{\pi}{180}$	-	$0.1\frac{\pi}{180}$
	δ_{iz}	$0.1\frac{\pi}{180}$	$0.1\frac{\pi}{180}$	-
noise std. dev. $\sigma_i[^\circ/s]$		$\frac{0.05}{\sqrt{T_s \cdot s^{-1}}}$	$\frac{0.05}{\sqrt{T_s \cdot s^{-1}}}$	$\frac{0.05}{\sqrt{T_s \cdot s^{-1}}}$

First Order Gauss Markov Process

Statistical errors with limited bandwidth, occurring in the respective sensor data, are suitably modeled by a first order Gauss Markov process. Such a band limited statistical process can e.g. be used as approximation for the output of the Kalman filter estimation error, since the Kalman filter error equation is in fact a strictly proper (i.e. without direct feed through) dynamic system which is excited by white noise and, as is

well-known, such systems transmit signals only up to a limited bandwidth. A 1st order Gauss Markov process is defined by

$$\dot{x} = 2\pi \cdot f_0 \cdot (-x + u) \tag{2.75}$$

where f_0 is the bandwidth in Hz and u is a continuous time white noise with a power spectral density (PSD) of Q_u. I.e. its autocorrelation is

$$R_u(t) = \lim_{T\to\infty} \frac{1}{2T} \cdot \int_{-T}^{T} u(t+\tau) \cdot u(\tau) \cdot d\tau = Q_u \cdot \delta(\tau) \tag{2.76}$$

and its power spectrum is

$$S_u(f) = \int_{-\infty}^{\infty} R_u(\tau) \cdot e^{-2\pi f_0 \tau} d\tau = Q_u \tag{2.77}$$

where $\delta(\tau)$ denotes Dirac's distribution. In the data sheets, statistical sensor errors are typically specified as standard deviation of a stochastic process, which is in our case the output of the Gauss Markov process. For implementation, the question to be clarified is, how to choose the PSD of u such that the output of the Gauss Markov process has the desired standard deviation. Due to [Ben10], the power spectral density of the system output x is related to the PSD of the input by

$$S_x(f) = |G(f)|^2 \cdot S_u(f) \tag{2.78}$$

where $G(f)$ is the transfer function of (2.75). Therefore, we get

$$S_x(f) = \frac{Q_u}{\left(f/f_0\right)^2 + 1} \tag{2.79}$$

Due to Parseval's theorem, the time averaged standard deviation of x amounts to the following expression.

$$\begin{aligned} \sigma_x^2 &= \lim_{T\to\infty} \frac{1}{2T} \cdot \int_{-T}^{T} x^2(\tau) d\tau = \int_{-\infty}^{\infty} S_x(f) df \\ &= \int_{-\infty}^{\infty} \frac{Q_u}{\left(f/f_0\right)^2 + 1} df = Q_u \cdot f_0 \cdot \left[\arctan\left(\frac{f}{f_0}\right) \right]_{-\infty}^{\infty} \\ &= Q_u \cdot f_0 \cdot \pi \end{aligned} \tag{2.80}$$

Since the simulation is in discrete time, it is necessary to have a discrete time approximation of the continuous time white noise signal u, which is, according to [Sim06], given by a zero mean Gaussian distributed sequence with a standard deviation of

$$\sigma_u^2 = \frac{Q_u}{T_s} \tag{2.81}$$

where T_s denotes the simulation sample time. Note that the white noise approximation is valid if the bandwidth of the Gauss Markov process is comparatively low to the

Nyquist frequency $f_{Ny}=1/(2T_s)$. Applying the white noise approximation to (2.80) and solving for σ_u yields.

$$\sigma_u = \frac{\sigma_x}{\sqrt{T_s \cdot f_0 \cdot \pi}} \tag{2.82}$$

Thus, we have obtained a relationship between the standard deviation of system input and output. The discrete implementation is depicted in Figure 2.6. Input and output signal are sequences, sampled with T_s and the continuous time integrator is replaced by the forward Euler discretization.

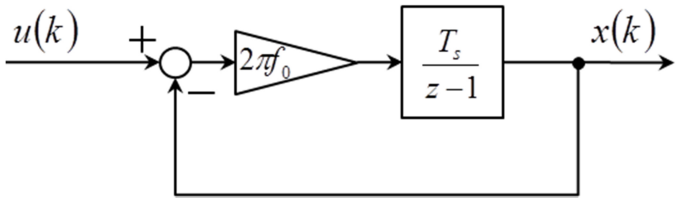


Figure 2.6 Discrete Approximation of a first Order Gauss Markov Process

Attitude

The aircraft attitude cannot be measured directly with gyroscope and accelerometer. However, the attitude can be computed by integration of the body-fixed rotation rates and correct the integration errors by a reference pitch and bank attitude, computed from the vertical gravity force, which is measured by the accelerometer and the heading, measured by the magnetometer. Such algorithms are denoted as attitude heading reference systems (AHRS) and are typically used in aviation applications. The Xsens unit however fuses all available sensors into a Kalman filter, which amongst others also estimates the attitude. As no detailed data about the estimation error are provided by the manufacturer, these quantities will be modeled in a rather phenomenological manner. For the sensor unit, we assume a deviated orientation relative to the body-fixed frame, given by the transformation matrix (2.66).

$$\mathbf{M}_{SO} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} = \mathbf{M}_{SB} \cdot \mathbf{M}_{BO} \tag{2.83}$$

Euler angles of the deviated sensor system are obtained from $\mathbf{M}_{SO}$ by consideration of the Euler angle parameterization of $\mathbf{M}_{BO}$.

$$\begin{aligned} \Psi_S &= \arctan_2\left(\frac{m_{12}}{m_{11}}\right) \\ \Theta_S &= \arcsin\left(-m_{13}\right) \\ \Phi_S &= \arctan_2\left(\frac{m_{23}}{m_{33}}\right) \end{aligned} \tag{2.84}$$

Since the attitude is the output of a Kalman filter, a dynamic bias is added to the clean measurements in form of a Gauss Markov process such that raw Euler measurements are

$$\begin{aligned} \Psi_{MEAS} &= \Psi + m_{\Psi} \\ \Theta_{MEAS} &= \Theta + m_{\Theta} \\ \Phi_{MEAS} &= \Phi + m_{\Phi} \end{aligned} \tag{2.85}$$

where m_{Ψ}, m_{Θ}, m_{Φ} represent the output of the Gauss Markov process. Alternatively, the sensor unit also offers a quaternion attitude representation, which is obtained from the Euler angles by equation (B.54) in the Appendix.

$$\begin{aligned} q_{0,MEAS} &= \cos\left(\Phi_{MEAS}/2\right)\cos\left(\Theta_{MEAS}/2\right)\cos\left(\Psi_{MEAS}/2\right)+\sin\left(\Phi_{MEAS}/2\right)\sin\left(\Theta_{MEAS}/2\right)\sin\left(\Psi_{MEAS}/2\right) \\ q_{1,MEAS} &= \sin\left(\Phi_{MEAS}/2\right)\cos\left(\Theta_{MEAS}/2\right)\cos\left(\Psi_{MEAS}/2\right)-\cos\left(\Phi_{MEAS}/2\right)\sin\left(\Theta_{MEAS}/2\right)\sin\left(\Psi_{MEAS}/2\right) \\ q_{2,MEAS} &= \cos\left(\Phi_{MEAS}/2\right)\sin\left(\Theta_{MEAS}/2\right)\cos\left(\Psi_{MEAS}/2\right)+\sin\left(\Phi_{MEAS}/2\right)\cos\left(\Theta_{MEAS}/2\right)\sin\left(\Psi_{MEAS}/2\right) \end{aligned} \tag{2.86}$$

Table 2.10 presents the parameters, used for the attitude sensor model.

Table 2.10 Sensor Errors for AHRS

static bias [°]			dynamic bias [°]			bandwidth [Hz]
θ_x	θ_y	θ_z	m_{Ψ}	m_{Θ}	m_{Φ}	
0.5	*0.5*	*1.0*	*1.0*	*1.0*	*2.0*	*0.1*

Kinematic Velocity

The kinematic velocity is measured in ECEF components. Also, the velocity measurement is subjected to errors due to sensor positions, deviating from the c.g.. Hence, the velocity at the aircraft reference point is corrupted by a lever arm term.

$$\left(\vec{\mathbf{V}}_K^S\right)_E^E = \mathbf{M}_{EB}\cdot\left[\left(\vec{\mathbf{V}}_K^R\right)_B^E + \left(\vec{\boldsymbol{\omega}}^{EB}\right)_B \times \left(\vec{\mathbf{r}}^{RS}\right)_B\right] \tag{2.87}$$

The kinematic velocity is estimated by the Kalman filter and as such, similar statistical errors are added as done for the attitude. Thus, the raw velocity measurements will be modeled by

$$\left(\vec{\mathbf{V}}_{K,MEAS}^S\right)_E^E = \left(\vec{\mathbf{V}}_K^S\right)_E^E + \vec{\mathbf{m}}_V \tag{2.88}$$

where $\vec{\mathbf{m}}_V$ is a 3 dimensional vector of Gauss Markov processes. Table 2.11 lists the parameters, used for the sensor model.

Table 2.11 Sensor Errors for Velocity

dynamic bias [m/s]			bandwidth [Hz]
x-axis	y-axis	z-axis	
0.5	*0.5*	*0.5*	*0.1*

Position

As the position is also as state of the Kalman filter, the sensor is modeled similarly as is done for attitude and velocity. The magnitude of position uncertainty is typically large compared to the scale of the aircraft and hence errors due to the deviated positioning of the GPS antenna are neglected. All errors are assumed to be incorporated within a Gauss Markov process $\mathbf{m}_r$, whose data are presented in Table 2.12.

$$\left(\vec{\mathbf{r}}^{S}_{MEAS}\right)_E = \left(\vec{\mathbf{r}}^{R}\right)_E + \vec{\mathbf{m}}_r \tag{2.89}$$

Table 2.12 Sensor Errors for Position

dynamic bias [m]			bandwidth[Hz]
x-axis	y-axis	z-axis	
2.5	*2.5*	*2.5*	*0.1*

As the sensor unit delivers the position information in WGS-84 coordinates the measurements in (2.89) have to be transformed. However a transformation from Cartesian ECEF to WGS-84 coordinates is not trivial due to the nonlinearity of the relationship, but solutions to the problem can e.g. be found in [Far99], [Hei82], [Kap96] and [Wen07].

2.5 Atmosphere

In aviation applications, the International Standard Atmosphere (ISA), which has been defined in 1975 and is established in DIN ISO 2533 is commonly utilized. It essentially describes an idealized dependence of the atmosphere states (air pressure, air density, air temperature) on the altitude. The ISA is divided in three parts

- troposphere: *-2km* to *11km*
- lower stratosphere: *11km* to *20km*
- upper stratosphere: *20km* to *32km*

which basically differ in the temperature-altitude-gradient. The troposphere has a constant negative temperature gradient, the lower stratosphere has a constant temperature and the upper stratosphere has a constant positive temperature gradient. Integration of Euler's hydrostatic law, using the ideal gas law as well as the assumed temperature gradients yields the following equations for the ISA troposphere with assumed constant gravity constant.

$$T(h) = T_s \cdot \left(1 - \frac{\gamma}{T_s} h\right) \quad , \quad p(h) = p_s \cdot \left(1 - \frac{\gamma}{T_s} h\right)^{\frac{n}{n-1}} \quad , \quad \rho(h) = \rho_s \cdot \left(1 - \frac{\gamma}{T_s} h\right)^{\frac{1}{n-1}} \tag{2.90}$$

where:

- $\gamma = -6.5 \cdot 10^{-3}\, {}^{K}\!/_{m}$: temperature gradient
- $T_s = 18815K$ air temperature at mean sea level (MSL)
- $p_s = 1.01325\, {}^{N}\!/_{m^2}$ air pressure at MSL
- $\rho_s = 1.225\, {}^{kg}\!/_{m^3}$: air density at MSL

- $n = 1.235$: polytropic exponent

Since the considered aircraft usually operates only in a low altitude range, the troposphere equations are sufficient for the simulation model and the assumption of constant gravity is valid.

Chapter 3

Novel Formulation of the Ultimate Boundedness Theorem

The Lyapunov stability framework is a central tool for design and analysis of adaptive control systems. Particularly in the last three decades, Lyapunov methods aroused interest of the adaptive control research community. Although it is state of the technology in these days, a comprehensive introduction to the fundamentals of Lyapunov stability is given in Appendix C. The proof that the states of a nonlinear dynamic system remain bounded is often based on a specific Lyapunov based theorem, referred to as “ultimate boundedness theorem” ([Kha02]). The current Chapter 3 is dedicated to a novel formulation of the ultimate boundedness theorem and the need for a reformulation is motivated in the following.

Khalil’s book ([Kha02]) became a standard work in the field of nonlinear system design and analysis and is frequently cited in various publications, amongst others, for its stability theorems. It is a fact that systems, for which asymptotic stability can be proofed, are – especially in aviation – of rather academical interest. In real aircraft control applications, there will always be some unmodeled dynamics or uncertainties that cannot be compensated by adaptation, which inhibits a proof of asymptotical stability. However it is still possible to proof boundedness of the system states under certain conditions. In [Kha02], the according theorem is formulated such that it allows the conclusion that there are conditions under which the system states are bounded. But quantitative values for the conditions and upper bounds on the system states are not computed explicitly.

Moreover, the theorems in [Kha02] are formulated for a generic nonlinear dynamic system with a single system state vector. Yet model reference adaptive control (MRAC) systems have a special structure where the whole state is usually partitioned into tracking and parameter error respectively.

In the following, the theorem for boundedness in [Kha02] is presented in a novel formulation as it is tailored to the special structure of MRAC systems, accounting for partitioning of the system states. Beside an explicit specification of the set of allowable initial conditions, such that boundedness of the system states is guaranteed, the theorem additionally provides explicit values for bounds, ultimate bounds (refer to

Definition C.4) and the time, at which the ultimate bound is reached. Hopefully, due to provision for the special MRAC structure, the bounds are less conservative than a generic formulation.

Particularly the latter feature is quite important for the design of flight control systems since explicit and tight bounds have to be guaranteed to certification authorities. This chapter is thus a central part of this work.

Theorem 3.1 Ultimate Boundedness

Consider the dynamic system $\dot{\mathbf{x}}(t)=\mathbf{f}(\mathbf{x},t)$, $\mathbf{x}(t_0)=\mathbf{x}_0$

where $\mathbf{f}:\mathcal{D}\times[0,\infty)\to\mathbb{R}^n$ is locally Lipschitz in $\mathbf{x}$ and piecewise continuous in t ,

$\mathcal{D}\subset\mathbb{R}^n$ is some region that contains the origin. Let

- the state vector be partitioned:
  $\mathbf{x}^T=\begin{pmatrix}\mathbf{x}_1^T & \mathbf{x}_2^T\end{pmatrix}$ where $\mathbf{x}_1\in\mathbb{R}^{n_1}$, $\mathbf{x}_2\in\mathbb{R}^{n_2}$, $n_1+n_2=n$
- constants $r_1>0$ and $r_2>0$ and sets
  $\overline{\mathcal{B}}_{r_1}=\left\{\mathbf{x}_1\in\mathbb{R}^{n_1}\,\middle|\,\|\mathbf{x}_1\|\le r_1\right\}$, $\overline{\mathcal{B}}_{r_2}=\left\{\mathbf{x}_2\in\mathbb{R}^{n_2}\,\middle|\,\|\mathbf{x}_2\|\le r_2\right\}$ such that $\overline{\mathcal{B}}_r:=\overline{\mathcal{B}}_{r_1}\times\overline{\mathcal{B}}_{r_2}\subset\mathcal{D}$
- a continuously differentiable function $V:\overline{\mathcal{B}}_r\times[0,\infty)\to\mathbb{R}$
- class $\mathcal{K}$ functions $\alpha_1(\|\mathbf{x}_1\|),\beta_1(\|\mathbf{x}_1\|)$ for $\mathbf{x}_1\in\overline{\mathcal{B}}_{r_1}$
- class $\mathcal{K}$ functions $\alpha_2(\|\mathbf{x}_2\|),\beta_2(\|\mathbf{x}_2\|)$ for $\mathbf{x}_2\in\overline{\mathcal{B}}_{r_2}$
- positive definite functions $W_1(\|\mathbf{x}_1\|),W_2(\|\mathbf{x}_2\|)$ for $\mathbf{x}\in\overline{\mathcal{B}}_r$
- a set $\mathcal{B}_\mu=\left\{\mathbf{x}\in\overline{\mathcal{B}}_r\,\middle|\,\|\mathbf{x}_1\|<\mu_1,\|\mathbf{x}_2\|<\mu_2\right\}$ for some $0<\mu_1<\rho_1$ and $0<\mu_2<\rho_2$

  where $\rho_1=\beta_1^{-1}\left(\frac{\underline{u}}{2}\right)$, $\rho_2=\beta_2^{-1}\left(\frac{\underline{u}}{2}\right)$ and $\underline{u}=\min(\alpha_1(r_1),\alpha_2(r_2))$

such that

1. (condition) $\alpha_1(\|\mathbf{x}_1\|)+\alpha_2(\|\mathbf{x}_2\|)\le V(\mathbf{x},t)\le\beta_1(\|\mathbf{x}_1\|)+\beta_2(\|\mathbf{x}_2\|)$ on $\overline{\mathcal{B}}_r\times[0,\infty)$
2. (condition) $\frac{\partial V(\mathbf{x},t)}{\partial t}+\frac{\partial V(\mathbf{x},t)}{\partial\mathbf{x}}\cdot\mathbf{f}(\mathbf{x},t)\le-\max(W_1(\mathbf{x}_1),W_2(\mathbf{x}_2))$ on $\overline{\mathcal{B}}_r\setminus\mathcal{B}_\mu$

If the initial conditions satisfy $\|\mathbf{x}_{1,0}\|\le\delta_1$ and $\|\mathbf{x}_{2,0}\|\le\delta_2$

where $\delta_1\in[0,\rho_1]$ and $\delta_2\in[0,\rho_2]$

Then $\|\mathbf{x}_1(t)\|\le\alpha_1^{-1}(\underline{u})$, $\|\mathbf{x}_2(t)\|\le\alpha_2^{-1}(\underline{u})$ for all $t\ge t_0$

and every solution of the system
is uniformly ultimately bounded
with ultimate bounds $b_1=\alpha_1^{-1}(v)$ for $\mathbf{x}_1$ and $b_2=\alpha_2^{-1}(v)$ for $\mathbf{x}_2$
and $v=\beta_1(\mu_1)+\beta_2(\mu_2)$

Proof

At first note that, for the admissible initial values, by condition 1, we have

$$V(\mathbf{x}_0, t_0) \leq \beta_1(\rho_1) + \beta_2(\rho_2).$$

Hence, by the definition of ρ_1, ρ_2:

$$V(\mathbf{x}_0, t_0) \leq \underline{u} \tag{3.1}$$

The set

$$\overline{\Omega}_{\underline{u},t} = \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, V(\mathbf{x}, t) \leq \underline{u} \right\}$$

is contained in $\overline{\mathcal{B}_r}$ since condition 1 of the theorem implies that $\|\mathbf{x}_1\| \leq \alpha_1^{-1}(\underline{u}) \leq \alpha_1^{-1}(\alpha_1(r_1)) = r_1$ and $\|\mathbf{x}_2\| \leq \alpha_2^{-1}(\underline{u}) \leq \alpha_2^{-1}(\alpha_2(r_2)) = r_2$. Furthermore, $\overline{\Omega}_{\underline{u},t}$ is an invariant set (i.e. if once the trajectory has entered or started within this set, it will remain there for all future time), since on the boundary of $\overline{\Omega}_{\underline{u}}$ which is defined by

$$\partial\overline{\Omega}_{\underline{u},t} = \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, V(\mathbf{x}, t) = \underline{u} \right\}$$

we have by condition 1 of the theorem that at least

- $\|\mathbf{x}_1\| \geq \beta_1^{-1}\left(\underline{u}/2\right) = \rho_1 > \mu_1$ or
- $\|\mathbf{x}_2\| \geq \beta_2^{-1}\left(\underline{u}/2\right) = \rho_2 > \mu_2$

since, imagine that simultaneously $\|\mathbf{x}_1\| < \rho_1$ and $\|\mathbf{x}_2\| < \rho_2$, then condition 1 implies

$$V(\mathbf{x}, t) \leq \beta_1(\|\mathbf{x}_1\|) + \beta_2(\|\mathbf{x}_2\|) < \beta_1(\rho_1) + \beta_2(\rho_2) = \frac{\underline{u}}{2} + \frac{\underline{u}}{2} = \underline{u}.$$

Hence $\partial\overline{\Omega}_{\underline{u},t}$ is contained in $\overline{\mathcal{B}_r} \setminus \mathcal{B}_\mu$ where $\dot{V} < 0$ according to condition 2. This implies that, if a trajectory, which starts within $\overline{\Omega}_{\underline{u},t}$ (which is fulfilled for admissible initial conditions), hits the boundary at some time instant then $\dot{V} < 0$ and hence the trajectory cannot leave the set, rendering $\overline{\Omega}_{\underline{u},t}$ invariant. Consequently, a trajectory, starting within $\overline{\Omega}_{\underline{u},t}$ also remains within $\overline{\mathcal{B}_r}$, since $\Omega_{\underline{u},t} \subset \overline{\mathcal{B}_r}$. Further, since $V(x,t) \leq \underline{u}$, condition 1 of the theorem implies

$$\alpha_1(\|\mathbf{x}_1\|) \leq \underline{u} \text{ and } \alpha_2(\|\mathbf{x}_2\|) \leq \underline{u}$$

which establishes the first bound of the theorem.

$$\|\mathbf{x}_1\| \leq \alpha_1^{-1}(\underline{u}) \text{ and } \|\mathbf{x}_2\| \leq \alpha_2^{-1}(\underline{u})$$

The following

$$\overline{\Omega}_{v,t} = \left\{ \mathbf{x} \in \overline{\mathcal{B}}_r \,\middle|\, V(\mathbf{x},t) \le v \right\}$$

is also an invariant set since, by condition 1 of the theorem and the definition of v, we have, on the boundary $\partial\overline{\Omega}_{v,t}$ that, at least

- $\|\mathbf{x}_1\| \ge \mu_1$ or
- $\|\mathbf{x}_2\| \ge \mu_2$.

This implies that $\partial\overline{\Omega}_{v,t} \subset (\overline{\mathcal{B}}_r \setminus \mathcal{B}_\mu)$ and, for the same reasons as stated recently, the set is rendered an invariant set. Figure 3.1 depicts the Lyapunov function over state for the case of only 1 vector partition while Figure 3.2 illustrates the Lyapunov level sets for the case of 2 vector partitions.

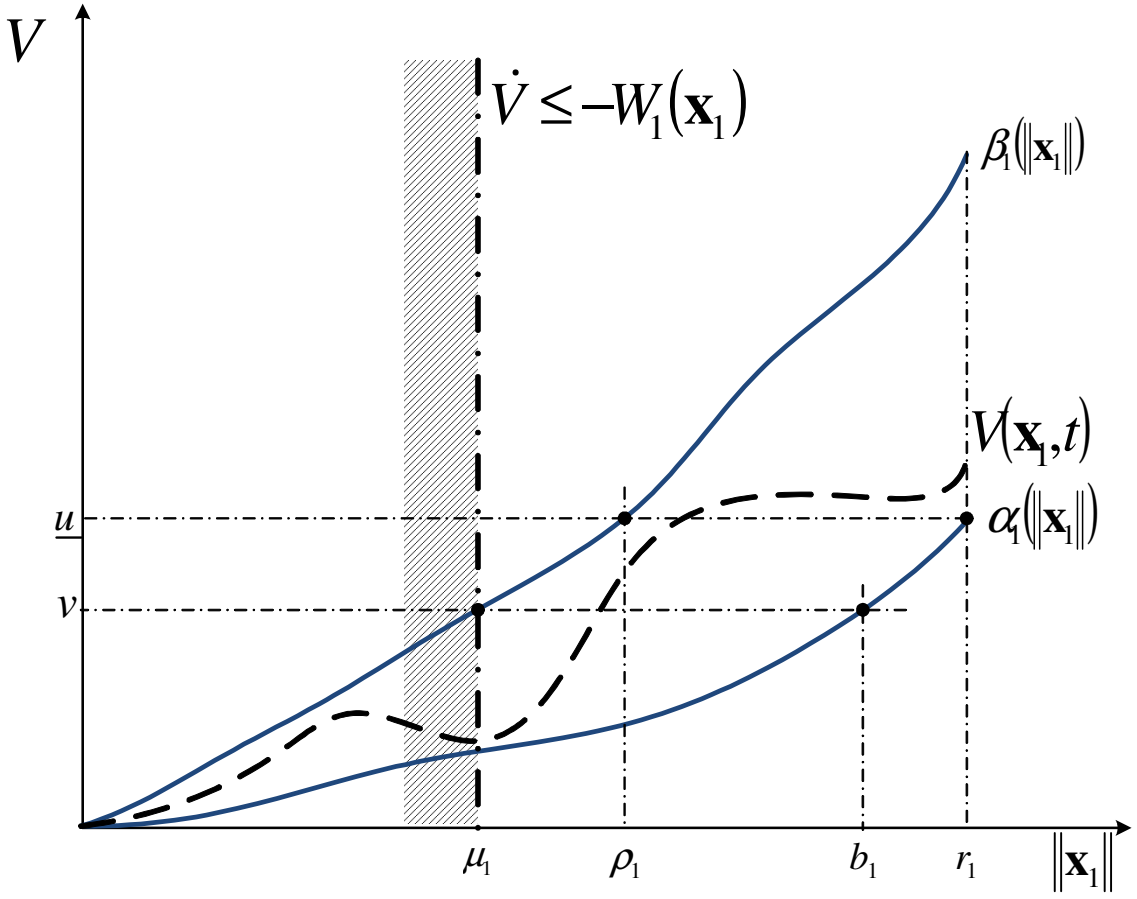


Figure 3.1 Lyapunov Function for 1 State Vector Partition

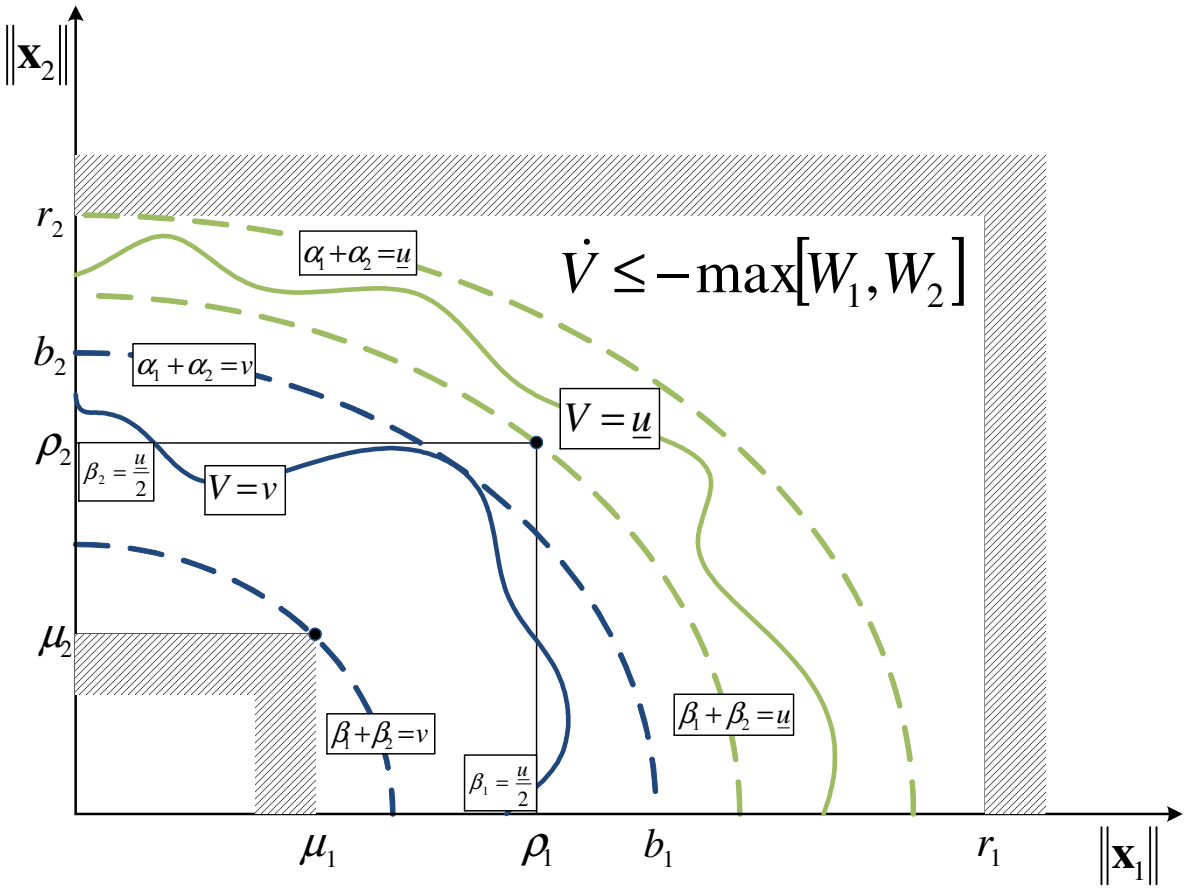


Figure 3.2 Lyapunov Level Sets in Case of 2 Vector Partitions

Next, it is shown by contradiction that any trajectory with allowable initial condition enters $\overline{\Omega}_{v,t}$ in some finite time. Therefore note that, by Lemma C.2, there are class $\mathcal{K}$ functions $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$ such that

$$\gamma_1\left(\|\mathbf{x}_1\|\right)\le W(\mathbf{x}_1) \text{ and } \gamma_2\left(\|\mathbf{x}_2\|\right)\le W_2(\mathbf{x}_2) \tag{3.2}$$

for $||\mathbf{x}_1||\le r_1$ and $||\mathbf{x}_2||\le r_2$. Now, let's restrict the initial conditions such that $||\mathbf{x}_{1,0}||\le\delta_1$ and $||\mathbf{x}_{2,0}||\le\delta_2$ for some $\delta_1\in[0,\rho_1]$ and $\delta_2\in[0,\rho_2]$ and assume $\mathbf{x}(t)$ remains in the set

$$\tilde{\Omega}_{v,t}=\left\{\mathbf{x}\in\overline{\mathcal{B}}_r\,\middle|\,V(\mathbf{x},t)>v\right\}$$

for some time interval $[t_0,t_1]$. Then (3.2) and condition 2 of the theorem imply that

$$\dot{V}(\mathbf{x},t)\le-\max(\gamma_1(\mu_1),\gamma_2(\mu_2)) \tag{3.3}$$

for $t\in[t_0,t_1]$ (also consider Figure 3.2 which shows that $\tilde{\Omega}_{v,t}\subset\overline{\mathcal{B}}_r\setminus\mathcal{B}_\mu$) and from condition 1 of the theorem one can conclude that

$$V(\mathbf{x}(t_0),t_0)\le\beta_1(\delta_1)+\beta_2(\delta_2) \tag{3.4}$$

Integration of the Lyapunov function, on inspection of (3.3) and (3.4) yields

$$V(\mathbf{x}(t_1),t_1)=V(\mathbf{x}(t_0),t_0)+\int_{t_0}^{t_1}\dot{V}(\mathbf{x}(\tau),\tau)d\tau\le\beta_1(\delta_1)+\beta_2(\delta_2)-\max(\gamma_1(\mu_1),\gamma_2(\mu_2))\cdot(t_1-t_0).$$

Then for

$$t_1 = t_0 + \frac{\beta_1(\delta_1) + \beta_2(\delta_2) - v}{\max(\gamma_1(\mu_1), \gamma_2(\mu_2))}$$

we have $V(\mathbf{x}(t_1), t_1) \leq v$ which is a contradiction and hence $\mathbf{x}(t)$ enters $\overline{\Omega}_{v,t}$ within $[t_0, t_1]$. Moreover condition 1 of the theorem implies that.

$$\|\mathbf{x}_1\| \leq b_1 =: \alpha_1^{-1}(v) \text{ and } \|\mathbf{x}_2\| \leq b_2 =: \alpha_2^{-1}(v) \tag{3.5}$$

if $\mathbf{x} \in \overline{\Omega}_{v,t}$. Hence for $||\mathbf{x}_{1,0}|| \leq \delta_1$ and $||\mathbf{x}_{2,0}|| \leq \delta_2$ for some $0 \leq \delta_1 \leq \rho_1$ and $0 \leq \delta_2 \leq \rho_2$ it can be concluded that

$$\|\mathbf{x}_1(t)\| \leq b_1 \text{ and } \|\mathbf{x}_2(t)\| \leq b_2$$

for all $t \geq t_0 + T(\delta_1, \delta_2, b_1, b_2)$ where

$$T(\delta_1, \delta_2, b_1, b_2) = \frac{\beta_1(\delta_1) + \beta_2(\delta_2) - v}{\max[\gamma_1(\mu_1), \gamma_2(\mu_2)]}. \tag{3.6}$$

So far we have shown a somehow “coupled” ultimate boundedness” meaning that, for a pair of initial conditions δ_1, δ_2 we obtain a pair of ultimate bounds b_1, b_2, which are both reached within the same elapsed time $T(\delta_1, \delta_2)$. But ultimate boundedness is only defined in an uncoupled way in Definition C.4. In order to comply with this definition, ultimate boundedness has to be established for each partial vector separately. Therefore we need triples $(b_1, \delta_1, T_1(\delta_1, b_1))$ and $(b_2, \delta_2, T_2(\delta_2, b_2))$ which are independent of each other. Therefore we have to find separate times T_1 and T_2 which only depend on δ_1 and δ_2 respectively. In order to obtain some T_1 for the first partition, independent of δ_2, we simple set the initial condition of the second part to the worst case value in (3.6), i.e. $\delta_2 = \rho_2$. In an analogous manner we obtain T_2 and result in the following ultimate boundedness statements:

$$\|\mathbf{x}_1(t)\| \leq b_1$$

for all $t \geq t_0 + T_1(\delta_1, b_1)$ and $||\mathbf{x}_{1,0}|| \leq \delta_1$ where

$$T_1(\delta_1, b_1) = \max\left[0, \frac{\beta_1(\delta_1) + \frac{u}{2} - v}{\max(\gamma_1(\mu_1), \gamma_2(\mu_2))}\right] \text{ and } b_1 = \alpha_1^{-1}(v)$$

and

$$\|\mathbf{x}_2(t)\| \leq b_2$$

for all $t \geq t_0 + T_1(\delta_2, b_2)$ and $||\mathbf{x}_{2,0}|| \leq \delta_2$ where

$$T_2(\delta_2, b_2) = \max\left[0, \frac{\beta_2(\delta_2) + \mu/2 - \nu}{\max(\gamma_1(\mu_1), \gamma_2(\mu_2))}\right] \text{ and } b_2 = \alpha_2^{-1}(\nu)$$

Note that we used the definition of ρ_1, ρ_2 which implies $\beta_1(\rho_1) = \beta_2(\rho_2) = \mu/2$ and T_1, T_2 are, of course, limited from below by zero, since for sufficiently small initial conditions, the state trajectory starts within the invariant set $\overline{\Omega}_{\nu,t}$ (Figure 3.2) and hence $||\mathbf{x}_{1,0}|| \leq b_1$ and $||\mathbf{x}_{2,0}|| \leq b_2$ are fulfilled at once.

□

Theorem 3.1 only delivers ultimate bounds b_i but there is no statement about the time T in which the bound is reached. But this of particular interest in the field of flight control, since it is a matter of performance, how fast transient dynamics will decay and performance guarantees in turn are an essential criterion for certification of flight control systems. Nevertheless, the proof to Theorem 3.1 contains some expression for T_i if the Lyapunov function derivative is bounded by some class $\mathcal{K}$ function as introduced in (3.2), which is summarized in the next corollary.

Corollary 3.1 Ultimate Boundedness with Explicit Convergence Rate

Consider the dynamic system $\dot{\mathbf{x}}(t)=\mathbf{f}(\mathbf{x},t)$, $\mathbf{x}(t_0)=\mathbf{x}_0$

where $\mathbf{f}:\mathcal{D}\times[0,\infty)\to\mathbb{R}^n$ is locally Lipschitz in $\mathbf{x}$ and piecewise continuous in t

$\mathcal{D}\subset\mathbb{R}^n$ is some region that contains the origin. Let

- the state vector be partitioned:
  $\mathbf{x}^T=\begin{pmatrix}\mathbf{x}_1^T & \mathbf{x}_2^T\end{pmatrix}$ where $\mathbf{x}_1\in\mathbb{R}^{n_1}$, $\mathbf{x}_2\in\mathbb{R}^{n_2}$
- constants $r_1>0$ and $r_2>0$ and sets
  $\overline{\mathcal{B}}_{r_1}=\left\{\mathbf{x}_1\in\mathbb{R}^{n_1}\,\middle|\,\|\mathbf{x}_1\|\le r_1\right\}$, $\overline{\mathcal{B}}_{r_2}=\left\{\mathbf{x}_2\in\mathbb{R}^{n_2}\,\middle|\,\|\mathbf{x}_2\|\le r_2\right\}$ such that $\overline{\mathcal{B}}_r:=\overline{\mathcal{B}}_{r_1}\times\overline{\mathcal{B}}_{r_2}\subset\mathcal{D}$
- a continuously differentiable function $V:\overline{\mathcal{B}}_r\times[0,\infty)\to\mathbb{R}$
- class $\mathcal{K}$ functions $\alpha_1(\|\mathbf{x}_1\|),\beta_1(\|\mathbf{x}_1\|),\gamma_1(\|\mathbf{x}_1\|)$ for $\mathbf{x}_1\in\overline{\mathcal{B}}_{r_1}$
- class $\mathcal{K}$ functions $\alpha_2(\|\mathbf{x}_2\|),\beta_2(\|\mathbf{x}_2\|),\gamma_2(\|\mathbf{x}_2\|)$ for $\mathbf{x}_2\in\overline{\mathcal{B}}_{r_2}$
- a set $\mathcal{B}_\mu=\left\{\mathbf{x}\in\overline{\mathcal{B}}_r\,\middle|\,\|\mathbf{x}_1\|<\mu_1,\|\mathbf{x}_2\|<\mu_2\right\}$ for some $0<\mu_1<\rho_1$ and $0<\mu_2<\rho_2$

  where $\rho_1=\beta_1^{-1}\left(\frac{\underline{u}}{2}\right)$, $\rho_2=\beta_2^{-1}\left(\frac{\underline{u}}{2}\right)$ and $\underline{u}=\min(\alpha_1(r_1),\alpha_2(r_2))$

such that

1. $\alpha_1(\|\mathbf{x}_1\|)+\alpha_2(\|\mathbf{x}_2\|)\le V(\mathbf{x},t)\le\beta_1(\|\mathbf{x}_1\|)+\beta_2(\|\mathbf{x}_2\|)$ on $\overline{\mathcal{B}}_r\times[0,\infty)$
2. $\dfrac{\partial V(\mathbf{x},t)}{\partial t}+\dfrac{\partial V(\mathbf{x},t)}{\partial\mathbf{x}}\cdot\mathbf{f}(\mathbf{x},t)\le-\max(\gamma_1(\|\mathbf{x}_1\|),\gamma_2(\|\mathbf{x}_2\|))$ on $\overline{\mathcal{B}}_r\setminus\mathcal{B}_\mu$

If the initial conditions satisfy $\|\mathbf{x}_{1,0}\|\le\delta_1$ and $\|\mathbf{x}_{2,0}\|\le\delta_2$

where $\delta_1\in[0,\rho_1]$ and $\delta_2\in[0,\rho_2]$

Then $\|\mathbf{x}_1(t)\|\le\alpha_1^{-1}(\underline{u})$, $\|\mathbf{x}_2(t)\|\le\alpha_2^{-1}(\underline{u})$ for all $t\ge t_0$

and every solution of the system
is uniformly ultimately bounded such that
$\|\mathbf{x}_1(t)\|\le b_1$ for all $t\ge t_0+T_1(\delta_1,b_1)$ and $\|\mathbf{x}_2(t)\|\le b_2$ for all $t\ge t_0+T_2(\delta_2,b_2)$

where $b_1=\alpha_1^{-1}(v)$, $b_2=\alpha_2^{-1}(v)$, $v=\beta_1(\mu_1)+\beta_2(\mu_2)$

$$T_1(\delta_1,b_1)=\max\left[0,\frac{\beta_1(\delta_1)+\underline{u}/2-v}{\max(\gamma_1(\mu_1),\gamma_2(\mu_2))}\right]$$

$$T_2(\delta_2,b_2)=\max\left[0,\frac{\beta_2(\delta_2)+\underline{u}/2-v}{\max(\gamma_1(\mu_1),\gamma_2(\mu_2))}\right]$$

So far we have restricted ourselves to the case of 2 partitions of the state vector. However, it can also be extended to a general case, where the state vector is divided into an arbitrary number of partitions.

Corollary 3.2 Ultimate Boundedness for Arbitrary Number of State Vector Partitions

Consider the dynamic system $\dot{\mathbf{x}}(t)=\mathbf{f}(\mathbf{x},t)$, $\mathbf{x}(t_0)=\mathbf{x}_0$

where $\mathbf{f}:\mathcal{D}\times[0,\infty)\to\mathbb{R}^n$ is locally Lipschitz in $\mathbf{x}$ and piecewise continuous in t

$\mathcal{D}\subset\mathbb{R}^n$ is some region that contains the origin. Let

- the state vector be partitioned:
  $\mathbf{x}^T=\begin{pmatrix}\mathbf{x}_1^T & \cdots & \mathbf{x}_k^T\end{pmatrix}$ where $\mathbf{x}_i\in\mathbb{R}^{n_i}$, $i=1,\ldots,k$, $k\in\mathbb{N}$
- constants $r_i>0$ and sets $\overline{\mathcal{B}}_{r_i}=\left\{\mathbf{x}_i\in\mathbb{R}^{n_i}\,\middle|\,\|\mathbf{x}_i\|\le r_i\right\}$, such that $\overline{\mathcal{B}}_r:=\overline{\mathcal{B}}_{r_1}\times\cdots\overline{\mathcal{B}}_{r_k}\subset\mathcal{D}$
- a continuously differentiable function $V:\overline{\mathcal{B}}_r\times[0,\infty)\to\mathbb{R}$
- class $\mathcal{K}$ functions $\alpha_i(\|\mathbf{x}_i\|),\beta_i(\|\mathbf{x}_i\|),\gamma_i(\|\mathbf{x}_i\|)$ for $\mathbf{x}_i\in\overline{\mathcal{B}}_{r_i}$
- a set $\mathcal{B}_\mu=\left\{\mathbf{x}\in\overline{\mathcal{B}}_r\,\middle|\,\|\mathbf{x}_i\|<\mu_i,i=1,\ldots,k\right\}$ for some $0<\mu_i<\rho_i$

  where $\rho_i=\beta_i^{-1}\left(\frac{\underline{u}}{k}\right)$ and $\underline{u}=\min_{i=1,\ldots,k}(\alpha_i(r_i))$

such that

1. $\sum_{i=1}^k\alpha_i(\|\mathbf{x}_i\|)\le V(\mathbf{x},t)\le\sum_{i=1}^k\beta_i(\|\mathbf{x}_i\|)$ on $\overline{\mathcal{B}}_r\times[0,\infty)$
2. $\frac{\partial V(\mathbf{x},t)}{\partial t}+\frac{\partial V(\mathbf{x},t)}{\partial\mathbf{x}}\cdot\mathbf{f}(\mathbf{x},t)\le-\max_{i=1,\ldots,k}(\gamma_i(\|\mathbf{x}_i\|))$ on $\overline{\mathcal{B}}_r\setminus\overline{\mathcal{B}}_\mu\times[0,\infty)$

If the initial conditions satisfy $\|\mathbf{x}_{i,0}\|\le\delta_i$ where $\delta_i\in[0,\rho_i]$

then $\|\mathbf{x}_i(t)\|\le\alpha_i^{-1}(\underline{u})$ for all $t\ge t_0$

and every solution of the system

is uniformly ultimately bounded such that $\|\mathbf{x}_i(t)\|\le b_i$ for all $t\ge t_0+T_i(\delta_i,b_i)$

where $b_i=\alpha_i^{-1}(v)$, $v=\sum_{i=1}^k\beta_i(\mu_i)$ and

$$T_i(\delta_i,b_i)=\max\left[0,\frac{\beta_i(\delta_i)+\frac{k-1}{k}\underline{u}-v}{\max_{j=1,\ldots,k}(\gamma_j(\mu_j))}\right]$$

Chapter 4

Background on Control Theory

This Chapter presents the theoretical background, needed for design of the flight control system. Thereby, it is intended to develop a control algorithm that accounts for the nonlinearities of the aircraft dynamics, on the one hand, which allows a full exploitation of the physical aircraft capabilities. Contrary, to linear control concepts, which merely allow the system to be controlled in a vicinity of a set-point, a nonlinear control strategy is globally valid within the whole flight envelope. Therefore, the control strategy that is chosen for a baseline controller is a reference model based nonlinear dynamic inversion (NDI) controller, whose basics are presented in section 4.1

On the other hand, particular focus is put onto tolerance against actuator failures, since this significantly increases survivability, which is particularly interesting for unmanned aerial systems, which should be capable of autonomous mission accomplishment. Hence the nonlinear baseline controller is augmented by an adaptive part that compensates for model uncertainties and sudden configuration changes which occur e.g. due to actuator failures and partial loss of the aircraft structure. The FSD Extreme Star is particularly suitable for test of such failure scenarios due to its highly redundant actuation.

In sections 4.2 and 4.3, the basic adaptive element is introduced, including a Lyapunov stability analysis. Section 4.4 presents a novel modification of the learning law that improves adaptation performance. It is in fact an enhancement of the q modification term ([Vol06] [Vol061]). Section 4.5 gives a short introduction to a method that allows the incorporation of actuators into the control algorithm, which enter the dynamics nonlinearly, using an online gradient minimization approach. The thrust vector controls are e.g. nonaffine, due to trigonometric functions. Most of the results, presented here are published ([Lav08], [Lav07b], [Lav09], [Lav07a]), but some minor extensions and generalizations are added. The incorporation of the introduced nonaffine-in-control algorithm into the whole flight control framework is derived in section 4.6. Finally, section 4.7 introduces a novel way for integration of the adaptive parameters in terms of its singular value decomposition, which provides the possibility to avoid singular matrices during integration. This is particularly interesting for the adaptive estimate of the control effectiveness, since has to be inverted within the NDI algorithm.

Notice that the Lyapunov stability analysis is based on the novel formulation of the ultimate boundedness theorem in Chapter 3, providing explicit values for bounds on the tracking and parameter errors as well as times when the bounds are reached.

4.1 Nonlinear Dynamic Inversion

Aim of NDI is finding a linearizing state feedback, such that the dynamic relationship between a newly defined pseudo control and the output is a linear one. Therefore, the dynamic system has to be transformed into Byrnes-Isidori normal form ([Mic90]) by means of a nonlinear state transformation, which is obtained by repeatedly deriving each output w.r.t. time until the input appears in the equations. NDI is a widespread concept in flight control and has a well-developed theory, which amongst others can be found in [Isi95], [Kha02], [Slo90] , [Hol04]. Nevertheless important facts on NDI are developed in the following.

4.1.1 Nonlinear State Transformation

Consider the following class of dynamical systems:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t)) + \mathbf{G}(\mathbf{x}(t))\cdot\mathbf{u}(t)\\ \mathbf{y}(t) &= \mathbf{h}(\mathbf{x}(t))\end{aligned} \tag{4.1}$$

with initial condition $\mathbf{x}(t_0)=\mathbf{x}_0$. Thereby $\mathbf{x}^T(t)=(x_1(t) \;\cdots\; x_n(t))\in\mathbb{R}^n$ is the state vector, $\mathbf{u}^T(t)=(u_1(t) \;\cdots\; u_m(t))\in\mathbb{R}^m$ is an external input consisting of piecewise continuous signals, $\mathbf{y}^T(t)=(y_1(t) \;\cdots\; y_m(t))\in\mathbb{R}^m$ is the output vector and $\mathbf{f}:\mathcal{D}\to\mathbb{R}^n$, $\mathbf{G}=[\mathbf{g}_1(\mathbf{x}) \;\cdots\; \mathbf{g}_m(\mathbf{x})]$, $\mathbf{g}_i:\mathcal{D}\to\mathbb{R}^n$, $i=1,\ldots,m$, $\mathbf{h}:\mathcal{D}\to\mathbb{R}^m$ are smooth (continuously differentiable up to any degree) vector fields on the open set $\mathcal{D}\subset\mathbb{R}^n$. Note that the class of systems is restricted to the case with the same number m of inputs and outputs. Further (4.1) has some stationary point $\mathbf{x}_s\in\mathcal{D}$ such that $\mathbf{f}(\mathbf{x}_s)=\mathbf{0}$. Without loss of generality, it is assumed that $\mathbf{h}(\mathbf{x}_S)=\mathbf{0}$ since, if $\mathbf{h}(\mathbf{x}_S)=\mathbf{y}_S\neq\mathbf{0}$, then another output is easily defined by $\bar{\mathbf{y}}=\mathbf{h}(\mathbf{x})-\mathbf{y}_S$, which fulfills the assumption. In order to find the nonlinear state transformation that transforms the system to normal form, each output y_i, $i=1,\ldots,p$ has to be derived w.r.t. time.

$$\dot{y}_i(t)=\frac{\partial h_i}{\partial\mathbf{x}}\cdot(\mathbf{f}(\mathbf{x})+\mathbf{G}(\mathbf{x})\mathbf{u}) \tag{4.2}$$

Using the Lie derivatives of Definition B.31, equation (4.2) is

$$\dot{y}_i(t)=L_{\mathbf{f}}h_i(\mathbf{x})+\underbrace{L_{\mathbf{G}}h_i(\mathbf{x})}_{=0}\mathbf{u}\,. \tag{4.3}$$

If $L_{\mathbf{G}}h_i(\mathbf{x})=\mathbf{0}$, (4.3) is derived once more w.r.t. time.

$$\ddot{y}_i(t) = L_{\mathbf{f}}^2 h_i(\mathbf{x}) + \underbrace{L_{\mathbf{G}}\left(L_{\mathbf{f}} h_i(\mathbf{x})\right)}_{=0}\mathbf{u} \tag{4.4}$$

This procedure is repeated until the control vector appears, let us say in the r_i^{th} derivative.

$$\begin{aligned} &\vdots \\ y_i^{(r_i-1)}(t) &= L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x}) + \underbrace{L_{\mathbf{G}}\left(L_{\mathbf{f}}^{r_i-2} h_i(\mathbf{x})\right)}_{=0}\mathbf{u} \\ y_i^{(r_i)}(t) &= L_{\mathbf{f}}^{r_i} h_i(\mathbf{x}) + \underbrace{L_{\mathbf{G}}\left(L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x})\right)}_{\neq 0}\mathbf{u} \end{aligned} \tag{4.5}$$

Doing so for every $y_i(t)$ yields

$$\begin{pmatrix} y_1^{(r_1)} \\ \vdots \\ y_p^{(r_m)} \end{pmatrix}(t) = \bar{\mathbf{a}}(\mathbf{x}) + \bar{\mathbf{B}}(\mathbf{x})\mathbf{u}(t) \tag{4.6}$$

where

$$\bar{\mathbf{a}}(\mathbf{x}) = \begin{pmatrix} L_{\mathbf{f}}^{r_1} h_1(\mathbf{x}) \\ \vdots \\ L_{\mathbf{f}}^{r_m} h_m(\mathbf{x}) \end{pmatrix} \in \mathbb{R}^m \quad , \quad \bar{\mathbf{B}}(\mathbf{x}) = \begin{bmatrix} \bar{\mathbf{b}}_1^T(\mathbf{x}) \\ \vdots \\ \bar{\mathbf{b}}_m^T(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} L_{\mathbf{G}}\left(L_{\mathbf{f}}^{r_1-1} h_1(\mathbf{x})\right) \\ \vdots \\ L_{\mathbf{G}}\left(L_{\mathbf{f}}^{r_m-1} h_m(\mathbf{x})\right) \end{bmatrix} \in \mathbb{R}^{m\times m} \quad , \quad \mathbf{r}^T = \begin{pmatrix} r_1 & \cdots & r_m \end{pmatrix} \tag{4.7}$$

and $\mathbf{r}$ is called vector *relative degree* of the system at $\mathbf{x}_0$, according to the following definition.

Definition 4.1 Vector Relative Degree

The system (4.1) is said to have a (vector) relative degree $\mathbf{r}^T = \begin{pmatrix} r_1 & \cdots & r_m \end{pmatrix}$, at $\mathbf{x}_0$ if

- $L_{\mathbf{G}} L_{\mathbf{f}}^k h_i(\mathbf{x}) = \mathbf{0}^T$ in a vicinity of $\mathbf{x_0}$ for $k = 0,\ldots,r_i - 2,\ i = 1,\cdots,m$
- The decoupling matrix $\bar{\mathbf{B}}(\mathbf{x})$ is regular at $\mathbf{x_0}$

The existence of a vector relative degree essentially guarantees the regularity of the decoupling matrix, which is preliminary to the existence of a linearizing state feedback as will be shown later. If, however, the decoupling matrix is singular, the concept of *dynamic extension* could serve as a method to gain a well-defined relative degree. For a detailed discussion see [Isi95].

If the system has a well-defined relative degree at some $\mathbf{x_0}$, we observe that, for $k = 0,\ldots,r_i - 1$, the k^{th} derivative of the i^{th} output is independent of the input

$$y_i^{(k)} = L_{\mathbf{f}}^k h_i(\mathbf{x}) \tag{4.8}$$

which motivates the following local nonlinear transformation

$$\zeta = \mathbf{\Phi}_{\zeta}(\mathbf{x}) \tag{4.9}$$

where

$$\mathbf{\Phi}_{\zeta}(\mathbf{x}) = \begin{pmatrix} \mathbf{\Phi}_1(\mathbf{x}) \\ \vdots \\ \mathbf{\Phi}_m(\mathbf{x}) \end{pmatrix}, \ \mathbf{\Phi}_i(\mathbf{x}) = \begin{pmatrix} h_i(\mathbf{x}) \\ L_{\mathbf{f}} h_i(\mathbf{x}) \\ \vdots \\ L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x}) \end{pmatrix}, \ i = 1, \ldots, m \tag{4.10}$$

and

$$\zeta = \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_m \end{pmatrix}, \ \zeta_i = \begin{pmatrix} \zeta_{i,1} \\ \zeta_{i,2} \\ \vdots \\ \zeta_{i,r_i} \end{pmatrix} = \begin{pmatrix} y_i \\ \dot{y}_i \\ \vdots \\ y_i^{(r_i-1)} \end{pmatrix}, \ i = 1, \ldots, m \tag{4.11}$$

Clearly $\dot{\zeta}_{i,k} = \zeta_{i,k+1}$ for $k = 1, \ldots, , r_i - 1$, which can be observed by the following derivation.

$$\begin{aligned} \frac{d(\zeta_{i,k})}{dt} &= \frac{d\left(L_{\mathbf{f}}^{k-1} h_i(\mathbf{x})\right)}{dt} = \frac{d\left(L_{\mathbf{f}}^{k-1} h_i(\mathbf{x})\right)}{d\mathbf{x}} \left(\mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u}\right) \\ &= L_{\mathbf{f}}^{k} h_i(\mathbf{x}) + \underbrace{L_{\mathbf{G}} L_{\mathbf{f}}^{k-1} h_i(\mathbf{x})}_{=0} \mathbf{u} = \zeta_{i,k+1} \end{aligned} \tag{4.12}$$

and for $k = r_i$:

$$\begin{aligned} \frac{d(\zeta_{i,r_i})}{dt} &= \frac{d\left(L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x})\right)}{dt} = \frac{d\left(L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x})\right)}{d\mathbf{x}} \left(\mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u}\right) \\ &= L_{\mathbf{f}}^{r_i} h_i(\mathbf{x}) + L_{\mathbf{G}} L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x}) \mathbf{u} \end{aligned} \tag{4.13}$$

Hence, with (4.12) and (4.13) we get the following differential equations for the transformed states (4.9).

$$\begin{aligned} &\dot{\zeta}_{1,1} = \zeta_{1,2} && && \dot{\zeta}_{m,1} = \zeta_{m,2} \\ &\dot{\zeta}_{1,2} = \zeta_{1,3} && \cdots && \dot{\zeta}_{m,2} = \zeta_{m,3} \\ &\vdots && && \vdots \\ &\dot{\zeta}_{1,r_1} = L_{\mathbf{f}}^{r_1} h_1(\mathbf{x}) + L_{\mathbf{G}} L_{\mathbf{f}}^{r_1-1} h_1(\mathbf{x}) \mathbf{u} && && \dot{\zeta}_{m,r_m} = L_{\mathbf{f}}^{r_m} h_m(\mathbf{x}) + L_{\mathbf{G}} L_{\mathbf{f}}^{r_m-1} h_m(\mathbf{x}) \mathbf{u} \end{aligned} \tag{4.14}$$

which is conveniently written in matrix vector notation

$$\dot{\zeta} = \mathbf{J} \cdot \zeta + \mathbf{H}\left[\bar{\mathbf{a}}(\mathbf{x}) + \bar{\mathbf{B}}(\mathbf{x}) \cdot \mathbf{u}\right] \tag{4.15}$$

where

$$\mathbf{J}=\begin{bmatrix}\mathbf{J}_1 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \ddots & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{J}_m\end{bmatrix},\ \mathbf{J}_i=\begin{bmatrix}0 & 1 & & 0\\ 0 & \ddots & \ddots & \\ \vdots & & \ddots & 1\\ 0 & \cdots & 0 & 0\end{bmatrix}\in\mathbb{R}^{r_i\times r_i},$$
$$\mathbf{H}=\begin{bmatrix}\mathbf{h}_{r_1} & \mathbf{0} & \cdots & \mathbf{0}\\ \mathbf{0} & \mathbf{h}_{r_2} & \ddots & \vdots\\ \vdots & \ddots & \ddots & \mathbf{0}\\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{h}_{r_m}\end{bmatrix},\ \mathbf{h}_{r_i}=\begin{pmatrix}0\\ \vdots\\ 0\\ 1\end{pmatrix}\in\mathbb{R}^{r_i} \tag{4.16}$$

and $\bar{\mathbf{a}}(\mathbf{x})$, $\bar{\mathbf{B}}(\mathbf{x})$ according to (4.7). Due to their special structure, matrices $\mathbf{H}$ and $\mathbf{J}$ have some properties that appear to be useful. The following identity for $\mathbf{H}$ can be easily verified.

$$\mathbf{H}^T\mathbf{H}=\mathbf{I}. \tag{4.17}$$

This in turn implies that its largest singular value $\bar{\sigma}_H=1$ and hence the matrix induced 2-norm evaluates to

$$\|\mathbf{H}\|_2=1. \tag{4.18}$$

Moreover, it can be easily verified that

$$\mathbf{J}_i^T\mathbf{J}_i=\begin{bmatrix}0 & \mathbf{0}^T\\ \mathbf{0} & \mathbf{I}\end{bmatrix} \tag{4.19}$$

i.e. the expression equals the identity matrix, with the 1,1 element set to 0. Therefore r_i-1 singular values evaluate to 1 and one singular value evaluates to 0 and hence.

$$\|\mathbf{J}\|_2=1 \tag{4.20}$$

Obviously equations (4.6) and (4.15) are related to each other in a way that (4.15) is the first order state space representation of the higher order system of differential equations (4.6). Since the linearizing state feedback is conducted based on the transformed system, it is necessary that the transformation is uniquely defined and hence also the inverse of the nonlinear state transformation (4.10) has to exist. This is the case, if $\mathbf{\Phi}_\zeta(\mathbf{x})$ is a diffeomorphism according to Definition B.38.

Remarks:

If $\mathbf{f}$: $\mathcal{X}\rightarrow\mathcal{Y}$, $\mathcal{X},\mathcal{Y}\subset\mathbb{R}^n$,is not uniquely invertible for the whole set $\mathcal{Y}$, it is still possible that it is, at least, a *local diffeomorphism*. I.e. for some $\mathbf{x}_0\in\mathcal{X}$ there is a vicinity $\mathcal{U}$ of $\mathbf{x}_0$ with $\mathbf{f}(\mathcal{U})=\mathcal{V}$ such that $\mathbf{f}:\mathcal{U}\rightarrow\mathcal{V}$ is a bijection. This practically means that $\mathbf{f}$ is a *local diffeomorphism* at $\mathbf{x}_0$ if:

- $\mathbf{f}$ is continuous at $\mathbf{x}_0$

- The Jacobian $\frac{\partial \mathbf{f}}{\partial \mathbf{x}}$ is regular at $\mathbf{x}_0$.

If $\mathcal{X}=\mathcal{Y}=\mathbb{R}^n$, $\mathbf{f}$ is said to be a *global diffeomorphism.* In case $r=n$, the transformation (4.9) already is a local diffeomorphism at $\mathbf{x_0}$. This is because the vector fields of (4.1) are assumed to be sufficiently smooth, and the linear independence of the rows of $\partial\boldsymbol{\Phi}_\zeta/\partial\mathbf{x}$ is assured by the existence of a vector relative degree as shown in the following theorem.

Theorem 4.1 Linear Independence of dΦ

If the system (4.1) has a well-defined vector relative degree $\mathbf{r}^T=\begin{pmatrix} r_1 & \cdots & r_m \end{pmatrix}$ at $\mathbf{x_0}$, the rows of $\frac{\partial \boldsymbol{\Phi}_\zeta}{\partial \mathbf{x}}$ with $\boldsymbol{\Phi}_\zeta(\mathbf{x})$ defined in (4.9) are linearly independent in a vicinity of $\mathbf{x}_0$.

The proof is presented in Appendix F.1. In case $r<n$, the mapping (4.9) is not a bijection. So, additional n-r nonlinear mappings $\boldsymbol{\Phi}_\eta^T(\mathbf{x})=\begin{pmatrix}\Phi_{\eta,1}(\mathbf{x}) & \dots & \Phi_{\eta,n-r}(\mathbf{x})\end{pmatrix}$ have to be found such that

$$\boldsymbol{\Phi}(\mathbf{x})=\begin{pmatrix}\boldsymbol{\Phi}_\zeta(\mathbf{x})\\ \boldsymbol{\Phi}_\eta(\mathbf{x})\end{pmatrix} \tag{4.21}$$

is a local diffeomorphism at $\mathbf{x}_0$. Then

$$\mathbf{z}=\boldsymbol{\Phi}(\mathbf{x}) \tag{4.22}$$

with $\mathbf{z}^T=\begin{pmatrix}\boldsymbol{\zeta}^T & \boldsymbol{\eta}^T\end{pmatrix}$, $\boldsymbol{\eta}^T=\begin{pmatrix}\eta_1 & \cdots & \eta_{n-r}\end{pmatrix}$ is locally invertible at $\mathbf{x_0}$ and, using dynamics (4.1) and transformation (4.22), the system can be transformed to *Byrnes-Isidori normal form*

$$\begin{aligned}\dot{\boldsymbol{\zeta}}&=\mathbf{J}\cdot\boldsymbol{\zeta}+\mathbf{H}\left[\mathbf{a}(\boldsymbol{\zeta},\boldsymbol{\eta})+\mathbf{B}(\boldsymbol{\zeta},\boldsymbol{\eta})\cdot\mathbf{u}\right]\\ \dot{\boldsymbol{\eta}}&=\mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta})+\mathbf{P}(\boldsymbol{\zeta},\boldsymbol{\eta})\cdot\mathbf{u}\end{aligned} \tag{4.23}$$

where, using definitions (4.7):

$$\begin{aligned}&\mathbf{a}(\boldsymbol{\zeta},\boldsymbol{\eta})=\begin{pmatrix}a_1(\boldsymbol{\zeta},\boldsymbol{\eta})\\ \vdots\\ a_m(\boldsymbol{\zeta},\boldsymbol{\eta})\end{pmatrix}=\bar{\mathbf{a}}\left(\boldsymbol{\Phi}^{-1}(\boldsymbol{\zeta},\boldsymbol{\eta})\right)\quad,\quad \mathbf{B}(\boldsymbol{\zeta},\boldsymbol{\eta})=\begin{bmatrix}\mathbf{b}_1^T(\boldsymbol{\zeta},\boldsymbol{\eta})\\ \vdots\\ \mathbf{b}_m^T(\boldsymbol{\zeta},\boldsymbol{\eta})\end{bmatrix}=\bar{\mathbf{B}}\left(\boldsymbol{\Phi}^{-1}(\boldsymbol{\zeta},\boldsymbol{\eta})\right)\\ &q_i(\boldsymbol{\zeta},\boldsymbol{\eta})=L_{\mathbf{f}}\Phi_{\eta,i}\left(\boldsymbol{\Phi}^{-1}(\boldsymbol{\zeta},\boldsymbol{\eta})\right)\quad,\quad \mathbf{p}_i^T(\boldsymbol{\zeta},\boldsymbol{\eta})=L_{\mathbf{G}}\Phi_{\eta,i}\left(\boldsymbol{\Phi}^{-1}(\boldsymbol{\zeta},\boldsymbol{\eta})\right)\ ,\ i=1,\dots,n-r\\ &\mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta})=\begin{pmatrix}q_1(\boldsymbol{\zeta},\boldsymbol{\eta})\\ \vdots\\ q_{n-r}(\boldsymbol{\zeta},\boldsymbol{\eta})\end{pmatrix}\quad,\quad \mathbf{P}(\boldsymbol{\zeta},\boldsymbol{\eta})=\begin{bmatrix}\mathbf{p}_1^T(\boldsymbol{\zeta},\boldsymbol{\eta})\\ \vdots\\ \mathbf{p}_{n-r}^T(\boldsymbol{\zeta},\boldsymbol{\eta})\end{bmatrix}\end{aligned} \tag{4.24}$$

For reasons, explained later, $\boldsymbol{\zeta}$ is called *external states*, while $\boldsymbol{\eta}$ is referred to as *internal states*. The naturally arising question is, whether additional mappings $\boldsymbol{\Phi}_\eta$ exist such that $\boldsymbol{\Phi}$ is a local diffeomorphism and, under which condition can it be guaranteed.

Theorem 4.2 Existence of an Input-Normalized Byrnes-Isidori Normal-Form

If the system (4.1) has a well-defined vector relative degree $\mathbf{r}=\begin{pmatrix} r_1 & \cdots & r_m \end{pmatrix}$, at $\mathbf{x_0}$,

it is always possible to find n-r mappings $\boldsymbol{\Phi}_\eta^T(\mathbf{x})=\begin{pmatrix}\Phi_{\eta,1}(\mathbf{x}) & \cdots & \Phi_{\eta,n-r}(\mathbf{x})\end{pmatrix}$

such that the mapping

$$\boldsymbol{\Phi}(\mathbf{x})=\begin{bmatrix}\boldsymbol{\Phi}_\zeta(\mathbf{x})\\ \boldsymbol{\Phi}_\eta(\mathbf{x})\end{bmatrix}$$

with $\boldsymbol{\Phi}_\zeta(\mathbf{x})$ defined in (4.9)

has a regular Jacobian and is hence a local diffeomorphism.

If additionally the distribution $G(\mathbf{x})=\operatorname{span}\{\mathbf{g}_1 \quad \cdots \quad \mathbf{g}_m\}$ is involutive

in a vicinity of $\mathbf{x}_0$, then $\boldsymbol{\Phi}_\eta(\mathbf{x})$ can be chosen such that

$L_{\mathbf{g}_i}\Phi_{\eta,j}(\mathbf{x})=0$ in a vicinity of $\mathbf{x}_0$, for $i=1,\ldots,m$, $j=1,\ldots,n-r$ and in this case the transformed system is called *input-normalized Byrnes-Isidori normal form*

The proof is presented in Appendix F.2, and the important facts about distributions are presented in Appendix B.3.

Remark

In the SISO case, it is always possible to find a transformation to input-normalized Byrnes-Isidori normal form, since a 1 dimensional distribution is always involutive from the fact, that the Lie product of the only base vector with itself equals zero and therefore belongs to the distribution.

Following Theorem 4.2, if the distribution $G(\mathbf{x})=\operatorname{span}\{\mathbf{g}_1(\mathbf{x}) \quad \cdots \quad \mathbf{g}_m(\mathbf{x})\}$ is involutive then (4.23) becomes

$$\begin{aligned}\dot{\boldsymbol{\zeta}}&=\mathbf{J}\boldsymbol{\zeta}+\mathbf{H}\left[\mathbf{a}(\mathbf{z})+\mathbf{B}(\mathbf{z})\cdot\mathbf{u}\right]\\ \dot{\boldsymbol{\eta}}&=\mathbf{q}(\mathbf{z}).\end{aligned} \tag{4.25}$$

The internal dynamics are independent of the input and the system is therefore also called input normalized Byrnes-Isidori normal form.

Remark

As system (4.1) has a steady state point $\mathbf{x}_S$, some conclusions could be made about the steady state point of the transformed system (4.23).

1. $\mathbf{\Phi}_\zeta(\mathbf{x}_S)=\mathbf{0}$, since these mappings are of the form $\frac{\partial L_\mathbf{f}^k h_i(\mathbf{x})}{\partial \mathbf{x}} \cdot \mathbf{f}(\mathbf{x})$ and $\mathbf{f}(\mathbf{x}_S)=\mathbf{0}$
2. $\mathbf{\Phi}_\eta(\mathbf{x})=\mathbf{0}$ can always be chosen such that $\mathbf{\Phi}_\eta(\mathbf{x}_S)=\mathbf{0}$ since, if one has found such mappings where $\mathbf{\Phi}_\eta(\mathbf{x}_S)=\boldsymbol{\eta}_S$, then $\overline{\mathbf{\Phi}}_\eta(\mathbf{x})=\mathbf{\Phi}_\eta(\mathbf{x})-\boldsymbol{\eta}_S$ is also a valid mapping.

These 2 facts basically imply that the transformed system can always be chosen such that $\mathbf{z}=\mathbf{0}$ is a stationary point.

4.1.2 Linearizing State Feedback

So far preparatory work was done for the actual goal of NDI, namely forcing the nonlinear and coupled system (4.1) to behave like a set of linear and decoupled SISO systems. Therefore new virtual inputs – so-called "pseudo controls", denoted with $\mathbf{v}^T(t)=(\nu_1(t) \quad \cdots \quad \nu_m(t))$- are introduced. Notice that the number of pseudo controls equals the number of outputs m, and it is intended that the i^{th} output is exclusively influenced by the i^{th} pseudo control. The following feedback law

$$\mathbf{u}(t)=\overline{\mathbf{B}}^{-1}(\mathbf{x})[\mathbf{v}(t)-\overline{\mathbf{a}}(\mathbf{x})]=\mathbf{B}^{-1}(\zeta,\boldsymbol{\eta})[\mathbf{v}(t)-\mathbf{a}(\zeta,\boldsymbol{\eta})] \tag{4.26}$$

transforms system (4.23) into

$$\begin{aligned}\dot{\zeta}(t)&=\mathbf{J}\cdot\zeta+\mathbf{H}\cdot\mathbf{v}(t)\\ \dot{\boldsymbol{\eta}}(t)&=\mathbf{q}(\zeta,\boldsymbol{\eta})+\mathbf{P}(\zeta,\boldsymbol{\eta})\cdot\left(\mathbf{B}^{-1}(\zeta,\boldsymbol{\eta})\cdot[\mathbf{v}(t)-\mathbf{a}(\zeta,\boldsymbol{\eta})]\right)\end{aligned} \tag{4.27}$$

or, if the system is in form (4.25) of an input-normalizes Byrnes-Isidori normal form

$$\begin{aligned}\dot{\zeta}&=\mathbf{J}\cdot\zeta+\mathbf{H}\cdot\mathbf{v}\\ \dot{\boldsymbol{\eta}}&=\mathbf{q}(\zeta,\boldsymbol{\eta})\end{aligned} \tag{4.28}$$

Due to the structure of $\mathbf{J}$ and $\mathbf{H}$ it is observed that the coupled nonlinear input/output dynamics, as desired, are transformed into m linear decoupled ones with each r_i poles in the origin. The structure of the resulting system is appropriately analyzed by a scalar expansion of the differential equation.

$$\begin{pmatrix}\dot{\zeta}_{i,1}\\ \vdots\\ \dot{\zeta}_{i,r_i}\end{pmatrix}=\begin{bmatrix}0 & 1 & & 0\\ \vdots & 0 & \ddots & \\ & & \ddots & 1\\ 0 & \cdots & & 0\end{bmatrix}\begin{pmatrix}\zeta_{i,1}\\ \vdots\\ \zeta_{i,r_i}\end{pmatrix}+\begin{pmatrix}0\\ \vdots\\ 0\\ 1\end{pmatrix}\nu_i \tag{4.29}$$

Obviously the states $\zeta_{i,j}$, $j=1,\ldots,r_i$ form a chain of integrators, while the last state $\zeta_{i,ri}$ is exclusively influenced by ν_i. Keeping in mind that $\zeta_{i,j}$ equal to the i^{th} output including its r_i-1 time derivatives, it reveals that the r_i^{th} time derivative of the i^{th} output is exclusively influenced by the i^{th} pseudo control. The output itself is consequently gained by r_i integrations of the corresponding pseudo control. Due to the fact that $\boldsymbol{\zeta}$ represents the system outputs and their respective time derivatives, it is nearby to call them *external*

states. Obviously $\boldsymbol{\eta}$, the second part of the feedback linearized system (4.27) / (4.28) becomes unobservable from the outputs for which reason the associated states are called *internal states*. The input/output relationship of (4.6), after feedback linearization, reads as

$$\begin{pmatrix} y_1^{(r_1)}(t) \\ \vdots \\ y_m^{(r_m)}(t) \end{pmatrix} = \begin{pmatrix} \nu_1(t) \\ \vdots \\ \nu_m(t) \end{pmatrix} \tag{4.30}$$

which again highlights the chain of integrators for each output and the collapse of the nonlinear coupled system into a set of linear decoupled SISO systems. The linearity of the resulting system also allows for a notation in Laplace domain.

$$\begin{pmatrix} y_1(s) \\ y_2(s) \\ \vdots \\ y_m(s) \end{pmatrix} = \begin{bmatrix} \frac{1}{s^{r_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{s^{r_2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{s^{r_m}} \end{bmatrix} \begin{pmatrix} \nu_1(s) \\ \nu_2(s) \\ \\ \nu_m(s) \end{pmatrix} \tag{4.31}$$

It follows from (4.31) that the external dynamics of the feedback linearized system are marginally stable, since the dynamics are linear with poles in the origin of the complex plane. However they can easily be stabilized by feedback of the respective part of the external state $\boldsymbol{\zeta}_i$ or rather of the respective output y_i including its time derivatives up to degree r_i-1 as will be shown later. [Hol04] developed a nice interpretation the linearizing state feedback as shown in Figure 4.1. There $\mathbf{F}(\mathbf{x},\mathbf{u})$ abstractly represents the system dynamics while $\mathbf{F}^{-1}(\mathbf{x},\mathbf{u})$ represents the inverse dynamics, which is actually the linearizing feedback (4.26). System dynamics in series with the inverse system dynamics result in an identity transfer behavior indicated by "1". The output of this serial connection equal the r_i^{th} time derivative $y_i^{(r_i)}$ and in order to obtain $y_i(t)$, $y_i^{(r_i)}$ has to be sent through a chain of r_i integrators.

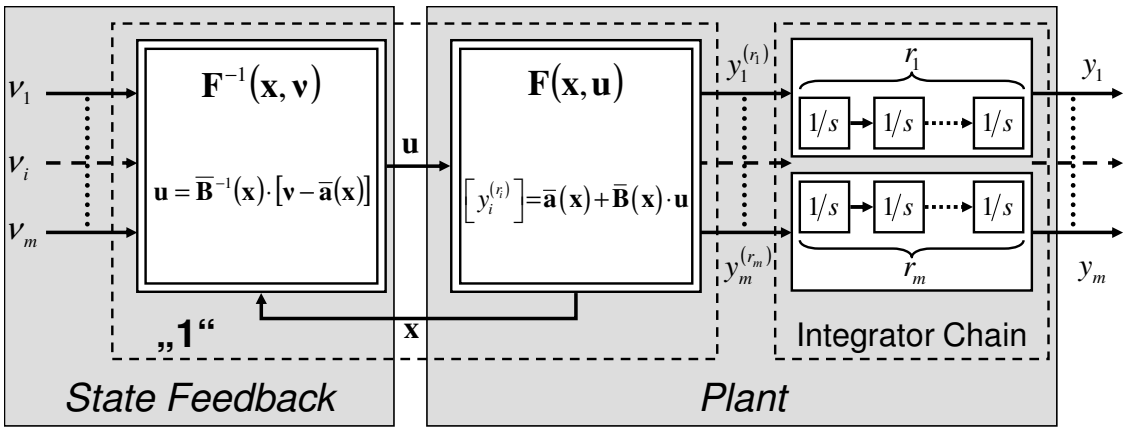


Figure 4.1 Graphical Interpretation of the Feedback Linearized System [Hol04]

In order to apply feedback linearization to real systems successfully, also the internal dynamics have to be stable since, otherwise, the feedback law (4.26) might grow unbounded. For this reason a closer look is taken at the internal dynamics in the following.

4.1.3 Stabilization and Internal Dynamics

The last section introduced the linearizing state feedback such that the dynamics of the external states collapse to set of m marginally stable integrator chains, which can be stabilized without difficulty. The dynamics of the internal states, however, are irremediable once the linearizing feedback is applied. But for success of the control task of real physical systems with limited control capacity it is inevitable that the states of the internal dynamics are at least bounded, since otherwise boundedness of the linearizing feedback (4.26) cannot be guaranteed, which lets the actuators of a real physical system run in saturation. At first, there is an interesting interpretation of the internal dynamics. Therefore consider the internal dynamics of (4.28) and the linearizing feedback (4.26) as dynamic system

$$\begin{aligned}\dot{\boldsymbol{\eta}}(t) &= \mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta})\\ \mathbf{u}(t) &= \mathbf{B}^{-1}(\boldsymbol{\zeta},\boldsymbol{\eta})\left[\mathbf{v}(t)-\mathbf{a}(\boldsymbol{\zeta},\boldsymbol{\eta})\right]\end{aligned} \tag{4.32}$$

where $\boldsymbol{\zeta}(t)$ is the input, $\boldsymbol{\eta}(t)$ is the state and $\mathbf{u}(t)$ is the output. Then, since $\boldsymbol{\zeta}$ and $\mathbf{u}$ represent the output including its time derivatives and input of the original system (4.1) respectively, (4.32) in a certain sense “inverts” the dynamic as output and input are interchanged. Therefore, (4.32) is also called *inverse dynamic*. In other words, if the external states are forced to follow some reference signal $\boldsymbol{\zeta}_R(t)$

$$\boldsymbol{\zeta}_R(t)=\begin{pmatrix}\zeta_{1,R}(t)\\ \vdots\\ \zeta_{m,R}(t)\end{pmatrix},\ \boldsymbol{\zeta}_{i,R}(t)=\begin{pmatrix}y_{i,R}(t)\\ \vdots\\ y_{i,R}^{(r_i-1)}(t)\end{pmatrix} \tag{4.33}$$

the internal dynamics provides the required plant input to achieve this objective. In order for the objective to be feasible at all, the following postulates have to be fulfilled.

1. The initial conditions have to be compliant, i.e. $\boldsymbol{\zeta}(0)=\boldsymbol{\zeta}_R(0)$ whereas $\boldsymbol{\eta}(0)$ can be chosen arbitrarily.
2. Each pseudo control equals to the r_i^{th} time derivative of the respective output, i.e. $\mathbf{v}(t)=\mathbf{v}_R(t)$ where $\mathbf{v}_R^T(t)=\begin{pmatrix}y_1^{(r_1)}(t) & \cdots & y_m^{(r_m)}(t)\end{pmatrix}$

If the conditions are fulfilled, the unique plant input that achieves $\boldsymbol{\zeta}(t)=\boldsymbol{\zeta}_R(t)$ is given by the following differential equations.

$$\begin{aligned}\dot{\boldsymbol{\eta}}_R(t) &= \mathbf{q}(\boldsymbol{\zeta}_R,\boldsymbol{\eta}_R)\\ \mathbf{u}(t) &= \mathbf{B}^{-1}(\boldsymbol{\zeta}_R,\boldsymbol{\eta}_R)\left[\mathbf{v}_R(t)-\mathbf{a}(\boldsymbol{\zeta}_R,\boldsymbol{\eta}_R)\right]\end{aligned} \tag{4.34}$$

A special case for the inverse dynamic is given when the output should remain zero for all times. The resulting dynamics which computes the plant input that achieves the goal, provided the initial conditions are compliant, is also called *zero dynamics*.

$$\dot{\boldsymbol{\eta}}_0(t) = \mathbf{q}(\mathbf{0}, \boldsymbol{\eta}_0) \tag{4.35}$$

The zero dynamics play an important role in stability proofs for the internal dynamics. In order to understand the progressive arguments, it is helpful to keep in mind the analogy of internal dynamics to linear systems, as developed in [Isi95]. Let us consider a linear SISO system that is described by a transfer function.

$$y(s) = k \frac{b_0 + b_1 s + \ldots + b_{n-r-1} s^{n-r-1} + s^{n-r}}{a_0 + a_1 s + \ldots + a_{n-1} s^{n-1} + s^n} u(s)$$

An associated state space representation in controllable canonical form is obtained by

$$\dot{\mathbf{x}}(t) = \mathbf{A} \cdot \mathbf{x} + \mathbf{b} \cdot u$$
$$y(t) = \mathbf{c}^T \cdot \mathbf{x}$$

where $\mathbf{A} = \begin{bmatrix} 0 & 1 & & \mathbf{0} \\ 0 & 0 & \ddots & \\ \vdots & \vdots & \ddots & 1 \\ -a_0 & -a_1 & \cdots & -a_{n-1} \end{bmatrix}$, $\mathbf{b} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ k \end{pmatrix}$, $\mathbf{c}^T = \begin{pmatrix} b_0 & \cdots & b_{n-r-1} & 1 & 0 & \cdots & 0 \end{pmatrix}$

Note that the relative degree of the transfer function, namely the difference in the degree of denominator and numerator polynomial is r. Conducting the transformation, described in the foregoing section, the external states are

$$\begin{aligned} \zeta_1 &= \mathbf{c}^T \mathbf{x} && = b_0 x_1 + b_1 x_2 + \ldots + b_{n-r-1} x_{n-r} + x_{n-r+1} \\ \zeta_2 &= \mathbf{c}^T \mathbf{A} \mathbf{x} && = b_0 x_2 + b_1 x_3 + \ldots + b_{n-r-1} x_{n-r+1} + x_{n-r+2} \\ &\vdots \\ \zeta_r &= \mathbf{c}^T \mathbf{A}^{r-1} \mathbf{x} && = b_0 x_r + b_1 x_{r+1} + \ldots + b_{n-r-1} x_{n-1} + x_n \end{aligned}$$

The internal states are appropriately chosen to

$$\begin{aligned} \eta_1 &= x_1 \\ \eta_2 &= x_2 \\ &\vdots \\ \eta_{n-r} &= x_{n-r}. \end{aligned}$$

and render the map

$$\begin{pmatrix} \zeta_1 \\ \vdots \\ \vdots \\ \zeta_r \\ \eta_1 \\ \vdots \\ \eta_{n-r} \end{pmatrix} = \begin{bmatrix} b_0 & \cdots & b_{n-r-1} & 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & b_0 & \cdots & b_{n-r-1} & 1 \\ 1 & & \mathbf{0} & 0 & \cdots & \cdots & 0 \\ & \ddots & & \vdots & & & \vdots \\ \mathbf{0} & & 1 & 0 & \cdots & \cdots & 0 \end{bmatrix} \cdot \begin{pmatrix} x_1 \\ \\ \\ \vdots \\ \\ \\ x_n \end{pmatrix}$$

a diffeomorphism, as the absolute value of the determinant of the involved matrix evaluates to 1, which can easily be verified by applying Laplace's formula for computation of the determinant, starting with the very right column.

Consider the dynamics of the external states

$$\begin{aligned} \dot{\zeta}_1 &= \mathbf{c}^T\mathbf{A}\mathbf{x} && = \zeta_2 \\ &\vdots \\ \dot{\zeta}_{r-1} &= \mathbf{c}^T\mathbf{A}^{r-1}\mathbf{x} && = \zeta_r \\ \dot{\zeta}_r &= \mathbf{c}^T\mathbf{A}^r\mathbf{x} + \mathbf{c}^T\mathbf{A}^{r-1}\mathbf{b}\cdot u \end{aligned}$$

which reads in matrix-vector notation as (* indicates a not further specified number)

$$\begin{pmatrix} \dot{\zeta}_1 \\ \vdots \\ \dot{\zeta}_r \end{pmatrix} = \begin{bmatrix} 0 & 1 & & \mathbf{0} \\ \vdots & 0 & \ddots & \\ 0 & & \ddots & 1 \\ * & * & \cdots & * \end{bmatrix} \cdot \begin{pmatrix} \zeta_1 \\ \vdots \\ \zeta_r \end{pmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ * & * & \cdots & * \end{bmatrix} \cdot \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_{n-r} \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ k \end{pmatrix} u \,.$$

Considering the special structure of $\mathbf{A}$ and the transformation rule for ζ_1, the internal dynamics become

$$\dot{\eta}_1 \quad = \dot{x}_1 \quad = x_2 \qquad = \eta_2$$

$$\begin{aligned} \dot{\eta}_{n-r-1} &= \dot{x}_{n-r-1} = x_{n-r} && = \eta_{n-r} \\ \dot{\eta}_{n-r} &= \dot{x}_{n-r} = x_{n-r+1} && = \zeta_1 - b_0 x_1 - \ldots - b_{n-r-1} x_{n-r} = \zeta_1 - b_0\eta_1 - \ldots - b_{n-r-1}\eta_{n-r} \end{aligned}$$

In matrix vector notation the internal dynamics become

$$\begin{pmatrix} \dot{\eta}_1 \\ \dot{\eta}_2 \\ \\ \dot{\eta}_{n-r} \end{pmatrix} = \begin{bmatrix} 0 & 1 & & \mathbf{0} \\ \vdots & 0 & \ddots & \\ 0 & & \ddots & 1 \\ -b_0 & -b_1 & \cdots & -b_{n-r-1} \end{bmatrix} \cdot \begin{pmatrix} \eta_1 \\ \eta_2 \\ \\ \eta_{n-r} \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \zeta_1$$

It reveals that the internal dynamics are in controllable canonical form and the coefficients in the last row of the system matrix are hence the coefficients of the characteristic polynomial. On the other hand, these coefficients are the same as in the numerator polynomial of the transfer function of the original system.

Summing up so far, in the linear case the internal dynamics are stable if the system is minimum phase, i.e. the associated transfer function has stable zeros. It is thus manifest to extend the concept of minimum phase to nonlinear systems according to the following definition ([Byr91], [Sva06]).

Definition 4.2 Minimum Phase System

If the system (4.1) has a well-defined vector relative degree, it is said to be a *minimum phase system*, if the equilibrium $\boldsymbol{\eta}_0 = 0$ of its zero dynamics (4.35) is locally asymptotically stable.

Remark

The term minimum phase system was first introduced by Byrnes and Isidori [Byr91] in a different statement that involves results from differential geometry. The original definition deals with a so-called *zero dynamics submanifold*, which is basically the set of original coordinates $\mathbf{x}$ such that the external states are zero, $\left\{\mathbf{x} \in \mathbb{R}^n \,\middle|\, \boldsymbol{\Phi}_\zeta(\mathbf{x}) = \mathbf{0}\right\}$ (which equivalently means that the output variables including their time derivatives are zero). Following the definition of [Byr91] (spoken in differential geometric language), if the system is minimum phase then there is some static feedback $\mathbf{u}^*(\mathbf{x})$ such that, if $\mathbf{x}(t)$ is on the zero dynamics submanifold then $\dot{\mathbf{x}}(t)$, i.e. $\mathbf{f}(\mathbf{x})+\mathbf{G}(\mathbf{x})\mathbf{u}^*(\mathbf{x})$, is tangent to it and $\mathbf{u}^*(\mathbf{x})$ renders the closed loop dynamics

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}) + \mathbf{G}(\mathbf{x})\mathbf{u}^*(\mathbf{x})$$

asymptotically stable. Roughly spoken there is some static feedback $\mathbf{u}^*(\mathbf{x})$ such that the closed loop system has a zero output $\mathbf{y}(t)=\mathbf{0}$ for $t \geq t_0$ and is asymptotically stable (for both external and internal states). The tangent property for $\dot{\mathbf{x}}(t)$ thereby simply means that $\mathbf{x}(t)$ remains on the zero dynamics manifold just as some trajectory on a surface. If the motion of some particle is tangent to a surface (imagine e.g. a sphere) then it will stay on the surface in the future. The definition in here is, however, taken from [Sva06]. But it turns out that the definition used here is implied by one given by Byrnes, since the existence of a relative degree assures the existence of an input, which forces the external states to remain at zero. If, additionally, the internal dynamics asymptotically approaches its equilibrium, the system in original coordinates also approaches its equilibrium $\mathbf{x}_S(t)$.

Minimum phase systems guarantee stability of the zero dynamics. It guarantees thus the stability of the overall system in the case where the output is identically kept at zero. But for a general external state, still no statement about the stability for the overall system can be made. The sticking point henceforth is to determine conditions under which the internal dynamics and with it the overall dynamics are stable.

Before the overall dynamics (4.28) have a chance to become stable, it is necessary that the external dynamics are stable. These can be stabilized by defining for the pseudo controls

$$\nu_i = -c_{i,0} \cdot \zeta_{i,1} - \ldots - c_{i,,r_i-2} \cdot \zeta_{i,r_i-1} - c_{i,,r_i-1} \cdot \zeta_{i,r_i} \tag{4.36}$$

such that, for $i = 1,\ldots,m$ the polynomials

$$s^{r_i} + c_{i,r_i-1} s^{r_i-1} \ldots + c_{i,1} s + c_{i,0}$$

have its roots in the open left half complex plane (l.h.c.p.). The dynamics (4.28) are

$$\begin{aligned} \dot{\boldsymbol{\zeta}} &= \left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right) \cdot \boldsymbol{\zeta} \\ \dot{\boldsymbol{\eta}} &= \mathbf{q}(\boldsymbol{\zeta}, \boldsymbol{\eta}) \end{aligned} \tag{4.37}$$

where $\mathbf{C}$ is a matrix that has the coefficients of the i^{th} polynomial stacked into the i^{th} row:

$$\mathbf{C}^T = \begin{bmatrix} \mathbf{c}_1^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{c}_2^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{c}_m^T \end{bmatrix} \tag{4.38}$$

where $\mathbf{c}_i^T = \begin{pmatrix} c_{i,0} & \cdots & c_{i,r_i-1} \end{pmatrix}$. Moreover the dynamics of $\boldsymbol{\zeta}_i$ – see (4.11) – are still decoupled from each other for $i = 1,\ldots,m$ and read as

$$\begin{pmatrix} \dot{\zeta}_{i,1} \\ \dot{\zeta}_{i,2} \\ \vdots \\ \dot{\zeta}_{i,r_i} \end{pmatrix} = \begin{bmatrix} 0 & 1 & & \mathbf{0} \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 \\ c_{i,0} & c_{i,1} & \cdots & c_{i,r_i-1} \end{bmatrix} \cdot \begin{pmatrix} \zeta_{i,1} \\ \zeta_{i,2} \\ \vdots \\ \zeta_{i,r_i} \end{pmatrix} \tag{4.39}$$

Equation (4.39) is obviously in controllable canonical form and the scalars of $\mathbf{c}_i$ are the coefficients of the characteristic polynomial, which was chosen to be stable. Hence the external dynamics is asymptotically stable. With the external dynamics stabilized, [Kha02] gives a lemma for stability of internal dynamics, which states that the internal dynamics are locally stable if the system is minimum phase. The proof is presented in Appendix F.3.

Lemma 4.1 Stability of Internal Dynamics

Assume that system (4.28)

$$\begin{aligned} \dot{\boldsymbol{\zeta}}(t) &= \left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right) \cdot \boldsymbol{\zeta}(t) \\ \dot{\boldsymbol{\eta}}(t) &= \mathbf{q}(\boldsymbol{\zeta}(t), \boldsymbol{\eta}(t)) \end{aligned}$$

where $\boldsymbol{\zeta}(t) \in \mathbb{R}^r, \boldsymbol{\eta}(t) \in \mathbb{R}^{n-r}, \mathbf{z}^T(t) = \begin{pmatrix} \boldsymbol{\zeta}^T(t) & \boldsymbol{\eta}^T(t) \end{pmatrix}$

has an equilibrium at $\mathbf{z} = \mathbf{0}$, , $\mathbf{J} - \mathbf{H}\mathbf{C}^T$ is Hurwitz and $\mathbf{q}(\boldsymbol{\zeta}, \boldsymbol{\eta})$ is continuously

differentiable on $\overline{\mathcal{B}_r} = \left\{ \mathbf{z} \in \mathbb{R}^n \;\middle|\; \|\mathbf{z}\| \le r \right\}$ for some $r>0$.

The equilibrium is (locally) asymptotically stable,
if the associated zero dynamics (4.35)

$$\dot{\boldsymbol{\eta}}_0 = \mathbf{q}(\mathbf{0}, \boldsymbol{\eta}_0)$$

are asymptotically stable

4.1.4 Plant-Model Deviations

Model Uncertainty

For the algorithm, derived so far, it is assumed that the plant dynamics are known exactly. This, however, is not the case for most applications. Particularly in flight control, identification of model parameters is a very time-consuming procedure nonetheless it is not possible to determine the dynamic model exactly. In some cases, the linearizing state feedback is simplified even consciously e.g. by neglecting some minor influences depending on parameters which could only be identified with high effort. Further sources of uncertainties are disturbances such as wind gust or inexact and noisy measurements. All these uncertainties are incorporated into analysis by defining assumed dynamics, marked with a "^"

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \hat{\mathbf{f}}(\mathbf{x}) + \hat{\mathbf{G}}(\mathbf{x})\mathbf{u} \\ \mathbf{y}(t) &= \hat{\mathbf{h}}(\mathbf{x}) \end{aligned} \tag{4.40}$$

which differ from the real dynamics (4.1) and which are used for feedback linearization. Conducting the nonlinear state transformation derived in section 4.1.1, we arrive at the Byrnes-Isidori normal form for the assumed system

$$\begin{aligned} \dot{\zeta}(t) &= \mathbf{J} \cdot \zeta + \mathbf{H} \cdot \left[\hat{\mathbf{a}}(\zeta, \boldsymbol{\eta}) + \hat{\mathbf{B}}(\zeta, \boldsymbol{\eta}) \cdot \mathbf{u} \right] \\ \dot{\boldsymbol{\eta}}(t) &= \hat{\mathbf{q}}(\zeta, \boldsymbol{\eta}) + \hat{\mathbf{P}}(\zeta, \boldsymbol{\eta})\mathbf{u} \end{aligned} \tag{4.41}$$

where $\hat{\mathbf{a}}(\zeta, \boldsymbol{\eta})$, $\hat{\mathbf{B}}(\zeta, \boldsymbol{\eta})$, $\hat{\mathbf{q}}(\zeta, \boldsymbol{\eta}), \hat{\mathbf{P}}(\zeta, \boldsymbol{\eta})$ are defined analogously to (4.24).

Actuators

Another source of uncertainty that occurs in reality is unmodeled dynamics. These are incorporated into the analysis by defining

$$\mathbf{u} = \mathbf{G}_A(\mathbf{u}_c, \zeta, \boldsymbol{\eta}) \tag{4.42}$$

where $\mathbf{u}_c$ is the command, obtained from the linearizing state feedback (4.26) and $\mathbf{u}$ is the actuator state, acting onto the plant. $\mathbf{G}_A(\mathbf{u}_c, \zeta, \boldsymbol{\eta})$ is thereby an abstract representation for some static (e.g. saturation) or dynamic relationship, which could be linear or nonlinear.

We assumed that the actuator dynamics also depend on the rigid body states. In aircraft application, of course, the aerodynamic forces and moments influence the actuator dynamics due to loads, imposed on the control surfaces. Taking into account model uncertainties as well as actuator dynamics, the linearizing state feedback (4.26) changes to:

$$\mathbf{u}_c = \hat{\mathbf{B}}^{-1}(\zeta,\boldsymbol{\eta})(\boldsymbol{\nu} - \hat{\mathbf{a}}(\zeta,\boldsymbol{\eta})) \tag{4.43}$$

Applying the linearizing state feedback (4.43) to the real dynamics yields

$$\begin{aligned}
\dot{\zeta} &= \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{a}(\zeta,\boldsymbol{\eta}) + \mathbf{B}(\zeta,\boldsymbol{\eta})\mathbf{u}\right] \\
&= \mathbf{J}\zeta + \mathbf{H}\left[\overset{+}{\hat{\mathbf{a}}(\zeta,\boldsymbol{\eta}) + \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u}} \overset{-}{- \hat{\mathbf{a}}(\zeta,\boldsymbol{\eta}) - \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u}} + \mathbf{a}(\zeta,\boldsymbol{\eta}) + \mathbf{B}(\zeta,\boldsymbol{\eta})\mathbf{u}\right] \\
&= \mathbf{J}\zeta + \mathbf{H}\left[\hat{\mathbf{a}}(\zeta,\boldsymbol{\eta}) + \overset{+}{\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u}_C} - \overset{-}{\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u}_C} + \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u} - \hat{\mathbf{a}}(\zeta,\boldsymbol{\eta}) - \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u} + \mathbf{a}(\zeta,\boldsymbol{\eta}) + \mathbf{B}(\zeta,\boldsymbol{\eta})\mathbf{u}\right] \\
&= \mathbf{J}\zeta + \mathbf{H}\left[\overbrace{\hat{\mathbf{a}}(\zeta,\boldsymbol{\eta}) + \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u}_C}^{\mathbf{y}} - \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\overbrace{(\mathbf{u}_C - \mathbf{u})}^{\Delta\mathbf{u}} - \overbrace{(\hat{\mathbf{a}}(\zeta,\boldsymbol{\eta}) - \mathbf{a}(\zeta,\boldsymbol{\eta}))}^{\tilde{\mathbf{a}}(\zeta,\boldsymbol{\eta})} - \overbrace{(\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta}) - \mathbf{B}(\zeta,\boldsymbol{\eta}))\mathbf{u}}^{\tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta})}\right].
\end{aligned}$$

The modeling error is defined as

$$\tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{u}) = \tilde{\mathbf{a}}(\zeta,\boldsymbol{\eta}) + \tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u} \tag{4.44}$$

and the control deficiency is

$$\Delta\mathbf{u} = \mathbf{u}_C - \mathbf{u}$$

With linearizing state feedback (4.43) and internal dynamics of (4.27) we arrive at

$$\begin{aligned}
\dot{\zeta} &= \mathbf{J}\zeta + \mathbf{H}\left[\boldsymbol{\nu} - \tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{u}) - \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u}\right] \\
\dot{\boldsymbol{\eta}} &= \mathbf{q}(\zeta,\boldsymbol{\eta}) + \mathbf{P}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^{-1}(\zeta,\boldsymbol{\eta})[\boldsymbol{\nu} - \hat{\mathbf{a}}(\zeta,\boldsymbol{\eta})] - \mathbf{P}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u}
\end{aligned} \tag{4.45}$$

Obviously, the internal dynamics (if not input-normalized) are excited by the linearizing state feedback and the control deficiency. The resulting external dynamics can be interpreted graphically ([Hol04]) as depicted in Figure 4.2. There are m decoupled systems, where each pseudo control is sent through a chain of r_i integrators, which is represented by $\mathbf{J}$ in (4.45). Additionally, the pseudo controls are disturbed by the augmented disturbance, defined as

$$\boldsymbol{\Delta}(\zeta,\boldsymbol{\eta},\mathbf{u},\Delta\mathbf{u}) = \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\cdot\Delta\mathbf{u} + \tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{u}) \tag{4.46}$$

Although the system is in fact intended to be decoupled by the linearizing feedback, these disturbances again couple the system, since each $\Delta_i(\zeta,\boldsymbol{\eta},\mathbf{u},\Delta\mathbf{u})$ is dependent on the whole state vector.

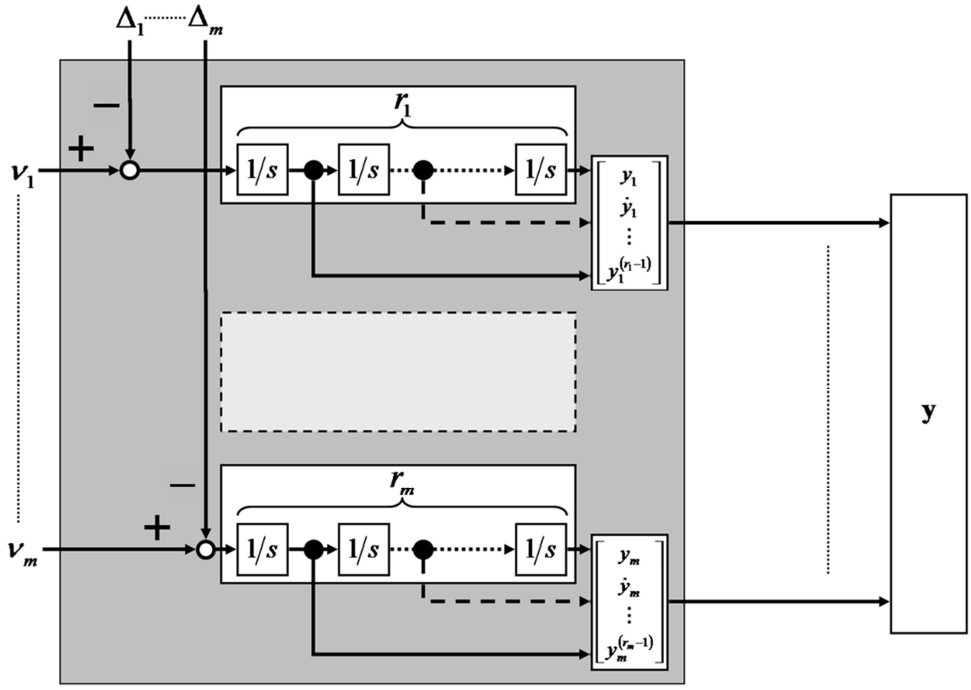


Figure 4.2 Feedback Linearized System with Augmented Disturbance [Hol04]

4.1.5 Tracking Control

For real life applications it is desirable that the output of the system $\mathbf{y}(t)$ tracks a reference trajectory. As indicated in (4.33), it is necessary that, for each scalar output $y_i(t)$, the time derivatives up to degree r_i exist and are known.

There are various ways to generate a sufficiently smooth trajectory including their time derivatives up to degree r_i. E.g. in case of controlling the flight path trajectory of an aircraft, the variable to be controlled is the position in space given by 3 independent coordinates. The trajectory usually is generated by some a flight management system, which delivers a trajectory e.g. in form of splines. In this case the time derivatives are obtained by analytical differentiation of the splines and generally the resulting dynamics are nonlinear for which reason it is called a nonlinear reference model. In other cases however it is sufficient to define a linear reference model given by a linear time invariant differential equation of order r_i.

$$y_{i,R}^{(r_i)} = -a_{i,r_i-1} \cdot y_{i,R}^{(r_i-1)} - \ldots - a_{i,1} \cdot \dot{y}_{i,R} - a_{i,0} \cdot y_{i,R} + a_{i,0} \cdot y_{i,c} \tag{4.47}$$

where $y_{i,R}(t)$ is the reference variable for the i^{th} system output and $y_{i,c}(t)$ the external command. Linear reference models are preferable since their dynamic behavior is easily predictable, scalable for inputs of different amplitudes and there are miscellaneous tools available for analysis of linear systems. Figure 4.3 illustrates, how to implement linear reference models and how to obtain the time derivatives up to degree r_i. Of course the $a_{i,k}$, k=1,...r_{i-1}, have to be chosen such that the reference dynamics are stable, i.e. the roots of the polynomial

$$s^{r_i} + a_{i,r_i-1} \cdot s^{r_i-1} + \ldots + a_{i,1} \cdot s + a_{i,0}$$

lie in the open l.h.c.p.. Note also that the coefficient of $y_{i,c}$ is chosen to be $a_{i,0}$ for steady state accuracy.

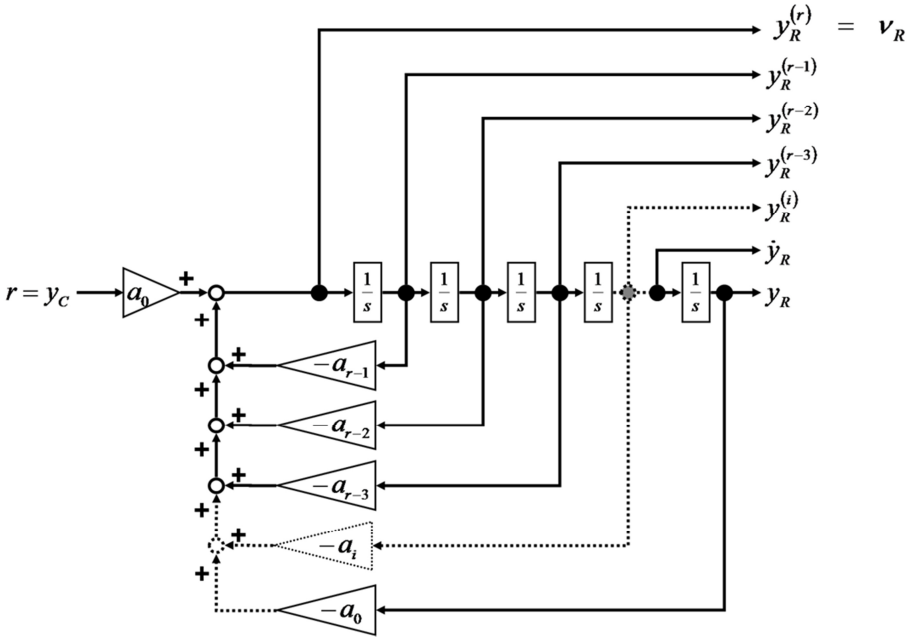


Figure 4.3 Linear Reference Model [Hol08]

Equation (4.47) is a linear ordinary differential equation of order r_i and can equivalently be written as state space model of r_i first order differential equations (e.g. [Wol74]) by defining the states

$$\begin{aligned} \zeta_{i,R,1} &= y_{i,R} \\ \zeta_{i,R2} &= \dot{y}_{i,R} \\ &\vdots \\ \zeta_{i,R,r_i} &= \dot{y}_{i,R}^{(r_i-1)} \end{aligned}$$

and the state vector $\zeta_{i,R}^T = \begin{pmatrix} y_{i,R} & \dot{y}_{i,R} & \cdots & y_{i,R}^{(r_i-1)} \end{pmatrix}$. Then the reference dynamics are

$$\begin{pmatrix} \dot{\zeta}_{i,R,1} \\ \dot{\zeta}_{i,R,2} \\ \vdots \\ \dot{\zeta}_{i,R,r_i} \end{pmatrix} = \begin{bmatrix} 0 & 1 & & \mathbf{0} \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 \\ -a_{i,0} & -a_{i,1} & \cdots & -a_{i,r_i-1} \end{bmatrix} \cdot \begin{pmatrix} \zeta_{i,R,1} \\ \zeta_{i,R,2} \\ \vdots \\ \zeta_{i,R,r_i} \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{i,0} \end{pmatrix} y_{i,c} \qquad (4.48)$$

The system matrix has a similar structure as $\mathbf{J}_i$, defined in (4.16). By defining

$$\mathbf{k}_i^T = \begin{pmatrix} a_{i,0} & a_{i,1} & \cdots & a_{i,r_i-1} \end{pmatrix}$$

the state space model is also written as

$$\dot{\zeta}_{i,R} = \left(\mathbf{J}_i - \mathbf{h}_i \mathbf{k}_i^T \right) \cdot \zeta_{i,R} + a_0 \mathbf{h}_i \cdot y_c$$

where $\mathbf{h}_i$ is defined in (4.16). Collecting all reference states into a single state space model yields

$$\begin{pmatrix} \dot{\zeta}_{1,R} \\ \dot{\zeta}_{2,R} \\ \\ \dot{\zeta}_{m,R} \end{pmatrix} = \left(\begin{bmatrix} \mathbf{J}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{J}_m \end{bmatrix} - \begin{bmatrix} \mathbf{h}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_2 & \ddots & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{h}_m \end{bmatrix} \cdot \begin{bmatrix} \mathbf{k}_1^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{k}_2^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{k}_m^T \end{bmatrix} \right) \cdot \begin{pmatrix} \zeta_{1,R} \\ \zeta_{2,R} \\ \vdots \\ \zeta_{m,R} \end{pmatrix}$$
$$+ \begin{bmatrix} \mathbf{h}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{h}_2 & \ddots & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{h}_m \end{bmatrix} \cdot \begin{bmatrix} a_{1,0} & 0 & \cdots & 0 \\ 0 & a_{2,0} & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & a_{m,0} \end{bmatrix} \cdot \begin{pmatrix} y_{1,c} \\ y_{2,c} \\ \vdots \\ y_{m,c} \end{pmatrix}$$

With definition of

$$\mathbf{K}^T = \begin{bmatrix} \mathbf{k}_1^T & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{k}_2^T & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{k}_m^T \end{bmatrix} \quad \mathbf{A}_0 = \operatorname{diag}\left(a_{1,0},\ldots,a_{m,0}\right)$$

$\mathbf{H}$ and $\mathbf{J}$ in (4.16), the state space model, containing the reference dynamics of all outputs, is conveniently written in the compact form

$$\dot{\zeta}_R = \left(\mathbf{J} - \mathbf{H}\cdot\mathbf{K}^T\right)\cdot\zeta_R + \mathbf{H}\cdot\mathbf{A}_0\cdot\mathbf{y}_c \tag{4.49}$$

where

$$\zeta_R^T = \begin{pmatrix} \zeta_{1,R}^T & \cdots & \zeta_{m,R}^T \end{pmatrix} \;,\; \mathbf{y}_c^T = \begin{pmatrix} y_{1,c} & \cdots & y_{m,c} \end{pmatrix}$$

are denoted as *reference state vector* and *reference command vector* respectively. As explained in section 4.1.3, in order to track a desired reference trajectory - in case that the assumed model perfectly matches reality, there are no neglected dynamics and initial conditions are compliant – it is sufficient to apply the r_i^{th} derivative of each reference trajectory to the associated pseudo control. However, due to various deviations introduced in 4.1.4, the suggested tracking algorithm will fail and results in an error between reference model and plant. In order to stabilize the error dynamics, additionally the errors between reference model and system output including their time derivatives up to degree r_i-1

$$e_i^{(k)} = y_{i,R}^{(k)} - y_i^{(k)} = \zeta_{i,R,k+1} - \zeta_{i,k+1} \tag{4.50}$$

k=0,1,...,r_i-1 have to be fed back to the pseudo control. In case of tracking, each pseudo control is of the following form

$$\nu_i = \nu_{i,R} + \nu_{i,E} \tag{4.51}$$

where $\nu_{i,R} := y_{i,R}^{(r_i)} = \dot{\zeta}_{i,r_i}$ is the *reference pseudo control* and

$$\nu_{i,E} = c_{r_i-1} \cdot e_i^{(r_i-1)} + \ldots + c_1 \cdot \dot{e}_i + c_0 \cdot e_i \tag{4.52}$$

is the stabilizing error feedback. It will be shown in a few lines that the coefficients $c_{i,k}$ can be chosen such that the error dynamics are stable. In order to derive the error dynamics, at first, define the error vectors associated with a single system output y_i as

$$\mathbf{e}_i^T = \begin{pmatrix} e_i & \dot{e}_i & \cdots & e_i^{(r_i-1)} \end{pmatrix}$$

and a compound error that stacks all errors into a vector:

$$\mathbf{e}^T = \begin{pmatrix} \mathbf{e}_1^T & \cdots & \mathbf{e}_m^T \end{pmatrix} = \boldsymbol{\zeta}_R - \boldsymbol{\zeta}$$

The pseudo control vector becomes according to (4.51)

$$\boldsymbol{\nu} = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_m \end{pmatrix} = \begin{pmatrix} \nu_{1,R} \\ \nu_{2,R} \\ \vdots \\ \nu_{m,R} \end{pmatrix} + \begin{pmatrix} \nu_{1,E} \\ \nu_{2,E} \\ \vdots \\ \nu_{m,E} \end{pmatrix} =: \boldsymbol{\nu}_R + \boldsymbol{\nu}_E \tag{4.53}$$

where the stabilizing error feedback $\boldsymbol{\nu}_E$, by consideration of (4.52), is

$$\boldsymbol{\nu}_E = \mathbf{C}^T \cdot \mathbf{e} \tag{4.54}$$

and $\mathbf{C}$ is defined in (4.38). Moreover, $\nu_{i,R}$ can also be written in terms of the reference state vector $\boldsymbol{\zeta}_R$. Therefore, note that the last line in (4.48) equals $\nu_{i,R}$ and hence

$$\nu_{i,R} = -\mathbf{k}_i^T \cdot \boldsymbol{\zeta}_{i,R} + a_{i,0} y_{i,c} .$$

The whole vector is

$$\boldsymbol{\nu}_R = -\mathbf{K}^T \cdot \boldsymbol{\zeta}_R + \mathbf{A}_0 \mathbf{y}_c \tag{4.55}$$

and allows for an alternative notation of the reference dynamics (4.49).

$$\dot{\boldsymbol{\zeta}}_R = \mathbf{J}\boldsymbol{\zeta}_R + \mathbf{H}\boldsymbol{\nu}_R \tag{4.56}$$

Applying the pseudo control (4.53) under consideration of (4.54) to the external dynamics (4.45) yields

$$\dot{\boldsymbol{\zeta}} = \mathbf{J} \cdot \boldsymbol{\zeta} + \mathbf{H} \cdot \left[\boldsymbol{\nu}_R + \mathbf{C}^T \cdot \mathbf{e} - \boldsymbol{\Delta}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}, \Delta\mathbf{u}) \right] \tag{4.57}$$

With the definition of the control error (4.49) and (4.56), (4.57) the error dynamics become

$$\begin{aligned} \dot{\mathbf{e}} &= \dot{\boldsymbol{\zeta}}_R - \dot{\boldsymbol{\zeta}} \\ &= \mathbf{J}\boldsymbol{\zeta}_R + \mathbf{H}\boldsymbol{\nu}_R - \mathbf{J}\boldsymbol{\zeta} - \mathbf{H}\left[\boldsymbol{\nu}_R + \mathbf{C}^T\mathbf{e} - \boldsymbol{\Delta}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}, \Delta\mathbf{u}) \right] \\ \dot{\mathbf{e}} &= \left(\mathbf{J} - \mathbf{H}\mathbf{C}^T \right) \cdot \mathbf{e} + \mathbf{H} \cdot \boldsymbol{\Delta}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}, \Delta\mathbf{u}) \end{aligned} \tag{4.58}$$

and the internal dynamics are

$$\dot{\boldsymbol{\eta}} = \mathbf{q}(\boldsymbol{\zeta}_R + \mathbf{e}, \boldsymbol{\eta}) + \mathbf{P}(\boldsymbol{\zeta}_R + \mathbf{e}, \boldsymbol{\eta})\hat{\mathbf{B}}^{-1}(\boldsymbol{\zeta}_R + \mathbf{e}, \boldsymbol{\eta})\left[\mathbf{v}_R + \mathbf{C}^T\mathbf{e} - \hat{\mathbf{a}}(\boldsymbol{\zeta}_R + \mathbf{e}, \boldsymbol{\eta})\right] - \mathbf{P}(\boldsymbol{\zeta}_R + \mathbf{e}, \boldsymbol{\eta})\Delta\mathbf{u}\,. \tag{4.59}$$

In order to render the error dynamics stable, the coefficients in $\mathbf{C}$ have to be chosen such that the polynomials

$$s^{r_i} + c_{i,r_i-1}s^{r_i-1} + \ldots + c_{i,1}s + c_{i,0} \tag{4.60}$$

have its roots in the open l.h.c.p., since then $\mathbf{J} - \mathbf{H}\mathbf{C}^T$ has block diagonal structure, where the diagonal blocks are in controllable canonical form

$$\mathbf{J}_i - \mathbf{h}_i \cdot \mathbf{c}_i^T = \begin{bmatrix} 0 & 1 & & \mathbf{0} \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & 1 \\ -c_{i,0} & -c_{i,1} & \cdots & -c_{i,r_i-1} \end{bmatrix}$$

and it is well-known, that the characteristic polynomial of such a matrix is given by (4.60). Hence the error dynamics consists of m decoupled subsystems of order r_i, each of which is in controllable canonical form ([Wol74]). If (4.60) has roots in the open l.h.c.p., also the eigenvalues of the system matrix are stable and with it the error dynamics (4.58). Figure 4.4 shows the block diagram of the whole tracking control system.

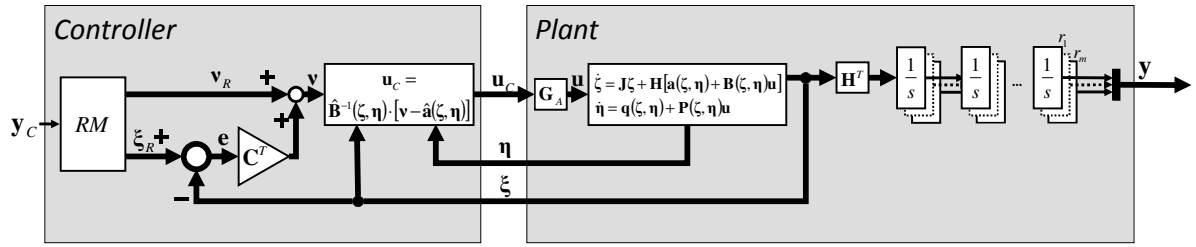


Figure 4.4 Tracking Control System [Hol04]

Furthermore, (4.58) highlights, that the error dynamics are excited by the augmented disturbance $\Delta(\boldsymbol{\zeta}, \boldsymbol{\eta}, \Delta\mathbf{u}, \mathbf{u})$, defined in (4.46). In section 4.3, an adaptive term will be added to the pseudo control (4.53) that partially compensates for these disturbances.

The question that now arises is, whether the tracking system – including internal dynamics – is stable, at least in absence of model uncertainties. Since the reference states are designed to be stable, it is admissible to consider the error dynamics (4.58). But also the internal dynamics are part of the whole system state and hence in absence of model uncertainties and disturbances, we get

$$\begin{aligned} \dot{\mathbf{e}}(t) &= \left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right)\cdot\mathbf{e}(t) \\ \dot{\boldsymbol{\eta}}(t)(t) &= \mathbf{q}(\boldsymbol{\zeta}_R(t) + \mathbf{e}(t), \boldsymbol{\eta}(t)) \end{aligned} \tag{4.61}$$

where we substituted $\boldsymbol{\zeta}=\boldsymbol{\zeta}_R+\mathbf{e}$. Lemma 4.1 implies that (4.61) is locally asymptotically stable in the case of zero reference state $\boldsymbol{\zeta}_R(t)\equiv\mathbf{0}$. But it is still left to show that it also holds for nonzero reference state. However, one can conjecture that $\boldsymbol{\zeta}_R$ is not allowed

to grow arbitrarily large in order not to violate stability of the overall system, which is stated next.

Theorem 4.3 Boundedness of Exactly Feedback Linearized Tracking System

Assume that the system

$$\dot{\mathbf{e}}(t) = \left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right)\cdot \mathbf{e}(t) \quad , \quad \mathbf{e}(t_0) = \mathbf{e}_0$$
$$\dot{\boldsymbol{\eta}}(t) = \mathbf{q}(\boldsymbol{\zeta}_R(t) + \mathbf{e}(t), \boldsymbol{\eta}(t)) \quad , \quad \boldsymbol{\eta}(t_0) = \boldsymbol{\eta}_0$$

where $\boldsymbol{\zeta}_R(t), \mathbf{e}(t) \in \mathbb{R}^r$, $\boldsymbol{\eta}(t) \in \mathbb{R}^{n-r}$, $\boldsymbol{\zeta}(t) = \boldsymbol{\zeta}_R(t) + \mathbf{e}(t)$, $\mathbf{z} = \begin{pmatrix} \boldsymbol{\zeta} \\ \boldsymbol{\eta} \end{pmatrix}$

has an equilibrium at $\mathbf{z} = \mathbf{0}$, $\mathbf{J} - \mathbf{H}\mathbf{C}^T$ is Hurwitz, $\mathbf{q}(\boldsymbol{\zeta}, \boldsymbol{\eta})$ is continuously differentiable on $\overline{\mathcal{B}_{r_0}} = \left\{\mathbf{z} \in \mathbb{R}^n \,\middle|\, \|\boldsymbol{\zeta}\|_2 \le r_\zeta, \|\boldsymbol{\zeta}\|_2 \le r_\eta\right\}$ for some $r_\zeta, r_\eta > 0$

and the zero dynamics $\dot{\boldsymbol{\eta}} = \mathbf{q}(\mathbf{0}, \boldsymbol{\eta})$ are asymptotically stable.

Then the system is uniformly ultimately bounded if $\|\boldsymbol{\zeta}_R(t)\| \le \overline{\zeta}$ for some sufficiently small $\overline{\zeta} > 0$

The proof is presented in Appendix F.4.

Unfortunately it is not possible to specify the rate of convergence, since we do not know the class $\mathcal{K}$ functions α_1 to α_4 which bound the Lyapunov function of the zero dynamics. If they were known, the rate of convergence could be specified. The result of this discussion is summarized in the following corollary.

Corollary 4.1 Boundedness of Exactly Feedback Linearized Tracking System with Rate of Convergence

Consider Theorem 4.3 and assume additionally that

- a symmetric positive definite matrix $\mathbf{P}$ solves $\left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right)^T \mathbf{P} + \mathbf{P}\left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right) = -\mathbf{I}$, where $\overline{\lambda}_P$, $\underline{\lambda}_P$ denote the maximum and minimum eigenvalue of $\mathbf{P}$.
- there are known class $\mathcal{K}$ functions $\alpha_1(x), \alpha_2(x), \alpha_3(x), \alpha_4(x)$, defined on $x \in [0, \rho]$, $0 < \rho < r_\eta$ and a positive definite function $V_0(\boldsymbol{\eta})$ such that $\alpha_1\left(\|\boldsymbol{\eta}\|_2\right) \le V_0(\boldsymbol{\eta}) \le \alpha_2\left(\|\boldsymbol{\eta}\|_2\right)$, $\left\|\frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right\|_2 \le \alpha_3\left(\|\boldsymbol{\eta}\|_2\right)$, $\frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \mathbf{q}(\mathbf{0}, \boldsymbol{\eta}) \le -\alpha_4\left(\|\boldsymbol{\eta}\|_2\right)$
- $\|\boldsymbol{\zeta}_R(t)\|_2 \le \overline{\zeta}$ such that $0 < \overline{\zeta} < \frac{\alpha_4\left(\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right)}{M \cdot \alpha_3(\rho)}$, where $\underline{u} = \min\left(\sqrt{\underline{\lambda}_P} \cdot \left(r_\zeta - \overline{\zeta}\right), \alpha_1(\rho)\right)$
- $\left\|\frac{\partial \mathbf{q}(\boldsymbol{\zeta}, \boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right\|_2 \le M$ on $\overline{\mathcal{B}_{r_0}}$ for some known $M>0$

- $\|\mathbf{e}_0\| \le \frac{\underline{u}}{2\sqrt{\bar{\lambda}_P}}$ and $\|\boldsymbol{\eta}_0\|_2 \le \alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)$.

Then

- $\|\mathbf{e}(t)\|_2 \le \|\mathbf{e}_0\|_2 \cdot \sqrt{\frac{\bar{\lambda}_P}{\underline{\lambda}_P}} \cdot \exp\left(-\frac{1}{2\bar{\lambda}_P} \cdot (t - t_0)\right)$
- $\|\boldsymbol{\eta}(t)\| \le b$ for all $t \ge t_0 + T(\delta, b)$ if $\boldsymbol{\eta}(t_0) \le \delta$ for some $\delta \in \left[0, \alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right]$, where

$$b = \alpha_1^{-1}\left(\frac{\sqrt{\bar{\lambda}_P}}{k_1} \cdot \left(M \cdot \alpha_3(\rho) \cdot \bar{\zeta} + k_2\right) + \alpha_2\left(\alpha_4^{-1}\left(M \cdot \alpha_3(\rho) \cdot \bar{\zeta} + k_2\right)\right)\right)$$

$$T(\delta, b) = \frac{\frac{\underline{u}}{2} - \frac{\sqrt{\bar{\lambda}_P}}{k_1} \cdot \left(M \cdot \alpha_3(\rho) \cdot \bar{\zeta} + k_2\right) + \alpha_2(\delta) - \alpha_2\left(\alpha_4^{-1}\left(M \cdot \alpha_3(\rho) \cdot \bar{\zeta} + k_2\right)\right)}{k_2}$$

and k_1, k_2 are chosen such that:

$$k_1 \ge \frac{2\sqrt{\bar{\lambda}_P}}{\underline{u}} \cdot \alpha_4\left(\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right) \quad , \quad 0 < k_2 < \alpha_4\left(\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right) - M \cdot \alpha_3(\rho) \cdot \bar{\zeta}$$

4.1.6 Pseudo Control Hedging

Reminiscent to (4.58), the tracking error dynamics are excited by two different sources. While the modeling error $\tilde{\mathbf{F}}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u})$ will be cancelled by an adaptive element, this cannot be done for the control deficiency. The control deficiency occurs, since capabilities of the actuators are physically limited, be it dynamics, rate- or position saturation. In other words, this term occurs, since the controller charges a reaction, the plant cannot follow.

But this particular term causes problems in the Lyapunov stability of the adaptive closed-loop system. The standard argument in literature is that the actuators are fast compared to the controlled dynamics – also referred to as *time scale separation* and, if the closed-loop dynamics are designed such that saturation will not become effective, the actuator dynamics and constraints can be neglected.

At the beginning of the 21st century, Johnson presented a new approach, which allows explicit consideration of the control deficiency term by modification of the reference model dynamics. Since references with detailed discussions are available for this approach ([Joh00a], [Joh00b], [Joh01] and [Hol04]), which is also referred to as *pseudo control hedging* (PCH) merely the crucial aspects will be pointed out here.

The main idea of PCH is to slow down the reference dynamics by the *expected reaction deficit*. Equation (4.56) could be considered as the desired reference

dynamics. But, since it is already expected, that the plant could not follow due to the actuator delay, the dynamics is slowed down by exactly this amount.

$$\dot{\zeta}_R = \mathbf{J}\zeta_R + \mathbf{H}\left(\mathbf{v}_R - \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u}\right) \tag{4.62}$$

The last term in (4.62) is also referred to as expected reaction deficit, since it is computed from the assumed decoupling matrix $\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})$, which differs from the real one. Computing the tracking error dynamics of the newly defined reference model and the external states in (4.45), obviously the term $\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u}$ is cancelled out, or one could say that the reaction deficit is “hidden” from the tracking error dynamics.

$$\dot{\mathbf{e}} = \left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right)\cdot\mathbf{e} + \mathbf{H}\cdot\tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{u}) \tag{4.63}$$

This is particularly beneficial, since the update of the adaptive parameters is based on the tracking error. Since the adaptive part cannot compensate for the control deficiency anyway, as stated above, it is reasonable that the error dynamics are not excited by this part. Figure 4.5 illustrates a tracking control system including the additional PCH part.

It is important to notice that PCH allows a Lyapunov stability analysis of the closed-loop adaptive system with explicit consideration of actuator dynamics and limitations, which assures boundedness of the tracking error. But, although the reference dynamics is designed as s stable system, PHC introduces an additional feedback from the plant and thus the reference model states cannot be considered stable a priori. Thus also the plant state can also grow unbounded even if the tracking error stays bounded. The stability proof of MRAC systems with PCH is still an open item and not solved for the general case until now. However there is a minor result ([Bie13]) that proves stability using results from $\mathcal{L}_1$ adaptive control ([Hov10]), if the actuator dynamics are described by a pure LTI system.

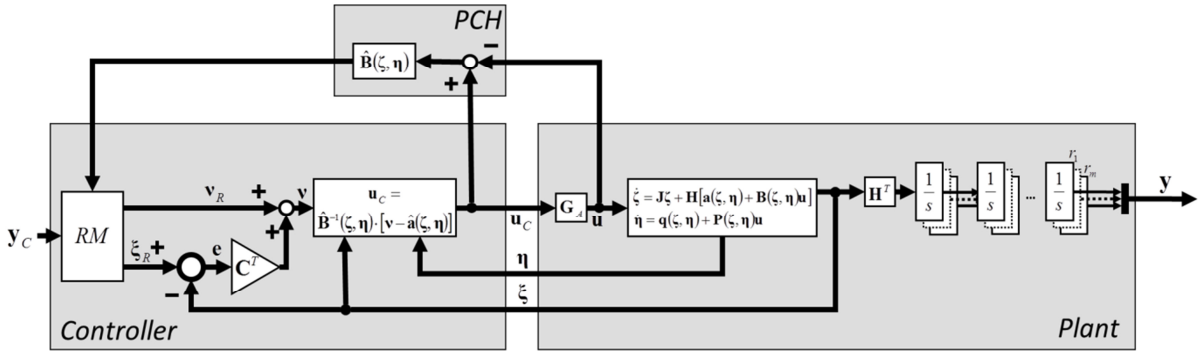


Figure 4.5 Tracking Control System with Pseudo Control Hedging

4.1.7 Model Extensions

Redundant Input Channels

So far, we considered the case of equal number of input and output channels. In this section, the whole algorithm is extended for the case that there are more inputs than output channels. Since the whole procedure of dynamic inversion has been presented in the last sections, this section will only highlight the differences to the case of equal number on input and output channels. Again, consider system (4.1),

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t)) + \mathbf{G}(\mathbf{x}(t)) \cdot \mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{h}(\mathbf{x}(t))\end{aligned} \tag{4.64}$$

but now the input vector has more entries than the output vector, i.e. $\mathbf{u}^T = (u_1(t), \ldots, u_p(t))$ and $p > m$. Applying the same procedure of differentiating each output until the input appears in the equation and, on this basis, defining the nonlinear state transformation accordingly, yields

$$\begin{aligned}\dot{\zeta} &= \mathbf{J} \cdot \zeta + \mathbf{H}\left[\mathbf{a}(\zeta, \boldsymbol{\eta}) + \mathbf{B}(\zeta, \boldsymbol{\eta}) \cdot \mathbf{u}\right] \\ \dot{\boldsymbol{\eta}} &= \mathbf{q}(\zeta, \boldsymbol{\eta}) + \mathbf{P}(\zeta, \boldsymbol{\eta}) \cdot \mathbf{u}\end{aligned} \tag{4.65}$$

with the difference that the decoupling matrix $\mathbf{B}(\zeta, \boldsymbol{\eta}) \in \mathbb{R}^{m \times p}$ is not quadratic anymore. This difference asks for a modified definition of relative degree.

Definition 4.3 Vector Relative Degree for Non-Quadratic Systems

The system (4.64) is said to have a (vector) relative degree $\mathbf{r}^T = (r_1 \quad \cdots \quad r_m)$, at $\mathbf{x}_0$ if

- $L_{\mathbf{G}}\left(L_{\mathbf{f}}^k h_i(\mathbf{x})\right) = \mathbf{0}^T$ in a vicinity of $\mathbf{x_0}$ for $k = 0, \ldots, r_i - 2,\ i = 1, \cdots, m$
- The decoupling matrix $\overline{\mathbf{B}}(\mathbf{x}_0)$ has full row rank $\mathbf{x_0}$

We observe that the regularity condition of the decoupling matrix is changed into a full row rank condition. It is hence assured that the linearizing state feedback condition

$$\overline{\mathbf{a}}(\mathbf{x}) + \overline{\mathbf{B}}(\mathbf{x})\mathbf{u} = \mathbf{v}$$

can still be solved, particularly there is an infinite number of $\mathbf{u}$, that solve the equation. If, however, the decoupling matrix does not have full row rank, then there is no solution of the above equation, if $\mathbf{v} - \overline{\mathbf{a}}(\mathbf{x})$ points out of the image space of $\overline{\mathbf{B}}(\mathbf{x})$.

In the quadratic case, the linearizing state feedback (4.26) involves the inverse of the decoupling matrix, which does not exist in the current situation. In order of obtain a modified linearizing state feedback, recall the intention of dynamic inversion. One has to find a static feedback $\mathbf{u}(\mathbf{v}, \mathbf{x})$ such that the external dynamics reduce to

$$\dot{\zeta} = \mathbf{J} \cdot \zeta + \mathbf{H} \cdot \mathbf{v}. \tag{4.66}$$

The existence of a relative degree assures that $\overline{\mathbf{B}}(\mathbf{x})$ has full row rank and we can employ the Moore-Penrose pseudoinverse instead of the matrix inverse, since the matrix $\overline{\mathbf{B}}(\mathbf{x})\cdot\overline{\mathbf{B}}^T(\mathbf{x})\in\mathbb{R}^{m\times m}$ is quadratic and regular. Due to model uncertainties, the linearizing state feedback is designed, based on an assumed model (indicated by ^) and the linearizing state feedback is

$$\begin{aligned}\mathbf{u}_C &= \hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})\cdot\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\cdot\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})\right]^{-1}\cdot\left[\mathbf{v}-\hat{\mathbf{a}}(\zeta,\boldsymbol{\eta})\right]\\ &= \hat{\mathbf{B}}^+(\zeta,\boldsymbol{\eta})\cdot\left[\mathbf{v}-\hat{\mathbf{a}}(\zeta,\boldsymbol{\eta})\right]\end{aligned} \tag{4.67}$$

where the upper case "+" denotes the pseudoinverse. Of course, also more sophisticated control allocation methods, such as cascaded pseudoinverse, could be used here. However, incorporation of such iterative methods into the Lyapunov stability analysis of the closed loop system is quite involved and an appropriate method is not known to the author.

In case of redundant actuators, we arrive at dynamics (4.45) where the inverse is replaced by the pseudoinverse

$$\begin{aligned}\dot{\zeta} &= \mathbf{J}\zeta+\mathbf{H}\left[\mathbf{v}-\tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{u})-\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u}\right]\\ \dot{\boldsymbol{\eta}} &= \mathbf{q}(\zeta,\boldsymbol{\eta})+\mathbf{P}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^+(\zeta,\boldsymbol{\eta})\left[\mathbf{v}-\hat{\mathbf{a}}(\zeta,\boldsymbol{\eta})\right]-\mathbf{P}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u}\end{aligned} \tag{4.68}$$

It is also well-known that the control input, computed by (4.67) is the one with the smallest vector 2-norm of all control inputs that feedback linearize the system (e.g. [Gel01]). Finally the tracking error dynamics (4.58) remain unchanged beside the fact that the internal dynamics contain the pseudoinverse of the decoupling matrix instead of the ordinary matrix inverse.

$$\begin{aligned}\dot{\mathbf{e}} &= \left(\mathbf{J}-\mathbf{H}\mathbf{C}^T\right)\mathbf{e}+\mathbf{H}\left[\tilde{\mathbf{F}}(\zeta_R+\mathbf{e},\boldsymbol{\eta},\mathbf{u})+\hat{\mathbf{B}}(\zeta_R+\mathbf{e},\boldsymbol{\eta})\Delta\mathbf{u}\right]\\ \dot{\boldsymbol{\eta}} &= \mathbf{q}(\zeta_R+\mathbf{e})+\mathbf{P}(\zeta_R+\mathbf{e},\boldsymbol{\eta})\hat{\mathbf{B}}^+(\zeta_R+\mathbf{e},\boldsymbol{\eta})\left[\mathbf{v}_R+\mathbf{C}^T\mathbf{e}-\hat{\mathbf{a}}(\zeta_R+\mathbf{e},\boldsymbol{\eta})\right]-\mathbf{P}(\zeta_R+\mathbf{e},\boldsymbol{\eta})\Delta\mathbf{u}\end{aligned} \tag{4.69}$$

Remark

Often physical systems comprise actuators with different authority. In this case it could be reasonable to use a weighted pseudoinverse, which accounts for the different actuator authorities. Let $\mathbf{B}\in\mathbb{R}^{m\times p}$, $\mathbf{y}\in\mathbb{R}^m$ and a symmetric positive definite matrix $\mathbf{R}$ of appropriate dimension. Then $\mathbf{u}=\mathbf{B}_R^+\cdot\mathbf{y}$ simultaneously solves the linear equation $\mathbf{Bu}=\mathbf{y}$ and minimizes $J=\mathbf{u}^T\mathbf{R}^{-1}\mathbf{u}$ where

$$\mathbf{B}_R^+=\mathbf{R}\mathbf{B}\cdot\left(\mathbf{B}^T\mathbf{R}\mathbf{B}\right)^{-1} \tag{4.70}$$

is the weighted pseudoinverse.

Nonaffine-in-Control Systems

Beside the aerodynamic control surfaces, which enter the dynamics linearly, the Extreme Star also comprises thrust vector control that enters the dynamics in a nonlinear manner. The dynamic inversion framework can also be extended to that case, by making the state dependent nonlinearity in (4.64) also dependent on the nonaffine controls.

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x},\mathbf{u}_N)+\mathbf{G}(\mathbf{x})\mathbf{u}\\ \mathbf{y} &= \mathbf{h}(\mathbf{x})\end{aligned} \tag{4.71}$$

where $\mathbf{u}_N \in \mathcal{U} \subset \mathbb{R}^{p_N}$ is a vector of nonaffine controls and $\mathbf{f}: \mathcal{D}\times\mathcal{U}\rightarrow\mathbb{R}^n$ is a nonlinear smooth map that describes the effect of the nonlinear controls onto the dynamics. For computation of relative degree we differentiate the i^{th} output.

$$\begin{aligned}\dot{y}_i(t) &= \frac{\partial h_i(\mathbf{x})}{\partial \mathbf{x}}\dot{\mathbf{x}} = \frac{\partial h_i(\mathbf{x})}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x},\mathbf{u}_N)+\frac{\partial h_i(\mathbf{x})}{\partial \mathbf{x}}\mathbf{G}(\mathbf{x})\mathbf{u}\\ &= L_{\mathbf{f}}h_i(\mathbf{x},\mathbf{u}_N)+L_{\mathbf{G}}h_i(\mathbf{x})\mathbf{u}\end{aligned}$$

For the affine part, we assumed that $L_{\mathbf{G}}h_i = \mathbf{0}^T$, i.e. the expression does not depend on the input. For a nonlinear analog, we assume, that the expression $L_{\mathbf{f}}h_i(\mathbf{x},\mathbf{u}_N)$ does not depend on $\mathbf{u}_N$, which can be expressed mathematically.

$$\frac{\partial L_{\mathbf{f}}h_i(\mathbf{x},\mathbf{u}_N)}{\partial \mathbf{u}_N} = \mathbf{0}$$

Note that one can also write $L_{\mathbf{f}}h_i(\mathbf{x})$, i.e. the expression only depends on the state vector. If the assumptions hold, the output is differentiated once more.

$$\begin{aligned}\dot{y}_i(t) &= L_{\mathbf{f}}h_i(\mathbf{x})\\ \ddot{y}_i(t) &= \left(\frac{\partial L_{\mathbf{f}}h_i(\mathbf{x})}{\partial \mathbf{x}}\right)(\mathbf{f}(\mathbf{x},\mathbf{u}_N)+\mathbf{G}(\mathbf{x})\mathbf{u})\\ &= L_{\mathbf{f}}^2h_i(\mathbf{x},\mathbf{u}_N)+L_{\mathbf{G}}L_{\mathbf{f}}h_i(\mathbf{x})\mathbf{u}\end{aligned}$$

and the procedure is repeated until, let's say until the r_i^{th} derivative, either

- $L_{\mathbf{G}}L_{\mathbf{f}}^{r_i-1}h_i(\mathbf{x}) \neq \mathbf{0}^T$ or
- $\frac{\partial}{\partial \mathbf{u}_N}\left[L_{\mathbf{f}}^{r_i}h_i(\mathbf{x},\mathbf{u}_N)\right] \neq \mathbf{0}^T$

and we obtain

$$y_i^{(r_i)}(r) = L_{\mathbf{f}}^{r_i}h_i(\mathbf{x},\mathbf{u}_N)+L_{\mathbf{G}}L_{\mathbf{f}}^{r_i-1}h_i(\mathbf{x})\mathbf{u}$$

Repeating this procedure for each output yields following expression

$$\begin{pmatrix} y_1^{(r_1)} \\ \vdots \\ y_m^{(r_m)} \end{pmatrix} = \bar{\mathbf{a}}(\mathbf{x},\mathbf{u}_N) + \bar{\mathbf{B}}(\mathbf{x})\mathbf{u} \tag{4.72}$$

where $\bar{\mathbf{B}}(\mathbf{x})$ is defined in (4.7) and

$$\bar{\mathbf{a}}(\mathbf{x},\mathbf{u}_N) = \begin{pmatrix} L_{\mathbf{f}}^{r_1} h_1(\mathbf{x},\mathbf{u}_N) \\ \vdots \\ L_{\mathbf{f}}^{r_m} h_m(\mathbf{x},\mathbf{u}_N) \end{pmatrix} \tag{4.73}$$

which motivates the definition of relative degree for nonaffine-in-control systems.

Definition 4.4 Vector Relative Degree for Nonaffine Systems

The system (4.71) is said to have a (vector) relative degree $\mathbf{r}^T = (r_1 \quad \cdots \quad r_m)$, at $\mathbf{x}_0$ if

- $L_{\mathbf{G}} L_{\mathbf{f}}^k h_i(\mathbf{x},\mathbf{u}_N) = \mathbf{0}^T$ and $\frac{\partial}{\partial \mathbf{u}_N} L_{\mathbf{f}}^{k+1} h_i(\mathbf{x},\mathbf{u}_N) = \mathbf{0}^T$

  in a vicinity of $\mathbf{x_0}$ and $\mathbf{u}_N \in \mathcal{U}$, for $k = 0,\ldots,r_i - 2,\ i = 1,\cdots,m$
- The decoupling matrix $\mathbf{\Gamma}(\mathbf{x},\mathbf{u}_N) = \left[\bar{\mathbf{B}}(\mathbf{x}) \quad \frac{\partial}{\partial \mathbf{u}_N} \bar{\mathbf{a}}(\mathbf{x},\mathbf{u}_N) \right]$

  has full row rank at $\mathbf{x_0}$.

References on relative degree for nonaffine systems are quite rare. However, in the field of chemical engineering, a definition of relative degree for nonaffine SISO systems is available ([Hen90], [Hen96]).

Definition 4.4 is a new extension of this definition to the MIMO case.

The existence of a relative degree requires the partial derivative of $L_{\mathbf{f}}^{k+1} h_i(\mathbf{x},\mathbf{u}_N)$ w.r.t. $\mathbf{u}_N$ to vanish in a vicinity of $\mathbf{x}_0$. It is not sufficient that the derivative has a zero at $\mathbf{x}_0$, but a variation of $\mathbf{u}_N$ is required not to have an effect onto the dynamics (locally) at all. This requirement can be considered as a generalization of the affine case, where the analogous requirement $L_{\mathbf{G}} L_{\mathbf{f}}^k h_i(\mathbf{x}) = \mathbf{0}^T$ could also be written as

$$\frac{\partial}{\partial \mathbf{u}} \left(L_{\mathbf{G}} L_{\mathbf{f}}^k h_i(\mathbf{x})\mathbf{u} \right) = \mathbf{0}^T$$

in a vicinity of $\mathbf{x}_0$. Particularly for $k=r_i-2$, in the nonaffine case the condition reads as

$$\frac{\partial}{\partial \mathbf{u}_N} \left(L_{\mathbf{f}}^{r_i - 1} h_i(\mathbf{x},\mathbf{u}_N) \right) = \mathbf{0}^T$$

which implies that

$$L_{\mathbf{G}} L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x}) = \frac{\partial L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x})}{\partial \mathbf{x}} \mathbf{G}(\mathbf{x})$$

is locally independent of $\mathbf{u}_N$ and so is $\overline{\mathbf{B}}(\mathbf{x})$. Thus, if a relative degree exists, one can define a local nonlinear state transformation, analogously to the affine case (4.9), (4.10) and due to the conditions for the existence of a relative degree, $\mathbf{\Phi}_\zeta(\mathbf{x})$ is independent of $\mathbf{u}_N$ although the map $\mathbf{f}(\mathbf{x},\mathbf{u}_N)$ is not. Considerations, analogous to the affine case, result in the dynamics for the transformed system.

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\left[\bar{\mathbf{a}}(\mathbf{x},\mathbf{u}_N) + \overline{\mathbf{B}}(\mathbf{x})\right] \tag{4.74}$$

with $\mathbf{J}$ and $\mathbf{H}$ defined in (4.16). The linear independence of the rows of $d\mathbf{\Phi}_\zeta$ is assured by Theorem 4.1, however in case that relative degree $r<n$, additional n-r maps $\mathbf{\Phi}_\eta(\mathbf{x})$ have to be found such that the Jacobian of

$$\mathbf{z} = \begin{pmatrix} \zeta \\ \eta \end{pmatrix} = \mathbf{\Phi}(\mathbf{x}) = \begin{pmatrix} \mathbf{\Phi}_\zeta(\mathbf{x}) \\ \mathbf{\Phi}_\eta(\mathbf{x}) \end{pmatrix}$$

is regular at $\mathbf{x}_0$, rendering the map a local diffeomorphism such that the inverse $\mathbf{x}=\mathbf{\Phi}^{-1}(\mathbf{z})$ exists uniquely in a vicinity of $\mathbf{x}_0$. Theorem 4.2 assures the existence of such a $\mathbf{\Phi}_\eta(\mathbf{x})$ and hence the dynamics of the Byrnes-Isidori normal-form are

$$\begin{aligned} \dot{\zeta} &= \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{a}(\zeta,\eta,\mathbf{u}_N) + \mathbf{B}(\zeta,\eta)\mathbf{u}\right] \\ \dot{\eta} &= \mathbf{q}(\zeta,\eta,\mathbf{u}_N) + \mathbf{P}(\zeta,\eta)\mathbf{u} \end{aligned} \tag{4.75}$$

where $\mathbf{B}(\zeta,\eta)$, $\mathbf{P}(\zeta,\eta)$ are defined in (4.24) and

$$\mathbf{a}(\zeta,\eta,\mathbf{u}_N) = \begin{pmatrix} a_1(\zeta,\eta,\mathbf{u}_N) \\ \vdots \\ a_m(\zeta,\eta,\mathbf{u}_N) \end{pmatrix} = \bar{\mathbf{a}}\left(\mathbf{\Phi}^{-1}(\zeta,\eta),\mathbf{u}_N\right) \tag{4.76}$$

$$\mathbf{q}(\zeta,\eta,\mathbf{u}_N) = \left.\frac{\partial \mathbf{\Phi}_\eta(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{\Phi}^{-1}(\zeta,\eta)} \cdot \mathbf{f}(\zeta,\eta,\mathbf{u}_N) = \begin{bmatrix} L_{\mathbf{f}}\Phi_{\eta,1}\left(\mathbf{\Phi}^{-1}(\zeta,\eta)\right) \\ \vdots \\ L_{\mathbf{f}}\Phi_{\eta,n-r}\left(\mathbf{\Phi}^{-1}(\zeta,\eta)\right) \end{bmatrix}. \tag{4.77}$$

In the affine case, a well-defined relative degree assures the existence of a linearizing state feedback due to the linear character of the equation to be inverted.

$$\mathbf{a}(\zeta,\eta) + \mathbf{B}(\zeta,\eta)\mathbf{u} = \mathbf{v}$$

If $\mathbf{B}(\zeta,\eta)$ has full row rank, there is always some $\mathbf{u}$ that satisfies the equation above. In the nonaffine case, the existence of a relative degree is not sufficient. In order to clarify this, assume that the system does not have any affine controls. Then, if the equation

$$\mathbf{a}(\zeta,\eta,\mathbf{u}_N) = \mathbf{v} \tag{4.78}$$

is solvable for the nonaffine controls $\mathbf{u}_N$, the system is feedback linearizable. This, however, is not necessarily the case for a general nonlinear map. Assume that the system has a relative degree at $\mathbf{x}_0$ and define $\boldsymbol{\zeta}^0=\boldsymbol{\Phi}_\zeta(\mathbf{x}_0)$, $\boldsymbol{\eta}^0=\boldsymbol{\Phi}_\eta(\mathbf{x}_0)$. If, for some $\boldsymbol{\nu}^0\in\mathbb{R}^m$, there exists $\mathbf{u}^0\in\mathcal{U}$ that solves (4.78) at $(\boldsymbol{\zeta}^0, \boldsymbol{\eta}^0)$, it is shown in Corollary B.2 that (4.78) can also be solved for $\mathbf{u}_N$ in a vicinity of $(\boldsymbol{\zeta}^0, \boldsymbol{\eta}^0, \boldsymbol{\nu}^0)$.

Taking all recent arguments into account, it can be stated that the existence of a relative degree assures **global** feedback linearizability for **affine** systems while, in the **nonaffine** case, feedback linearizability is only assured **locally**, if it has a relative degree at $\mathbf{x}_0$ and, additionally, equation (4.78) is solved for some $(\boldsymbol{\zeta}^0, \boldsymbol{\eta}^0, \mathbf{u}^0, \boldsymbol{\nu}^0)$

Now we switch back to the full system with affine as well as nonaffine controls. For the considered aircraft, it is possible to separate a pure state dependent part and a part that depends on $\mathbf{u}_N$.

$$\mathbf{a}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N)=\mathbf{a}_x(\boldsymbol{\zeta},\boldsymbol{\eta})+\mathbf{g}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N) \tag{4.79}$$

If the pure affine part of the decoupling matrix already has full row rank, then the system is feedback linearizable anyway, since the equation

$$\mathbf{a}_x(\boldsymbol{\zeta},\boldsymbol{\eta})+\mathbf{g}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N)+\mathbf{B}(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{u}=\boldsymbol{\nu} \tag{4.80}$$

can always be fulfilled for some $(\mathbf{u}, \mathbf{u}_N)$, e.g. by fixing $\mathbf{u}_N$ and solving for $\mathbf{u}$, using the pseudoinverse. Moreover, for redundant input channels, there is an infinite number of controls that solve (4.80) if the affine part of the decoupling matrix has full row rank. Therefore, in section 4.1.8, two concepts for incorporation of redundant affine and nonaffine controls into the control channel will be presented.

Incorporation of Disturbances

Real physical systems are also often subjected to external disturbances, such as wind in case of aircraft. In the equations derived so far, disturbances can be incorporated easily into the framework. Assume that the state nonlinearity of system (4.1) additionally depends on some constant external disturbance vector $\mathbf{d}\in\mathbb{R}^d$.

$$\begin{aligned}\dot{\mathbf{x}}(t)&=\mathbf{f}(\mathbf{x},\mathbf{d})+\mathbf{G}(\mathbf{x})\mathbf{u}\\ \mathbf{y}&=\mathbf{h}(\mathbf{x})\end{aligned} \tag{4.81}$$

The procedure of determining the relative degree takes a very similar form

$$\dot{y}_i(t)=\frac{\partial h_i(\mathbf{x})}{\partial\mathbf{x}}\dot{\mathbf{x}}=\frac{\partial h_i(\mathbf{x})}{\partial\mathbf{x}}\mathbf{f}(\mathbf{x},\mathbf{d})+\frac{\partial h_i(\mathbf{x})}{\partial\mathbf{x}}\mathbf{G}(\mathbf{x})\mathbf{u}=L_\mathbf{f}h_i(\mathbf{x},\mathbf{d})+\underbrace{L_\mathbf{G}h_i(\mathbf{x})\mathbf{u}}_{=0}$$

If the last term vanishes, the output is differentiated once more

$$\ddot{y}_i(t)=\frac{\partial L_\mathbf{f}h_i(\mathbf{x},\mathbf{d})}{\partial\mathbf{x}}\mathbf{f}(\mathbf{x},\mathbf{d})+\underbrace{\frac{\partial L_\mathbf{f}h_i(\mathbf{x},\mathbf{d})}{\partial\mathbf{d}}\dot{\mathbf{d}}}_{=0}+\underbrace{\frac{\partial L_\mathbf{f}h_i(\mathbf{x},\mathbf{d})}{\partial\mathbf{x}}\mathbf{G}(\mathbf{x})\mathbf{u}}_{=0}$$

Again, if the last term is zero again, necessarily the disturbance dependent term also has to vanish, which is the case if

$$\frac{\partial L_{\mathbf{f}} h_i(\mathbf{x},\mathbf{d})}{\partial \mathbf{d}} = \mathbf{0}^T . \tag{4.82}$$

Note that this implies that $L_{\mathbf{f}} h_i(\mathbf{x},\mathbf{d})$ is independent of the disturbance, i.e. it is justified to write $L_{\mathbf{f}} h_i(\mathbf{x})$. Provided the influence of the disturbance vanishes in the subsequent derivatives, the output is derived w.r.t. time until the input appears.

$$y_i^{(r_i)}(t) = \frac{\partial L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x})}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x},\mathbf{d}) + \frac{\partial L_{\mathbf{f}}^{r_i-1} h_i(\mathbf{x})}{\partial \mathbf{x}} \mathbf{G}(\mathbf{x})\mathbf{u} = L_{\mathbf{f}}^{r_i} h_i(\mathbf{x},\mathbf{d}) + \underbrace{L_{\mathbf{G}} L_{\mathbf{f}}^{r_i-1}(\mathbf{x})}_{\neq 0}\mathbf{u}$$

It is hence necessary that the derivatives of the output that are not influenced by the control be not influenced by the disturbance, too. For existence of a relative degree we additionally require

$$\frac{\partial L_{\mathbf{f}}^{k} h_i(\mathbf{x})}{\partial \mathbf{d}} = \mathbf{0}^T \ , \ k = 1,\ldots,r_i - 1 . \tag{4.83}$$

However, it is allowed that

$$\frac{\partial L_{\mathbf{f}}^{r_i} h_i(\mathbf{x})}{\partial \mathbf{d}} \neq \mathbf{0}^T . \tag{4.84}$$

Remark

> Conditions (4.83) and (4.84) are associated with a relative degree of the system with external disturbances. Particularly, if (4.83) holds and the disturbance can be measured, a feedback linearization can be designed, that rejects the disturbance. It can hence be stated that a system is feedback linearizable, if the relative degree w.r.t. to the disturbance is equal or higher than the relative degree w.r.t. the controls. More information on this topic can be found in [Isi95], [Nij90], [Hen90].

In a similar manner, as without disturbance, the system can be transformed to Byrnes-Isidori normal form

$$\begin{aligned} \dot{\zeta}(t) &= \mathbf{J}\cdot\zeta + \mathbf{H}\cdot\left[\mathbf{a}(\zeta,\boldsymbol{\eta},\mathbf{d}) + \mathbf{B}(\zeta,\boldsymbol{\eta})\cdot\mathbf{u}\right] \\ \dot{\boldsymbol{\eta}}(t) &= \mathbf{q}(\zeta,\boldsymbol{\eta},\mathbf{d}) + \mathbf{P}(\zeta,\boldsymbol{\eta})\mathbf{u} \end{aligned} \tag{4.85}$$

where the state dependent nonlinearities $\mathbf{a}(\zeta,\boldsymbol{\eta},\mathbf{d})$ and $\mathbf{q}(\zeta,\boldsymbol{\eta},\mathbf{d})$ now additionally depend on the disturbance. If the disturbance can be measured, the linearizing state feedback is easily obtained.

$$\mathbf{u}_C = \hat{\mathbf{B}}^{+}(\zeta,\boldsymbol{\eta})\left[\mathbf{v} - \hat{\mathbf{a}}(\zeta,\boldsymbol{\eta},\mathbf{d})\right] \tag{4.86}$$

The linearizing state feedback is, of course, performed with an assumed model, that differs from the real plant and, accounting for actuators, as has been done in the preceding sections, the feedback linearized plant dynamics read as

$$\begin{aligned}\dot{\zeta} &= \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{v} - \tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{u},\mathbf{d}) - \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u}\right] \\ \dot{\boldsymbol{\eta}} &= \mathbf{q}(\zeta,\boldsymbol{\eta},\mathbf{d}) + \mathbf{P}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^{-1}(\zeta,\boldsymbol{\eta})[\mathbf{v} - \hat{\mathbf{a}}(\zeta,\boldsymbol{\eta})] - \mathbf{P}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u}\end{aligned} \tag{4.87}$$

and, consequently, the modeling error becomes dependent on the disturbance.

$$\begin{aligned}\tilde{\mathbf{a}}(\zeta,\boldsymbol{\eta},\mathbf{d}) &= \hat{\mathbf{a}}(\zeta,\boldsymbol{\eta},\mathbf{d}) - \mathbf{a}(\zeta,\boldsymbol{\eta},\mathbf{d}) \\ \tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{d}) &= \tilde{\mathbf{a}}(\zeta,\boldsymbol{\eta},\mathbf{d}) + \tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u}\end{aligned} \tag{4.88}$$

4.1.8 Propositions for Feedback Linearization of Nonaffine Systems

In this section, we will consider systems that comprise all extensions, introduced in section 4.1.7, namely redundant affine as well as nonaffine input channels and external disturbances.

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x},\mathbf{u}_N,\mathbf{d}) + \mathbf{G}(\mathbf{x})\mathbf{u} \tag{4.89}$$

If it is transformed to Byrnes-Isidori normal form, we obtain

$$\begin{aligned}\dot{\zeta}(t) &= \mathbf{J}\cdot\zeta + \mathbf{H}\cdot[\mathbf{a}_x(\zeta,\boldsymbol{\eta}) + \mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) + \mathbf{B}(\zeta,\boldsymbol{\eta})\cdot\mathbf{u}] \\ \dot{\boldsymbol{\eta}}(t) &= \mathbf{q}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) + \mathbf{P}(\zeta,\boldsymbol{\eta})\mathbf{u}\end{aligned} \tag{4.90}$$

where we have separated a pure state dependent nonlinearity according to (4.79) and we assume that $\mathbf{d}$ is available for measurement. In the following, two strategies are proposed for feedback linearization of such systems.

Nonaffine Controls Used for Linearizing State Feedback

At first, introduce a virtual control, which equals the actual effect of the nonaffine controls onto the dynamics.

$$\mathbf{w} = \mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) \tag{4.91}$$

The external dynamics also read as

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{a}_x(\zeta,\boldsymbol{\eta}) + \begin{bmatrix}\mathbf{B}(\zeta,\boldsymbol{\eta}) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{w}\end{pmatrix}\right] \tag{4.92}$$

and, in order to compute the linearizing state feedback, we use the pseudoinverse of the new augmented decoupling matrix $\begin{bmatrix}\mathbf{B}(\zeta,\boldsymbol{\eta}) & \mathbf{I}\end{bmatrix}$.

$$\begin{pmatrix}\mathbf{u}_C\\ \mathbf{w}_C\end{pmatrix} = \begin{bmatrix}\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})\\ \mathbf{I}\end{bmatrix}\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta}) + \mathbf{I}\right]^{-1}[\mathbf{v} - \hat{\mathbf{a}}_x(\zeta,\boldsymbol{\eta})] \tag{4.93}$$

and $\mathbf{u}_{N,C}$ will be henceforth denoted as the control that satisfies

$$\mathbf{w}_C = \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_{N,C},\mathbf{d}).$$

Notice that the linearizing state feedback is designed, based on the assumed model (indicated by "^"). Then, accounting for actuators, the external dynamics become

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{a}_x(\zeta,\boldsymbol{\eta}) + \begin{bmatrix}\mathbf{B}(\zeta,\boldsymbol{\eta}) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\end{pmatrix}\right]$$

$$= \mathbf{H}\left[\hat{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) + \begin{bmatrix}\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta}) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u}_C\\ \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_{N,C},\mathbf{d})\end{pmatrix}\right] - \mathbf{H}\left[\hat{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) + \begin{bmatrix}\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta}) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u}_C\\ \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_{N,C},\mathbf{d})\end{pmatrix}\right]$$

$$+ \mathbf{H}\begin{bmatrix}\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta}) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u}\\ \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\end{pmatrix} - \mathbf{H}\begin{bmatrix}\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta}) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u}\\ \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\end{pmatrix}$$

$$+ \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{a}_x(\zeta,\boldsymbol{\eta}) + \begin{bmatrix}\mathbf{B}(\zeta,\boldsymbol{\eta}) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u}\\ \mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\end{pmatrix}\right]$$

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\mathbf{v} - \mathbf{H}[\hat{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) - \mathbf{a}_x(\zeta,\boldsymbol{\eta})]$$
$$- \mathbf{H}\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u}_C - \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u} + \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_{N,C},\mathbf{d}) - \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]$$
$$- \mathbf{H}\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u} - \mathbf{B}(\zeta,\boldsymbol{\eta})\mathbf{u} + \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]$$

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\mathbf{v} - \mathbf{H}\left[\tilde{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) + \tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u} + \tilde{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right] - \mathbf{H}\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u} + \Delta\mathbf{w}\right]$$

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{v} - \tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{u},\mathbf{u}_N) - \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u} + \Delta\mathbf{w}\right] \tag{4.94}$$

where we have defined

$$\tilde{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) = \hat{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) - \mathbf{a}_x(\zeta,\boldsymbol{\eta}) \tag{4.95}$$
$$\tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta}) = \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta}) - \mathbf{B}(\zeta,\boldsymbol{\eta}) \tag{4.96}$$
$$\tilde{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) = \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) \tag{4.97}$$
$$\Delta\mathbf{u} = \mathbf{u}_C - \mathbf{u} \quad , \quad \Delta\mathbf{w} = \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_{N,C},\mathbf{d}) - \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}). \tag{4.98}$$

Thereby

$$\tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{u},\mathbf{u}_N,\mathbf{d}) = \tilde{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) + \tilde{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) + \tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u} \tag{4.99}$$

is the model deviation and

$$\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u} + \Delta\mathbf{w} \tag{4.100}$$

is the expected reaction deficit due to actuators. However, this approach leaves an open question. Is there some $\mathbf{u}_{N,C}$ that produces the demanded $\mathbf{w}_C$? The in-depth

discussion of this issue is postponed to section 4.5. To this end, some crude arguments on the design process of control systems are discussed in the following.

In the theoretical handling of control-affine dynamic inversion, there is typically no consideration of actuator saturation at first, but it is assumed that $\mathbf{u}$ can be arbitrarily large. This, however, does not hold in reality, since every actuator is saturated at some point. PCH includes this fact into analysis to a certain extent, though this point remains somehow unclean. For real systems, the bandwidth of the reference models and gains of the error feedback are decreased to such an extent that actuators will not run into saturation in any flight condition.

This also motivates an analogous assumption for nonaffine systems. Of course, for physical systems, the set of achievable virtual controls is bounded (in our case the virtual controls will be the moments, produced by the propulsion system). For the real system, the nonlinear map $\mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N)$ is actually restricted to $\mathbf{u}_N \in \mathcal{U} \subset \mathbb{R}^{p_N}$ and also the set $\mathcal{V} \subset \mathbb{R}^m$ of achievable virtual controls $\mathbf{w}$ is bounded. For stability analysis, we virtually extend the set of admissible $\mathbf{u}_N$ to $\mathbb{R}^{pn}$ such that every $\mathbf{w} \in \mathbb{R}^m$ can be achieved. Then, using Lyapunov stability analysis, the controller is designed such that the controller only demands virtual controls, the real system can produce, i.e. $\mathbf{u}_N$ is restricted to $\mathcal{U}$. Then, the controller is applied to the real system but, since the controller is designed such that $\mathbf{u}_N$ never leaves $\mathcal{U}$, the real plant will react exactly the same way as the plant, assumed for stability analysis (Figure 4.6).

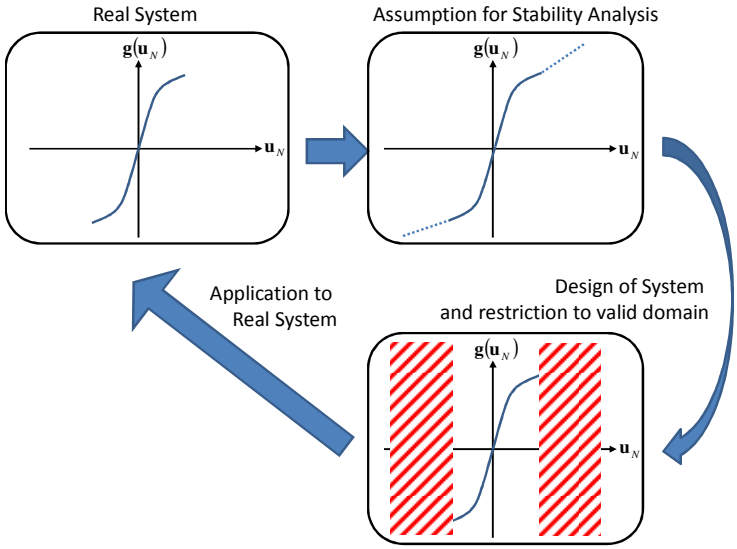


Figure 4.6 System Design Procedure in Case of Nonlinear-in-Control Systems

A further desirable property is that, to any virtual control $\mathbf{w} \in \mathbb{R}^m$ there is a unique $\mathbf{u}_N$, i.e. the inverse $\mathbf{u}_N = \mathbf{g}^{-1}(\zeta,\boldsymbol{\eta},\mathbf{w})$ exists. However, this is only possible, if $\mathbf{u}_N$ and $\mathbf{w}$ have the same dimensionality, i.e. $p_N = m$. However, this is not sufficient for the existence of an inverse and additional conditions have to be fulfilled. Even if the inverse

exists, it is often the case that an explicit expression for it is unknown. A solution to this issue is presented in section 4.5, where a novel approach tunes the desired control in an online algorithm, using a time scale separated gradient minimization approach.

Nonaffine Controls Excluded from Linearizing State Feedback

In some cases, the nonaffine controls are not suitable for control, since they do not provide a minimum control effectiveness uniformly over the whole operated state space. This is e.g. the case for trust vector controls using propeller engine. Since the trust, produced by the propellers, decreases with aerodynamic velocity, the trust vanishes at some velocity for a fixed thrust lever position. In this case, the propeller cannot produce any moments relative to the aircraft c.g. and hence cannot contribute to the aircraft control. However, turbo jet engines approximately provide constant thrust over the flight envelope and are more suitable for thrust vector control, providing a minimum control authority, uniformly over the whole flight envelope. As the Extreme Star uses propellers, thrust vectoring will not be used for linearizing state feedback. Instead, the nonaffine controls are prescribed from extern, i.e. we demand

$$\hat{\mathbf{g}}\left(\zeta,\boldsymbol{\eta},\mathbf{u}_{N,C}\right)=\mathbf{w}_C \tag{4.101}$$

for some external $\mathbf{w}_C$. An algorithm that finds $\mathbf{u}_{N,C}$ that produces $\mathbf{w}_C$, such as introduced in section 4.5, has to be implemented separately. Figure 4.7 shows block diagrams of both approaches.

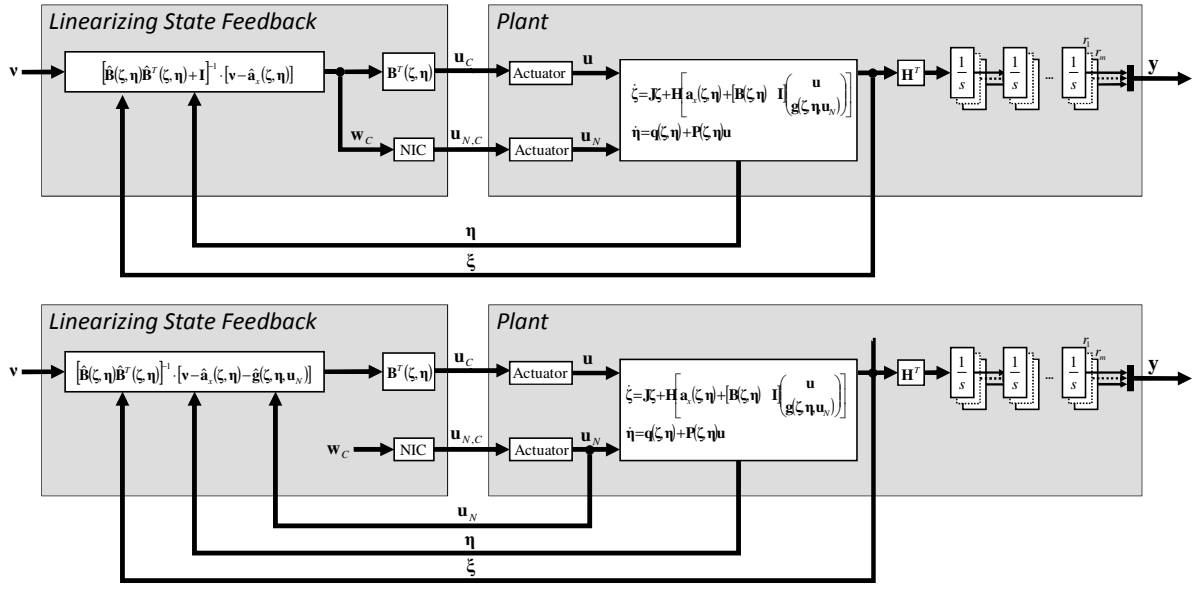


Figure 4.7 Propositions for Linearizing State Feedback of Nonaffine-in-Control Systems

Alternatively, $\mathbf{u}_{N,C}$ could also be prescribed directly. Since the nonaffine controls are not used for control, their effect onto the dynamics has to be canalled out by the linearizing state feedback.

$$\mathbf{u}_C=\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})\right]^{-1}\left(\mathbf{v}-\hat{\mathbf{a}}_x(\zeta,\boldsymbol{\eta})-\hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right) \tag{4.102}$$

The external $\mathbf{w}_C$ could be used to optimize some secondary objectives. One possible option is the use of the nonaffine controls to zero (4.79), i.e.

$$\mathbf{w}_C = -\hat{\mathbf{a}}_x(\zeta, \eta). \tag{4.103}$$

For an ideal feedback linearized plant (without model deviations and actuators), the linearizing state feedback reduces to

$$\mathbf{u} = \mathbf{B}^T(\zeta, \eta)\left[\mathbf{B}(\zeta, \eta)\mathbf{B}^T(\zeta, \eta)\right]^{-1}\mathbf{v}.$$

I.e. if the idealized plant is forced to remain at the stationary point $\zeta = \mathbf{0}$, which implies for the pseudo control $\mathbf{v} = \mathbf{0}$, no control action of the affine controls is needed. Transferred to the aircraft, this means that the nonaffine controls, namely the thrust vectoring, is used to produce trim moments such that the aircraft remains in steady state flight condition without any action of the control surfaces necessary.

Applying (4.102) to the plant (4.90), the external dynamics adopt a form similar to (4.94)

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{v} - \tilde{\mathbf{F}}(\zeta, \eta, \mathbf{u}, \mathbf{u}_N, \mathbf{d}) - \hat{\mathbf{B}}(\zeta, \eta)\Delta\mathbf{u}\right] \tag{4.104}$$

with the model deviation defined in (4.99).

4.2 Model Reference Adaptive Control

A conventional NDI algorithm has been developed so far. However, the derived equations are highly dependent on the dynamic model. Particularly in aviation, it takes plenty of effort to identify the model parameters, and if identified, they are still subjected to uncertainty. The equations, derived in section 4.1 already incorporate model uncertainties by distinguishing the real dynamics from the assumed dynamics, used for dynamic inversion. As result, the error dynamics (4.58) / (4.69) are excited by the augmented modeling error $\boldsymbol{\Delta}$.

In order to cancel parts of the modeling error and thereby improve performance, an adaptive part will be added to the algorithm. The method of choice is “Model Reference Adaptive Control” (MRAC).

4.2.1 Historical Note and MRAC Architectures

Investigations on adaptive systems started in the beginning second half of the 20th century, as there was need of fast and accurate control even in uncertain and changing environmental conditions. At first, different areas of engineering independently developed their own adaptive algorithms and own terminology, but the various approaches were unified in the following decades.

First flight experiments with adaptive control systems were accomplished in the 1960s, however, without thorough analysis of closed loop stability. This lead to a fatal crash of the X15A in 1967 and, as a consequence, adaptive flight control systems were shifted

out of focus for some years. This crash, however, lead to the insight that an exact mathematical stability analysis of the closed loop system is indispensable. In 1980, the research community refocused on adaptive flight control after Narendra provided a mathematical stability proof for MRAC systems([Nar80]). Important results pertaining to MRAC were consecutively summarized by Narendra and Annaswamy in 1989 [Nar05] and in the following decades, emphasis was put on performance and robustness of adaptive systems in presence of unmatched uncertainties and unmodeled dynamics, which bore various modifications of the parameter update equations. Today MRAC in its diverse occurrences and their stability analysis, using Lyapunov methods, could be considered as state of the technology and there are several, quite comprehensive references available, such as [Kha02], [Slo90], [Nar05]. In order to keep those readers, not familiar with MRAC, on track, a brief overview about the main ideas of MRAC is presented.

The standard MRAC approach is based on plants described by linear state space models, whose matrices are assumed unknown and the controller is designed with estimates of the unknown ideal controller parameters. In order to have a measure for the performance of the closed loop system, the plant states are compared to a reference model, and the error is used to update the parameter estimates. Figure 4.8 shows a block diagram of a standard MRAC architecture where the system matrix $\mathbf{A}$ is assumed unknown. The system, illustrated here, is also referred to as direct MRAC, since the controller parameters are update directly. Contrary to this, the indirect MRAC approach estimates the plant parameters and computes controller gains on these estimates in a second step.

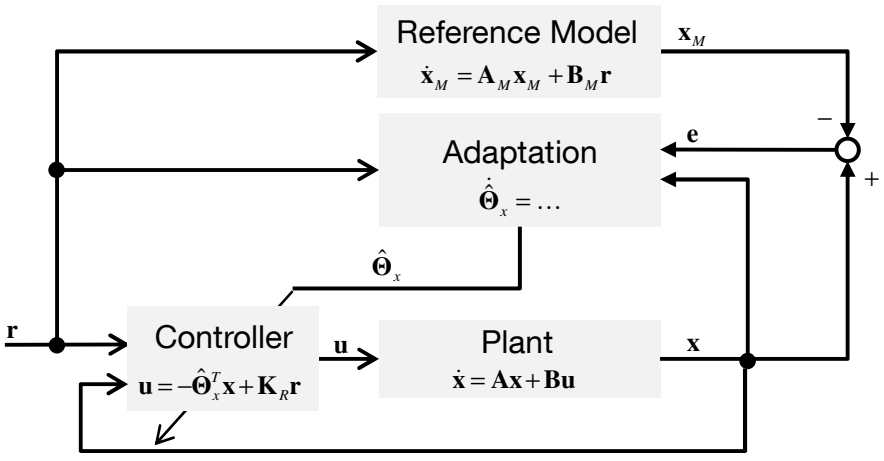


Figure 4.8 Standard MRAC

The reference model has to be designed such that $\mathbf{A}_M$ is, of course, stable and fulfills the so-called matching condition for all possible system matrices $\mathbf{A}$, i.e. there exists an ideal controller parameter Θ_x such that

$$\mathbf{A}_M = \mathbf{A} + \mathbf{B}\Theta_x^T$$

and the feed-forward gain has to be chosen, such that

$$\mathbf{B}_M = \mathbf{B}\mathbf{K}_R .$$

If these conditions are fulfilled, update laws can be designed using Lyapunov's theory. There are also approaches for the case, when the input matrix $\mathbf{B}$ is not known exactly, however some information about the structure of $\mathbf{B}$ has to be known.

In our case, the MRAC structure is slightly different, since the adaptive part is set up on an NDI baseline controller. In section 4.1.5 the non-adaptive tracking controller has been introduced, whose block diagram is depicted in Figure 4.4. In review of equation (4.45), plant and dynamic inversion controller could also be drawn as in Figure 4.9 where the linearized and decoupled dynamic of the ideal dynamic inversion algorithm is disturbed by the modeling error $\tilde{\mathbf{F}}(\zeta,\boldsymbol{\eta},\mathbf{u})$ and the control deficiency $\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\Delta\mathbf{u}$.

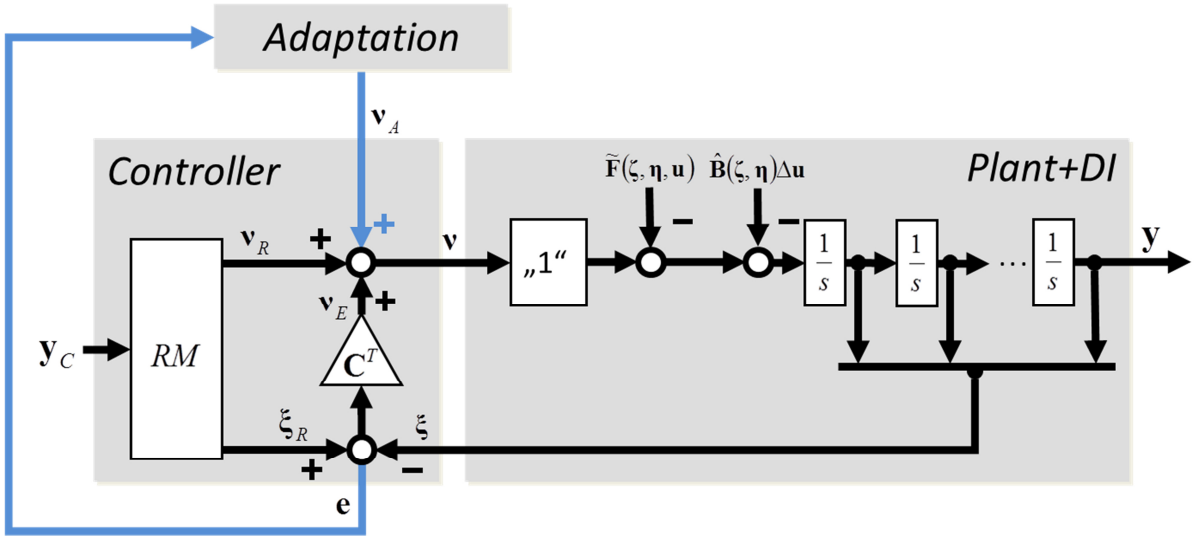


Figure 4.9 Dynamic Inversion Based MRAC

In order to make the whole scheme adaptive, an addition signal $\mathbf{v}_A$ is added to the pseudo control, and the update of the adaptive parameters is accomplished, using the same tracking error, as is used for the stabilizing error feedback in the conventional part of the controller. However, as stated in section 4.1.6, the explicit consideration of actuator causes problems in the stability proof, so in the remainder, we will neglect actuators in the stability proof, assuming that they are sufficiently fast. According to (4.58), the tracking error dynamics take the following form:

$$\dot{\mathbf{e}} = \left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right)\mathbf{e} + \tilde{\mathbf{a}}(\zeta,\boldsymbol{\eta}) + \tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta})\mathbf{u} \qquad (4.105)$$

where $\tilde{\mathbf{a}}(\zeta,\boldsymbol{\eta})$ denote the state dependent uncertainty and $\tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta})$ is the input dependent uncertainty. This equation will be the basis for adaptive control algorithm, developed subsequently.

4.2.2 Parameterization and Adaptive Compensation

Purpose of the adaptive part is the partial cancellation of the modeling error that excites the tracking error dynamics (4.58). In order to achieve this, the model deviations have to be parameterized. There are several strategies available.

If there isn't any information about the structure of the of the uncertainty available, neural networks are usually employed, since they comprise the property to approximate any smooth map arbitrarily close, provided the number of neurons is sufficiently large ([Fun89], [Hor89]). Successful flight experiments were accomplished with neural network based adaptive flight control systems in the 1990s ([Bri01], [Bri98], [Joh00a], [Joh00c]). However, this approach suffers from the fact, that the number of adaptive parameter is quite large, if the unmatched approximation error is needed to be kept small, which results in a slow adaptation rate in general. This is, though, the price to be paid for universal approximation of any uncertainty.

If there is some knowledge on the structure of the uncertainty that e.g. evolves from physical modeling, it is strongly recommended to employ this knowledge and keeping the number of adaptive parameters small. In MRAC systems, it is common practice to model uncertainties linearly in the parameters by some expression of the form

$$\boldsymbol{\Delta}(\mathbf{x}) = \boldsymbol{\Theta}^T \boldsymbol{\varphi}(\mathbf{x})$$

where $\boldsymbol{\Delta}(\mathbf{x}) \in \mathbb{R}^m$ is a state dependent uncertainty, $\boldsymbol{\Theta} \in \mathbb{R}^{s \times m}$ is a matrix of unknown, constant parameters and $\boldsymbol{\varphi}(\mathbf{x}) \in \mathbb{R}^s$ is a vector of known state dependent nonlinear functions also referred to as *regressors*.

For further considerations, we assume the dynamics to comprise all extensions introduced in section 4.1.7, namely redundant affine controls $\mathbf{u} \in \mathbb{R}^p$, $p \geq m$, nonaffine controls $\mathbf{u}_N \in \mathbb{R}^{p_N}$ and an external disturbance $\mathbf{d} \in \mathbb{R}^d$. Influences of actuators are neglected, since they cannot be canalled out by the adaptive part, assuming that they are fast enough. We arrive at the following external dynamics regardless whether we use the nonaffine controls for the linearizing state feedback or not (refer to equations (4.94), (4.104)).

$$\dot{\boldsymbol{\zeta}} = \mathbf{J}\boldsymbol{\zeta} + \mathbf{H}\left[\mathbf{v} - \tilde{\mathbf{a}}_x(\boldsymbol{\zeta}, \boldsymbol{\eta}) - \tilde{\mathbf{g}}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) - \tilde{\mathbf{B}}(\boldsymbol{\zeta}, \boldsymbol{\eta})\mathbf{u}\right] \qquad (4.106)$$

Alternatively PCH could hide the influence of actuators from the error dynamics (section 4.1.6), but then, the reference dynamics cannot be considered stable a priori, which causes problems to the stability proof, too.

State Dependent Uncertainty

The state dependent uncertainty is modeled linearly in the parameters. For the linearizing state feedback, we use

$$\hat{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) = \hat{\mathbf{K}}_x^T \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta}) \tag{4.107}$$

where $\hat{\mathbf{K}}_x \in \mathbb{R}^{s_x \times m}$ is a matrix of known constant parameters, $\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta}) \in \mathbb{R}^{s_x}$ is a column vector of known state dependent regressors. $\hat{\mathbf{K}}_x$ is a guess for the true and unknown parameter matrix $\mathbf{K}_x$ such that

$$\mathbf{a}_x(\zeta,\boldsymbol{\eta}) = \mathbf{K}_x^T \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta}) \tag{4.108}$$

Then

$$\boldsymbol{\Theta}_x = \mathbf{K}_x - \hat{\mathbf{K}}_x, \tag{4.109}$$

is the deviation between assumed and true state nonlinearity and the model uncertainty in (4.106) is

$$\tilde{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) = -\boldsymbol{\Theta}_x^T \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta}) \tag{4.110}$$

Affine Control Effectiveness

Adaptive cancellation of uncertainties in the input channel is not as straight forward as for state dependent uncertainty since the linearizing state feedback requires the existence of the pseudoinverse of the estimated decoupling matrix $\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})$([Lav09], [Bie10]). As the estimate is adapted by a dynamic equation, measures have to be taken that avoid the adaptive estimate to become singular. Therefore, in the following two parameterizations are introduced.

Absolute Scaling

In the first approach, the decoupling matrices of true and assumed models are related to each other by.

$$\mathbf{B}(\zeta,\boldsymbol{\eta}) = \mathbf{B}_L(\zeta,\boldsymbol{\eta})\boldsymbol{\Lambda}_L \quad \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta}) = \mathbf{B}_L(\zeta,\boldsymbol{\eta})\hat{\boldsymbol{\Lambda}}_L \tag{4.111}$$

where $\mathbf{B}_L(\zeta,\boldsymbol{\eta})$ is some known state dependent matrix of the same dimension as $\mathbf{B}(\zeta,\boldsymbol{\eta}), \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})$. $\boldsymbol{\Lambda}_L$ is an unknown quadratic *control effectiveness matrix*, associated with the true decoupling matrix and $\hat{\boldsymbol{\Lambda}}_L$ in an adaptive estimation of $\boldsymbol{\Lambda}_L$. With

$$\tilde{\boldsymbol{\Lambda}}_L = \hat{\boldsymbol{\Lambda}}_L - \boldsymbol{\Lambda}_L \tag{4.112}$$

the affine modeling error is

$$\tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta}) = \mathbf{B}_L(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Lambda}}_L. \tag{4.113}$$

The adaptive update of the estimated control effectiveness has to be restricted such that the matrix product $\mathbf{B}(\zeta,\boldsymbol{\eta})\hat{\boldsymbol{\Lambda}}_L$ results in

- a regular matrix for equal number of controls and outputs to be controlled (square decoupling matrix)
- a matrix with full row rank for larger number of controls than outputs (rectangular decoupling matrix with more columns than rows)

Provided that $\mathbf{B}_L(\zeta,\boldsymbol{\eta})$ fulfills the rank condition, in adaptive control the conditions above are typically fulfilled by restricting $\hat{\boldsymbol{\Lambda}}_L$ be a diagonal matrix with positive diagonal entries, which implies a considerable limitation of uncertainties in the control channel, that can be compensated for.

Interpreted physically, the diagonal entries of $\boldsymbol{\Lambda}_L$ determine the effectiveness of the single control surfaces. In order to clarify this, recall that the columns of $\mathbf{B}_L$ represent the effective direction of the single control surfaces. (Note that the dependent variables $(\zeta,\boldsymbol{\eta})$ are dropped for readability.)

$$\mathbf{B}_L=\begin{bmatrix}\mathbf{b}_{L1} & \cdots & \mathbf{b}_{Lp}\end{bmatrix}$$

For a diagonal control effectiveness matrix $\boldsymbol{\Lambda}_L=\mathrm{diag}(\lambda_{L1},\cdots,\lambda_{L1})$, we arrive at:

$$\mathbf{B}_L\boldsymbol{\Lambda}_L=\begin{bmatrix}\lambda_{L1}\mathbf{b}_{L1} & \cdots & \lambda_{Lp}\mathbf{b}_{Lp}\end{bmatrix}$$

Hence, the columns of $\mathbf{B}_L$ are scaled by the respective λ_{Li}. With this diagonal constraint it is however not possible that some control surface changes its effective direction. As aerodynamic models are usually subjected to uncertainties, the assumed effective direction of control surface might differ from the true one. E.g. aileron and rudder deflection both effect aerodynamic moments in both roll and yaw axis, where the ratio between roll and yaw axis has some distinct value. Now, the identified model could result in a slightly different ratio for the respective control surfaces and as a result, a diagonal control effectiveness matrix could not map the true control effectiveness (Figure 4.10).

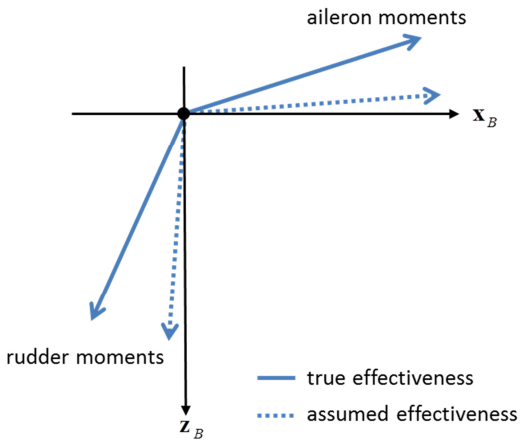


Figure 4.10 Deviation between True and Assumed Control Effectiveness Direction

However, a linear combination of assumed aileron and rudder effectiveness could map the true effective directions, which requires that the diagonal constraint is dropped.

In order to overcome this diagonal constraint, a novel algorithm for the adaptation of the control effectiveness is introduced in section 4.7, where the update of $\hat{\mathbf{\Lambda}}_L$ itself is transformed into an update of its singular value decomposition (SVD) while the singular values are limited away from zero avoiding singularity. The SVD update approach could even include cases where the effective direction of some control is inverted.

Relative Scaling

In nominal configuration and in case of exact system identification, one could omit $\mathbf{\Lambda}_L$ and choose $\mathbf{B}(\zeta,\boldsymbol{\eta})=\mathbf{B}_L(\zeta,\boldsymbol{\eta})$. This motivates a relative scaling where the adaptive parameter only contains the deviation from the nominal case.

$$\mathbf{B}(\zeta,\boldsymbol{\eta})=\mathbf{B}_L(\zeta,\boldsymbol{\eta})\cdot\left[\mathbf{I}+\mathbf{\Lambda}_L\right]\quad\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})=\mathbf{B}_L(\zeta,\boldsymbol{\eta})\cdot\left[\mathbf{I}+\hat{\mathbf{\Lambda}}_L\right] \tag{4.114}$$

For absolute scaling the case $\mathbf{\Lambda}_L=\mathbf{0}$ causes problems since, the inverse of the decoupling matrix does not exist. For relative scaling, there is no problem at this point, since $\mathbf{B}_L(\zeta,\boldsymbol{\eta})\left[\mathbf{I}+\mathbf{\Lambda}_L\right]$ can be inverted, if $\mathbf{B}_L(\zeta,\boldsymbol{\eta})$ fulfills the rank condition. By Theorem B.22, $\mathbf{I}+\hat{\mathbf{\Lambda}}_L$ is a regular matrix, if $\bar{\sigma}\left(\hat{\mathbf{\Lambda}}_L\right)<1$. Thus, if the adaptive $\hat{\mathbf{\Lambda}}_L$ is restricted such that the condition holds, the control effectiveness matrix will not become singular. However, this restriction excludes the case that a control changes its effective direction. This becomes obvious if we consider scalar control effectiveness with relative scaling

$$\hat{b}=b_L\left(1+\hat{\lambda}_L\right)\quad,\quad b=b_L\left(1+\lambda_L\right)$$

If the true effectiveness b is the inverted nominal one, i.e. $b=-b_L$, then we have $\lambda_L=-2$. This implies $\bar{\sigma}(\lambda_L)=2$ and hence the adaptive estimate cannot converge to the true value due to its restriction. For relative scaling, the modeling error $\tilde{\mathbf{B}}(\zeta,\boldsymbol{\eta})$ takes the same form as for absolute scaling (4.112), (4.113).

Nonaffine Controls

Beside aerodynamic control surfaces, which approximately enter the dynamics linearly, the Extreme Star also provides thrust vectoring which enter the dynamics in a nonlinear manner, as it involves trigonometric functions of the thrust vectoring angles. The nonlinear map $\mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$ that describes the effect of the nonlinear controls onto the external dynamics is assumed to be linearized parameterized

$$\mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})=\mathbf{K}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})+\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) \tag{4.115}$$

where $\mathbf{K}_N \in \mathbb{R}^{m\times s_N}$ is a matrix of unknown constant parameters; $\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\in\mathbb{R}^{s_N}$ is a vector of known regressors and $\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$ is a term, collecting all unmatched contributions, i.e. contributions that cannot be mapped by the chosen regressors. The assumed nonlinear map is defined accordingly.

$$\hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})=\left(\hat{\mathbf{K}}_N^T+\hat{\boldsymbol{\Theta}}_N^T\right)\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) \tag{4.116}$$

with the assumed and known parameter matrix $\hat{\mathbf{K}}_N$, and $\hat{\boldsymbol{\Theta}}_N$ is an adaptive estimation for the unknown deviation between assumed and real parameters.

$$\boldsymbol{\Theta}_N=\mathbf{K}_N-\hat{\mathbf{K}}_N \tag{4.117}$$

According to (4.115) – (4.117), the nonaffine modeling error is

$$\begin{aligned}\tilde{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})&=\left(\hat{\mathbf{K}}_N^T+\hat{\boldsymbol{\Theta}}_N^T-\mathbf{K}_N^T\right)\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})-\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\\ \tilde{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})&=\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})-\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}).\end{aligned} \tag{4.118}$$

where

$$\tilde{\boldsymbol{\Theta}}_N=\hat{\boldsymbol{\Theta}}_N-\boldsymbol{\Theta}_N \tag{4.119}$$

is the parameter-estimation-error.

External Dynamics

The adaptive parameters $\hat{\boldsymbol{\Theta}}_N$, $\hat{\boldsymbol{\Lambda}}_L$ have already been included into the linearizing state feedback (refer to linearizing state feedback (4.102) and parameterizations (4.111), (4.116)) and hence the control dependent uncertainties are cancelled out, if the adaptive parameters converge to the true values. Using uncertainty parameterizations (4.110), (4.113), (4.117)-(4.119), the external states dynamics (4.106) are

$$\dot{\zeta}=\mathbf{J}\zeta+\mathbf{H}\left[\mathbf{v}+\boldsymbol{\Theta}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})-\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})-\mathbf{B}_L(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Lambda}}_L\mathbf{u}+\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]. \tag{4.120}$$

In order to cancel the state dependent term, we choose for the pseudo control

$$\mathbf{v}=\mathbf{v}_R+\mathbf{v}_E+\mathbf{v}_A \tag{4.121}$$

where $\mathbf{v}_E$ is the error feedback, defined in (4.54), $\mathbf{v}_R$ is the reference pseudo control, defined in (4.55) and

$$\mathbf{v}_A=-\hat{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta}) \tag{4.122}$$

is the adaptive term and $\hat{\boldsymbol{\Theta}}_x$ is an adaptive estimation for $\boldsymbol{\Theta}_x$. With pseudo control (4.121), the external dynamics read as

$$\dot{\zeta}=\mathbf{J}\zeta+\mathbf{H}\left[\mathbf{v}_R+\mathbf{v}_E-\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})-\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})-\mathbf{B}_L(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Lambda}}_L\mathbf{u}+\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right] \tag{4.123}$$

where

$$\tilde{\boldsymbol{\Theta}}_x = \hat{\boldsymbol{\Theta}}_x - \boldsymbol{\Theta}_x . \tag{4.124}$$

4.3 Variant 1: Nonaffine Controls Excluded from Linearizing State Feedback

In this section, the whole adaptive controller structure is developed and a stability analysis is presented. Thereby, we use the approach, introduced in section 4.1.8, where the nonaffine controls are not used for linearizing state feedback but they are prescribed from extern (equations (4.102), (4.103) and Figure 4.7).

4.3.1 Tracking Error Dynamics

We start with the transformed system (4.90) and apply linearizing state feedback (4.102) which reads as

$$\mathbf{u} = \hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})\right]^{-1}\left(\mathbf{v} - \hat{\mathbf{a}}_x(\zeta,\boldsymbol{\eta}) - \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right). \tag{4.125}$$

With parameterizations of section 4.2.2, we obtain dynamics (4.123) of the feedback linearized plant. The reference dynamics are given by (4.55), (4.56). With the *tracking error*

$$\mathbf{e} = \zeta_R - \zeta \tag{4.126}$$

we get

$$\begin{aligned}
\dot{\mathbf{e}} &= \mathbf{J}\zeta_R + \mathbf{H}\dot{\mathbf{v}}_R - \mathbf{J}\zeta - \mathbf{H}\left[\dot{\mathbf{v}}_R + \mathbf{v}_E - \tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta}) - \tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\mathbf{u} + \boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right] \\
\dot{\mathbf{e}} &= \mathbf{J}\mathbf{e} - \mathbf{H}\left[\mathbf{C}^T\mathbf{e} - \tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta}) - \tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{B}_L(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Lambda}}_L\mathbf{u} + \boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right] \\
\dot{\mathbf{e}} &= \mathbf{A}_E\mathbf{e} + \mathbf{H}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta}) + \tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) + \mathbf{B}_L(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Lambda}}_L\mathbf{u} - \boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]
\end{aligned} \tag{4.127}$$

where $\mathbf{A}_E = \mathbf{J} - \mathbf{H}\mathbf{C}^T$ is the system matrix of the tracking error dynamics and $\mathbf{C}$ is chosen such that $\mathbf{A}_E$ is stable.

4.3.2 Lyapunov Stability Analysis

By Lemma C.4, there is a positive symmetric definite matrix $\mathbf{P}_E$ that satisfies the Lyapunov equation

$$\mathbf{A}_E^T\mathbf{P}_E + \mathbf{P}_E\mathbf{A}_E = -\mathbf{Q}_E \tag{4.128}$$

where $\mathbf{Q}_E$ is any symmetric positive definite matrix. The tracking error $\mathbf{e}$ as well as the adaptive parameter-estimation-errors $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Theta}}_N$ and $\tilde{\boldsymbol{\Lambda}}_L$ are states of the system and a Lyapunov function candidate is

$$V\left(\mathbf{e},\tilde{\boldsymbol{\Theta}}_x,\tilde{\boldsymbol{\Theta}}_N,\tilde{\boldsymbol{\Lambda}}_L\right)=\mathbf{e}^T\mathbf{P}_E\mathbf{e}+\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\gamma_x^{-1}\boldsymbol{\Gamma}_x^{-1}\tilde{\boldsymbol{\Theta}}_x+\tilde{\boldsymbol{\Theta}}_N^T\gamma_N^{-1}\boldsymbol{\Gamma}_N^{-1}\tilde{\boldsymbol{\Theta}}_N+\tilde{\boldsymbol{\Lambda}}_L\gamma_L^{-1}\boldsymbol{\Gamma}_L^{-1}\tilde{\boldsymbol{\Lambda}}_L^T\right] \tag{4.129}$$

where γ_x, γ_N and γ_L are positive scalars, denoted as learning rates and $\boldsymbol{\Gamma}_x$, $\boldsymbol{\Gamma}_N$ and $\boldsymbol{\Gamma}_L$ are symmetric positive definite matrices of appropriate dimension. As will be explained later, these matrices take the role of a weighting for the Frobenius norm of the parameter-estimation-errors. This potentially gives room for reduction of conservativeness in the stability analysis. The time derivative along the trajectory is:

$$\begin{aligned}
\dot{V}&=\dot{\mathbf{e}}^T\mathbf{P}_E\mathbf{e}+\mathbf{e}^T\mathbf{P}_E\dot{\mathbf{e}}\\
&\quad+\mathrm{tr}\left[\dot{\tilde{\boldsymbol{\Theta}}}_x^T\gamma_x^{-1}\boldsymbol{\Gamma}_x^{-1}\tilde{\boldsymbol{\Theta}}_x+\tilde{\boldsymbol{\Theta}}_x^T\gamma_x^{-1}\boldsymbol{\Gamma}_x^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_x\right]+\mathrm{tr}\left[\dot{\tilde{\boldsymbol{\Theta}}}_N^T\gamma_N^{-1}\boldsymbol{\Gamma}_N^{-1}\tilde{\boldsymbol{\Theta}}_N+\tilde{\boldsymbol{\Theta}}_N^T\gamma_N^{-1}\boldsymbol{\Gamma}_N^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_N\right]+\mathrm{tr}\left[\dot{\tilde{\boldsymbol{\Lambda}}}_L\gamma_L^{-1}\boldsymbol{\Gamma}_L^{-1}\tilde{\boldsymbol{\Lambda}}_L^T+\tilde{\boldsymbol{\Lambda}}_L\gamma_L^{-1}\boldsymbol{\Gamma}_L^{-1}\dot{\tilde{\boldsymbol{\Lambda}}}_L^T\right]\\
\dot{V}&=\overbrace{\left\{\mathbf{A}_E\mathbf{e}+\mathbf{H}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})+\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})+\mathbf{B}_L(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Lambda}}_L\mathbf{u}-\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]\right\}^T}^{\dot{\mathbf{e}}^T}\mathbf{P}_E\,\mathbf{e}\\
&\quad+\mathbf{e}^T\mathbf{P}_E\overbrace{\left\{\mathbf{A}_E\mathbf{e}+\mathbf{H}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})+\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})+\mathbf{B}_L(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Lambda}}_L\mathbf{u}-\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]\right\}}^{\dot{\mathbf{e}}}\\
&\quad+\mathrm{tr}\left[\dot{\tilde{\boldsymbol{\Theta}}}_x^T\gamma_x^{-1}\boldsymbol{\Gamma}_x^{-1}\tilde{\boldsymbol{\Theta}}_x+\tilde{\boldsymbol{\Theta}}_x^T\gamma_x^{-1}\boldsymbol{\Gamma}_x^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_x\right]+\mathrm{tr}\left[\dot{\tilde{\boldsymbol{\Theta}}}_N^T\gamma_N^{-1}\boldsymbol{\Gamma}_N^{-1}\tilde{\boldsymbol{\Theta}}_N+\tilde{\boldsymbol{\Theta}}_N^T\gamma_N^{-1}\boldsymbol{\Gamma}_N^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_N\right]+\mathrm{tr}\left[\dot{\tilde{\boldsymbol{\Lambda}}}_L\gamma_L^{-1}\boldsymbol{\Gamma}_L^{-1}\tilde{\boldsymbol{\Lambda}}_L^T+\tilde{\boldsymbol{\Lambda}}_L\gamma_L^{-1}\boldsymbol{\Gamma}_L^{-1}\dot{\tilde{\boldsymbol{\Lambda}}}_L^T\right]\\
\dot{V}&=\overbrace{\mathbf{e}^T\left(\mathbf{A}_E^T\mathbf{P}_E+\mathbf{P}_E\mathbf{A}_E\right)\mathbf{e}}^{-\mathbf{Q}_E}\\
&\quad+\left[\boldsymbol{\varphi}^T(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Theta}}_x+\boldsymbol{\omega}^T(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\tilde{\boldsymbol{\Theta}}_N+\mathbf{u}^T\tilde{\boldsymbol{\Lambda}}_L^T\mathbf{B}_L^T(\zeta,\boldsymbol{\eta})-\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]\mathbf{H}^T\mathbf{P}_E\,\mathbf{e}\\
&\quad+\left[\mathbf{e}^T\mathbf{P}_E\mathbf{H}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})+\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})+\mathbf{B}_L(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Lambda}}_L\mathbf{u}-\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]\right]\\
&\quad+\mathrm{tr}\left[\dot{\tilde{\boldsymbol{\Theta}}}_x^T\gamma_x^{-1}\boldsymbol{\Gamma}_x^{-1}\tilde{\boldsymbol{\Theta}}_x+\tilde{\boldsymbol{\Theta}}_x^T\gamma_x^{-1}\boldsymbol{\Gamma}_x^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_x\right]+\mathrm{tr}\left[\dot{\tilde{\boldsymbol{\Theta}}}_N^T\gamma_N^{-1}\boldsymbol{\Gamma}_N^{-1}\tilde{\boldsymbol{\Theta}}_N+\tilde{\boldsymbol{\Theta}}_N^T\gamma_N^{-1}\boldsymbol{\Gamma}_N^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_N\right]+\mathrm{tr}\left[\dot{\tilde{\boldsymbol{\Lambda}}}_L\gamma_L^{-1}\boldsymbol{\Gamma}_L^{-1}\tilde{\boldsymbol{\Lambda}}_L^T+\tilde{\boldsymbol{\Lambda}}_L\gamma_L^{-1}\boldsymbol{\Gamma}_L^{-1}\dot{\tilde{\boldsymbol{\Lambda}}}_L^T\right]
\end{aligned}$$

Since the expressions in dashed lines are scalars, they are equal to its transpose and also the trace of a matrix is equal to the trace of its transpose and hence

$$\begin{aligned}
\dot{V}=-\mathbf{e}^T\mathbf{Q}_E\mathbf{e}\quad&-2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})+\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})+\mathbf{B}_L(\zeta,\boldsymbol{\eta})\tilde{\boldsymbol{\Lambda}}_L\mathbf{u}-\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]\\
&+2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\gamma_x^{-1}\boldsymbol{\Gamma}_x^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_x\right]+2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\gamma_N^{-1}\boldsymbol{\Gamma}_N^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_N\right]+2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\gamma_L^{-1}\boldsymbol{\Gamma}_L^{-1}\dot{\tilde{\boldsymbol{\Lambda}}}_L^T\right]
\end{aligned}$$

Clearly, $\mathbf{e}^T\mathbf{P}\mathbf{H}$ is a row vector and the term in square brackets is a column vector. Using the dyadic product property of the trace operator, we obtain:

$$\begin{aligned}
\dot{V}=-\mathbf{e}^T\mathbf{Q}_E\mathbf{e}-2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})&+2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\left(\gamma_x^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_x+\boldsymbol{\Gamma}_x\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})\mathbf{e}^T\mathbf{P}_E\mathbf{H}\right)\right]\\
&+2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\Gamma}_N^{-1}\left(\gamma_N^{-1}\dot{\tilde{\boldsymbol{\Theta}}}_N+\boldsymbol{\Gamma}_N\boldsymbol{\omega}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\mathbf{e}^T\mathbf{P}_E\mathbf{H}\right)\right]\\
&+2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\left(\gamma_L^{-1}\dot{\tilde{\boldsymbol{\Lambda}}}_L^T+\boldsymbol{\Gamma}_L\mathbf{u}\mathbf{e}^T\mathbf{P}_E\mathbf{H}\mathbf{B}_L(\zeta,\boldsymbol{\eta})\right)\right]
\end{aligned} \tag{4.130}$$

For the parameter update, we choose the standard update law with parameter projection (refer to Appendix D.1)

$$\begin{aligned}
\dot{\hat{\boldsymbol{\Theta}}}_x &= \gamma_x \operatorname{Proj}\left[\hat{\boldsymbol{\Theta}}_x, -\boldsymbol{\Gamma}_x \boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta}) \mathbf{e}^T \mathbf{P}_E \mathbf{H} - \mathbf{M}_x\right]_{\varepsilon_x, \theta_{x,\max}, \boldsymbol{\Gamma}_x} \\
\dot{\hat{\boldsymbol{\Theta}}}_N &= \gamma_N \operatorname{Proj}\left[\hat{\boldsymbol{\Theta}}_N, -\boldsymbol{\Gamma}_N \boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) \mathbf{e}^T \mathbf{P}_E \mathbf{H} - \mathbf{M}_N\right]_{\varepsilon_N, \theta_{N,\max}, \boldsymbol{\Gamma}_N} \\
\dot{\hat{\boldsymbol{\Lambda}}}_L^T &= \gamma_L \operatorname{Proj}\left[\hat{\boldsymbol{\Lambda}}_L^T, -\boldsymbol{\Gamma}_L \mathbf{u} \mathbf{e}^T \mathbf{P}_E \mathbf{H} \mathbf{B}_L(\boldsymbol{\zeta},\boldsymbol{\eta}) - \mathbf{M}_L^T\right]_{\varepsilon_L, \lambda_{L,\max}, \boldsymbol{\Gamma}_L}
\end{aligned} \tag{4.131}$$

where $\mathbf{M}_x$, $\mathbf{M}_N$, $\mathbf{M}_L$ are modification terms to be specified. Further, we assume bounds on the true parameters.

Assumption A: *Bound on true parameters*

$$\left\|\boldsymbol{\Theta}_x\right\|_{\boldsymbol{\Gamma}_x} \le \theta_x \qquad \left\|\boldsymbol{\Theta}_N\right\|_{\boldsymbol{\Gamma}_N} \le \theta_N \qquad \left\|\boldsymbol{\Lambda}_L^T\right\|_{\boldsymbol{\Gamma}_L} \le \lambda_L \tag{4.132}$$

for some $\theta_x, \theta_N, \lambda_L > 0$ and symmetric positive definite matrices $\boldsymbol{\Gamma}_x, \boldsymbol{\Gamma}_N, \boldsymbol{\Gamma}_L$

Notice that the bounds are formulated in terms of the weighed rather than the classical Frobenius norm. The following values are chosen for the projection parameters.

$$\varepsilon_x > 0 \qquad \theta_{x,\max} = 2\theta_x\sqrt{1+\varepsilon_x} \tag{4.133}$$

$$\varepsilon_N > 0 \qquad \theta_{N,\max} = 2\theta_N\sqrt{1+\varepsilon_N} \tag{4.134}$$

$$\varepsilon_L > 0 \qquad \lambda_{L,\max} = 2\lambda_L\sqrt{1+\varepsilon_N} \tag{4.135}$$

i.e. the projection is only activated, if the weighted Frobenius norm of the respective parameter estimate exceeds the double of bounds (4.132) on the weighted Frobenius norms of the ideal parameters (refer to Appendix D.1). Therefore, an upper bound on the ideal parameter has to be known. The use of weighted Frobenius norm in the projection operator allows a customization of the set to which the adaptive parameters are restricted by the projection operator. This is a chance to construct less conservative parameter bounds than it can be done in case of classical Frobenius norm.

This fact can be intuitively explained for $\boldsymbol{\Theta} \in \mathbb{R}^2$. Of course, the true parameter is unknown, but one might have an idea about some region, where the true parameter is located, as depicted in Figure 4.11.

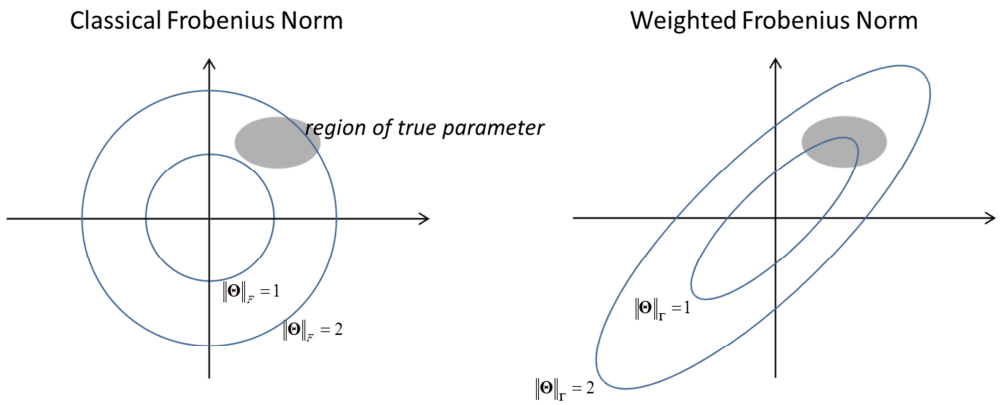


Figure 4.11 Weighted Frobenius Norm

Without weight, the lines of constant norm are circles and a parameter bound, using classical Frobenius norm would be $\|\boldsymbol{\Theta}\|_F \leq 2$. As the true parameters are supposed to be located in a certain direction, a weighting matrix could buckle the lines of constant norm in that direction, leading to a smaller bound for the true parameters. For the remainder, the indices, associated with projection operator are dropped for readability. Moreover, we also have to assume bounds on the unmatched uncertainty.

Assumption B*: Boundedness of unmatched uncertainty*

$$\|\boldsymbol{\delta}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d})\|_2 \leq D \tag{4.136}$$

for some $D \geq 0$ in a set

$$(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) \in \overline{\mathcal{B}}_{\bar{\zeta}} \times \mathcal{D}_\eta \times \mathcal{U} \times \mathcal{D} \tag{4.137}$$

Thereby $\overline{\mathcal{B}}_{\bar{\zeta}} = \left\{ \boldsymbol{\zeta} \in \mathbb{R}^r \,\middle|\, \|\boldsymbol{\zeta}\|_2 \leq \bar{\zeta} \right\}$ for some $\bar{\zeta} > 0$, $\mathcal{D}_\eta$ is a set, where the internal states can be restricted to, $\mathcal{U}$ is a convex set containing the origin and $\mathcal{D}$ is a set where $\mathbf{d}$ can be restricted to.

The Lyapunov analysis will prove boundedness of the tracking error $\mathbf{e}$. However, assumption B requires a bound for the plant state $\boldsymbol{\zeta}$, which is connected to the tracking error by

$$\boldsymbol{\zeta} = \boldsymbol{\zeta}_R - \mathbf{e}. \tag{4.138}$$

Since the reference model is stable, we can assure that

$$\|\boldsymbol{\zeta}_R(t)\|_2 \leq \bar{\zeta}_R \quad \text{for all} \quad t \geq t_0 \tag{4.139}$$

for sufficiently small exogenous inputs and initial conditions and some $\bar{\zeta}_R > 0$. Hence constraints on $\boldsymbol{\zeta}$ in assumption B is fulfilled, if the Lyapunov analysis assures

$$\|\mathbf{e}(t)\|_2 \leq r_e \quad \text{for all} \quad t \geq t_0 \tag{4.140}$$

where r_e satisfies

$$r_e + \bar{\zeta}_R \leq \bar{\zeta}. \tag{4.141}$$

Since we do not use the nonaffine controls for feedback linearization, but prescribe values from extern, it is easy to confine those to the valid set $\mathcal{U}$. The external disturbances cannot be influenced in general. However, restriction of disturbances to a bounded set is achieved by defining operating conditions under which the controlled system is operated. As wind and gusts are a main contribution to external disturbances for aircraft, a limit on $\mathbf{d}$ is assured, if the operation of the system is only allowed for appropriate wind conditions. The restriction of the internal states is the most problematic one. However, some remarks and potential solutions on this problem are given by remark 2 at the end of this section. We define:

$$\mathbf{Y}_x = -\boldsymbol{\Gamma}_x \boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{e}^T\mathbf{P}_E\mathbf{H} - \mathbf{M}_x,\ \mathbf{Y}_N = -\boldsymbol{\Gamma}_N \boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{M}_N,\ \mathbf{Y}_L^T = -\boldsymbol{\Gamma}_L \mathbf{u}\mathbf{e}^T\mathbf{P}_E\mathbf{H} - \mathbf{M}_L^T$$

and obtain:

$$\begin{aligned}\dot{V} = &-\mathbf{e}^T\mathbf{Q}_E\mathbf{e} - 2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) + 2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_N^{-1}\left(\mathrm{Proj}\left(\hat{\boldsymbol{\Theta}}_x,\mathbf{Y}_x\right) - \mathbf{Y}_x\right) - \tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\mathbf{M}_x\right] \\ &+ 2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\Gamma}_N^{-1}\left(\mathrm{Proj}\left(\hat{\boldsymbol{\Theta}}_N,\mathbf{Y}_N\right) - \mathbf{Y}_N\right) - \boldsymbol{\Theta}_N^T\boldsymbol{\Gamma}_N^{-1}\mathbf{M}_N\right] \\ &+ 2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\left(\mathrm{Proj}\left(\hat{\boldsymbol{\Lambda}}_L^T,\mathbf{Y}_L^T\right) - \mathbf{Y}_L^T\right) - \tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\mathbf{M}_L^T\right]\end{aligned}$$

With property (D.5) of the projection operator, derived in Appendix D, we arrive at

$$\dot{V} \le -\mathbf{e}^T\mathbf{Q}_E\mathbf{e} - 2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - 2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\mathbf{M}_x\right] - 2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\Gamma}_N^{-1}\mathbf{M}_N\right] - 2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\mathbf{M}_L^T\right] \tag{4.142}$$

In order to make the adaptation robust w.r.t unmatched uncertainties, we choose switching σ-modification (refer to Appendix D.2).

$$\begin{aligned}\mathbf{M}_x &= \sigma_x\,\overline{f}_{\varepsilon_x,\theta_x,\boldsymbol{\Gamma}_x}\left(\hat{\boldsymbol{\Theta}}_x\right)\hat{\boldsymbol{\Theta}}_x \\ \mathbf{M}_N &= \sigma_N\,\overline{f}_{\varepsilon_N,\theta_N,\boldsymbol{\Gamma}_N}\left(\hat{\boldsymbol{\Theta}}_N\right)\hat{\boldsymbol{\Theta}}_N \\ \mathbf{M}_L &= \sigma_L\,\overline{f}_{\varepsilon_L,\lambda_L,\boldsymbol{\Gamma}_L}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\hat{\boldsymbol{\Lambda}}_L\end{aligned} \tag{4.143}$$

Applying σ-modification to (4.142) yields.

$$\begin{aligned}\dot{V} \le &-\mathbf{e}^T\mathbf{Q}_E\mathbf{e} \quad -2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta} \\ &-2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\hat{\boldsymbol{\Theta}}_x\right] - 2\sigma_N\bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\Gamma}_N^{-1}\hat{\boldsymbol{\Theta}}_N\right] - 2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\hat{\boldsymbol{\Lambda}}_L^T\right]\end{aligned} \tag{4.144}$$

where the indices of $\bar{f}(\cdot)$ and arguments of $\boldsymbol{\delta}$ have been dropped for readability. In order to proof boundedness, we have to find an upper bound on $\dot{V}$. The first term is negative definite since $\mathbf{Q}_E$ is a positive definite matrix. However, the second term, induced by the unmatched uncertainty, is indefinite why stability cannot be concluded. However, at least, the system states are bounded if $\dot{V}$ is negative definite outside of a bounded set according to Corollary 3.2. Therefore, using (4.112), (4.119), (4.124), matrix induced norms, Cauchy-Schwartz inequality, Theorem B.20, assumption A with equation (4.133) – (4.135) and assumption B, we obtain an upper bound on $\dot{V}$.

$$\begin{aligned}\dot{V} \le &-\underline{\lambda}_{Q_E}\left\|\mathbf{e}\right\|_2^2 \quad +2\overline{\lambda}_{P_E}D\left\|\mathbf{e}\right\|_2 \\ &-2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left[\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 - \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\theta_{x,\max}\right] - 2\sigma_N\bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left[\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 - \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\theta_{N,\max}\right] \\ &-2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left[\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 - \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\lambda_{L,\max}\right]\end{aligned} \tag{4.145}$$

Note that, for the term involving the unmatched uncertainty, we used Cauchy-Schwartz inequality and matrix induced *2*-norm.

$$\left|\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta}\right| \le \left\|\mathbf{e}\right\|_2\left\|\mathbf{P}_E\mathbf{H}\boldsymbol{\delta}\right\|_2 \le \left\|\mathbf{e}\right\|_2\left\|\mathbf{P}_E\mathbf{H}\right\|_2\left\|\boldsymbol{\delta}\right\|_2$$

Since $\mathbf{P}_E$ is symmetric and positive definite, we have $\|\mathbf{P}_E\|_2 = \overline{\lambda}_{P_E}$; by inequality (1.2) and (4.18) we obtain

$$\|\mathbf{P}_E\mathbf{H}\|_2 \le \|\mathbf{P}_E\|_2 \|\mathbf{H}\|_2 \le \overline{\lambda}_{P_E}. \tag{4.146}$$

The Lyapunov function can be divided into four parts each of which only depends on one variable.

$$\begin{aligned}
h_e\left(\|\mathbf{e}\|_2\right) &= -\underline{\lambda}_{Q_E}\|\mathbf{e}\|_2^2 + 2\overline{\lambda}_{P_E} D\|\mathbf{e}\|_2 \\
h_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\right) &= -2\sigma_x \bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left[\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 - \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\theta_{x,\max}\right] \\
h_N\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\right) &= -2\sigma_N \bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left[\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 - \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\theta_{N,\max}\right] \\
h_L\left(\left\|\tilde{\boldsymbol{\Lambda}}_L\right\|_{\boldsymbol{\Gamma}_L}\right) &= -2\sigma_L \bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left[\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 - \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\lambda_{L,\max}\right]
\end{aligned} \tag{4.147}$$

All of them are concave parabolas and hence can be bounded from above. The bound is obtained by some lines of simple computations.

$$h_{e,\max} = \frac{\left(\overline{\lambda}_{P_E} D\right)^2}{\underline{\lambda}_{Q_E}} \qquad h_x = \frac{\sigma_x}{2}\theta_{x,\max}^2 \qquad h_{N,\max} = \frac{\sigma_N}{2}\theta_{N,\max}^2 \qquad h_{L,\max} = \frac{\sigma_L}{2}\lambda_{L,\max}^2 \tag{4.148}$$

With these limits, one can construct four functions, each of which only depends on the norm of one part of the whole state and which establish an upper bound on $\dot{V}$. The functions are obtained by taking the respective function of (4.147) and the worst-case bounds in (4.148) for the remaining functions.

$$\begin{aligned}
\gamma_e\left(\|\mathbf{e}\|_2\right) &= \underline{\lambda}_{Q_E}\|\mathbf{e}\|_2^2 - 2\overline{\lambda}_{P_E} D\|\mathbf{e}\|_2 - \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_N\theta_{N,\max}^2 + \sigma_L\lambda_{L,\max}^2\right) \\
\gamma_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\right) &= 2\sigma_x \bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left[\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 - \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\theta_{x,\max}\right] - \frac{\left(\overline{\lambda}_{P_E} D\right)^2}{\underline{\lambda}_{Q_E}} - \frac{1}{2}\left(\sigma_N\theta_{N,\max}^2 + \sigma_L\lambda_{L,\max}^2\right) \\
\gamma_N\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\right) &= 2\sigma_N \bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left[\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 - \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\theta_{N,\max}\right] - \frac{\left(\overline{\lambda}_{P_E} D\right)^2}{\underline{\lambda}_{Q_E}} - \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_L\lambda_{L,\max}^2\right) \\
\gamma_L\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\right) &= 2\sigma_L \bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left[\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 - \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\lambda_{L,\max}\right] - \frac{\left(\overline{\lambda}_{P_E} D\right)^2}{\underline{\lambda}_{Q_E}} - \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_N\theta_{N,\max}^2\right)
\end{aligned} \tag{4.149}$$

Using notation of Corollary 3.2, the Lyapunov function time derivative is bounded from above by

$$\dot{V} \le -\max\left[\gamma_e\left(\|\mathbf{e}\|_2\right), \gamma_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\right), \gamma_N\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\right), \gamma_L\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\right)\right]. \tag{4.150}$$

Further r_e, r_x, r_N, r_L in Corollary 3.2, have to be found, i.e. the radii of the balls

$$\|\mathbf{e}\|_2 \le r_e \,,\; \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} \le r_x \,,\; \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} \le r_N \,,\; \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} \le r_L$$

in which the Lyapunov function is enclosed by class $\mathcal{K}$ functions α_i and β_i. According to (4.141) we require

$$r_e \le \overline{\zeta} - \overline{\zeta}_R \,. \tag{4.151}$$

There is no restriction for the radii of the parameter-estimation-errors and hence we set

$$r_x, r_N, r_L = \infty \,. \tag{4.152}$$

In order to find the class $\mathcal{K}$ functions, which establish lower and upper bounds for the Lyapunov function candidate α_i and β_i according to Corollary 3.2, recall (4.129). With calculus of norms, we obtain

$$\begin{aligned}
&\alpha_e\left(\|\mathbf{e}\|_2\right) = \underline{\lambda}_{P_E}\|\mathbf{e}\|_2^2 && \beta_e\left(\|\mathbf{e}\|_2\right) = \overline{\lambda}_{P_E}\|\mathbf{e}\|_2^2 \\
&\alpha_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\right) = \beta_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\right) = \gamma_x^{-1}\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 && \alpha_N\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_N}\right) = \beta_N\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_N}\right) = \gamma_N^{-1}\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 \\
&\alpha_L\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\right) = \beta_L\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\right) = \gamma_L^{-1}\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 \,.
\end{aligned} \tag{4.153}$$

Furthermore, according to Corollary 3.2, we define

$$\underline{u} = \min\left[\alpha_e(r_e)\,,\, \alpha_x(r_x)\,,\, \alpha_N(r_N)\,,\, \alpha_L(r_L)\right] = \underline{\lambda}_{P_E} r_e^2 \,. \tag{4.154}$$

Notice that $\alpha_x(r_x)=\alpha_N(r_N)=\alpha_L(r_L)=\infty$ (refer to (4.152), (4.153)) and hence $\alpha_e(r_e)$ is the relevant value (4.154). Moreover, the radii, determining the sets of admissible initial conditions are

$$\rho_e = \sqrt{\frac{\underline{\lambda}_{P_E}}{\overline{\lambda}_{P_E}}}\frac{r_e}{2}, \rho_x = \sqrt{\gamma_x \underline{\lambda}_{P_E}}\,\frac{r_e}{2} \,,\; \rho_N = \sqrt{\gamma_N \underline{\lambda}_{P_E}}\,\frac{r_e}{2} \,,\; \rho_L = \sqrt{\gamma_L \underline{\lambda}_{P_E}}\,\frac{r_e}{2} \quad . \tag{4.155}$$

Then, if

$$\left\|\hat{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} \ge \theta_x \,,\; \left\|\hat{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} \ge \theta_N \,,\; \left\|\hat{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} \ge \lambda_L \tag{4.156}$$

we have

$$\bar{f}\left(\left\|\hat{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2\right) = \bar{f}\left(\left\|\hat{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2\right) = \bar{f}\left(\left\|\hat{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2\right) = 1$$

and hence γ_i in (4.149) are concave parabolas

$$\begin{aligned}
&\gamma_e = \underline{\lambda}_{Q_E}\|\mathbf{e}\|_2^2 - 2\overline{\lambda}_{P_E} D\|\mathbf{e}\|_2 - V_e && \gamma_x = 2\sigma_x\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 - 2\sigma_x\theta_x\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} - V_x - V_D \\
&\gamma_N = 2\sigma_N\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 - 2\sigma_N\theta_N\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} - V_N - V_D && \gamma_L = 2\sigma_L\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 - 2\sigma_L\lambda_L\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} - V_L - V_D
\end{aligned} \tag{4.157}$$

where we have abbreviated

$$V_e = \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_N\theta_{N,\max}^2 + \sigma_L\lambda_{L,\max}^2\right) \quad V_x = \frac{1}{2}\left(\sigma_N\theta_{N,\max}^2 + \sigma_L\lambda_{L,\max}^2\right) \quad V_D = \frac{\left(\overline{\lambda}_{P_E}D\right)^2}{\underline{\lambda}_{Q_E}}$$
$$V_N = \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_L\lambda_{L,\max}^2\right) \qquad V_L = \frac{1}{2}\left(\sigma_x\theta_x^2 + \sigma_N\theta_{N,\max}^2\right). \tag{4.158}$$

Thus, the functions (4.157) are rendered negative if the respective argument is sufficiently large, let us say

$$\|\mathbf{e}\|_2 \geq \mu_e\,,\ \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} \geq \mu_x\,,\ \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} \geq \mu_N\,,\ \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} \geq \mu_L$$

with μ_e, μ_x, μ_N, μ_L, according to Corollary 3.2. These are obtained, by choosing some small $K_0>0$ and determining μ_i such that

$$\gamma_e(\mu_e) = K_0 \qquad \gamma_x(\mu_x) = K_0 \qquad \gamma_N(\mu_N) = K_0 \qquad \gamma_L(\mu_L) = K_0\,. \tag{4.159}$$

Then, $\gamma_i(\cdot)$ are class $\mathcal{K}$ functions within the set $\overline{\mathcal{B}}_r \backslash \mathcal{B}_\mu$ ($\overline{\mathcal{B}}_r$, $\mathcal{B}_\mu$ according to notation of Corollary 3.2) and hence $\dot{V}$ is rendered negative definite in $\overline{\mathcal{B}}_r \backslash \mathcal{B}_\mu$ according to (4.150); μ_i evaluate to:

$$\mu_e = \frac{\overline{\lambda}_{P_E}D}{\underline{\lambda}_{Q_E}} + \sqrt{\left(\frac{\overline{\lambda}_{P_E}D}{\underline{\lambda}_{Q_E}}\right)^2 + \frac{V_e + K_0}{\underline{\lambda}_{Q_E}}} \qquad \mu_x = \frac{\theta_{x,\max}}{2} + \sqrt{\left(\frac{\theta_{x,\max}}{2}\right)^2 + \frac{V_D + V_x + K_0}{2\sigma_x}}$$
$$\mu_N = \frac{\theta_{N,\max}}{2} + \sqrt{\left(\frac{\theta_{N,\max}}{2}\right)^2 + \frac{V_D + V_N + K_0}{2\sigma_N}} \qquad \mu_L = \frac{\lambda_{L,\max}}{2} + \sqrt{\left(\frac{\lambda_{L,\max}}{2}\right)^2 + \frac{V_D + V_L + K_0}{2\sigma_L}} \tag{4.160}$$

Note that μ_x, μ_N and μ_L were computed under assumption (4.156) and (4.160) is only valid if this assumption is true in the whole set

$$\overline{\mathcal{B}}_r \backslash \mathcal{B}_\mu = \left\{\left(\mathbf{e}, \tilde{\boldsymbol{\Theta}}_x, \tilde{\boldsymbol{\Theta}}_N, \tilde{\boldsymbol{\Lambda}}_L\right) \middle| \mu_e \leq \|\mathbf{e}\|_2 \leq r_e\,, \mu_x \leq \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} \leq r_x, \mu_N \leq \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} \leq r_N, \lambda_L \leq \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} \leq r_L\right\}.$$

But, since $\mu_x>\theta_{x,max}$, $\mu_N>\theta_{N,max}$ and $\mu_L>\lambda_{L,max}$ (according to (4.160)), this requirement is fulfilled by definition of $\theta_{x,max}$, $\theta_{N,max}$, $\lambda_{L,max}$ in equations (4.133) - (4.135). According to Corollary 3.2, sufficient condition for ultimate boundedness is that $\mu_i < \rho_i$, $i=e,x,N,L$. Hence, for $\mu_e < \rho_e$, we obtain

$$\frac{\overline{\lambda}_{P_E}D}{\underline{\lambda}_{Q}} + \sqrt{\left(\frac{\overline{\lambda}_{P_E}D}{\underline{\lambda}_{Q_E}}\right)^2 + \frac{V_e + K_0}{\underline{\lambda}_{Q_E}}} < \sqrt{\frac{\underline{\lambda}_{P_E}}{\overline{\lambda}_{P_E}}}\frac{r_e}{2}$$

or, if we take the square

$$\left(\frac{\overline{\lambda}_{P_E}D}{\underline{\lambda}_{Q_E}}\right)^2 + \frac{V_e + K_0}{\underline{\lambda}_{Q_E}} < \left(\sqrt{\frac{\underline{\lambda}_{P_E}}{\overline{\lambda}_{P_E}}}\frac{r_e}{2}\right)^2 - 2\sqrt{\frac{\underline{\lambda}_{P_E}}{\overline{\lambda}_{P_E}}}\frac{r_e}{2}\frac{\overline{\lambda}_{P_E}D}{\underline{\lambda}_{Q_E}} + \left(\frac{\overline{\lambda}_{P_E}D}{\underline{\lambda}_{Q_E}}\right)^2$$

$$V_e + K_0 < k_e(r_e) := \frac{\underline{\lambda}_{Q_E}\underline{\lambda}_{P_E}}{4\overline{\lambda}_{P_E}} r_e^2 - \sqrt{\underline{\lambda}_{P_E}\overline{\lambda}_{P_E}}\, D\, r_e . \tag{4.161}$$

I.e., since $V_e+K_0>0$ but can be chosen arbitrarily small by appropriate choice of σ_x, σ_x , σ_x, $K_0>0$, it is necessary and sufficient for the existence of a solution to (4.161) that $k_e(r_e)>0$. This results in a lower bound for r_e.

$$r_e > \underline{e}^{(0)} := \frac{4\overline{\lambda}_{P_E}}{\underline{\lambda}_{Q_E}} \sqrt{\frac{\overline{\lambda}_{P_E}}{\underline{\lambda}_{P_E}}} D \tag{4.162}$$

and, if we substitute V_e in (4.161), we gain a condition for the σ-modification gains.

$$\theta_{x,\max}^2\sigma_x + \theta_{N,\max}^2\sigma_N + \lambda_{L,\max}^2\sigma_L + 2K_0 < 2k_e(r_e) \tag{4.163}$$

As the modification gains are restricted to positive values, (4.163) has a solution, only if condition (4.162) is fulfilled, since this implies $k_e(r_e)>0$. In a similar way, consider the second condition $\mu_x <\rho_x$,

$$\frac{\theta_{x,\max}}{2} + \sqrt{\left(\frac{\theta_{x,\max}}{2}\right)^2 + \frac{V_D + V_x + K_0}{2\sigma_x}} < \sqrt{\gamma_x \underline{\lambda}_{P_E}}\, \frac{r_e}{2}$$

which reads, after taking squares

$$\cancel{\left(\frac{\theta_{x,\max}}{2}\right)^2} + \frac{V_D + V_x + K_0}{2\sigma_x} < \left(\sqrt{\gamma_x \underline{\lambda}_{P_E}}\, \frac{r_e}{2}\right)^2 - 2\sqrt{\gamma_x \underline{\lambda}_{P_E}}\, \frac{r_e}{2}\frac{\theta_{x,\max}}{2} + \cancel{\left(\frac{\theta_{x,\max}}{2}\right)^2}$$

$$\frac{V_D + V_x + K_0}{\sigma_x} < k_x(r_e, \gamma_x) := \frac{\gamma_x \underline{\lambda}_{P_E}}{2} r_e^2 - \sqrt{\gamma_x \underline{\lambda}_{P_E}}\, \theta_{x,\max} r_e \tag{4.164}$$

Now, $K_0>0$ and $V_x>0$ can be made arbitrarily small by choosing the σ-modification gains appropriately. However $V_D>0$ is determined by the uncertainty bound D, which is considered to be fixed. Nevertheless, the left hand side of (4.164) can be made arbitrarily small by choosing σ_x sufficiently large. Hence the inequality is solved if $k_x(r_e, \gamma_x)>0$, which is in turn fulfilled, if γ_x is chosen sufficiently large

$$\gamma_x > \frac{4\theta_{x,\max}^2}{r_e^2\, \underline{\lambda}_{P_E}} =: \underline{\gamma}_x^{(0)} . \tag{4.165}$$

Moreover, if we substitute V_x in (4.164), we obtain a second condition for the σ-modification gains and K_0.

$$-2k_x(r_e, \gamma_x)\sigma_x + \theta_{N,\max}^2\sigma_N + \lambda_{L,\max}^2\sigma_L + 2K_0 < -2V_D \tag{4.166}$$

Analogously, for the remaining conditions, we require

$$\frac{V_D + V_N + K_0}{\sigma_N} < k_N(r_e, \gamma_N) := \frac{\gamma_N \underline{\lambda}_{P_E}}{2} r_e^2 - \sqrt{\gamma_N \underline{\lambda}_{P_E}}\, \theta_{N,\max} r_e \tag{4.167}$$

$$\frac{V_D+V_L+K_0}{\sigma_L} < k_L(r_e,\gamma_L) := \frac{\gamma_L \underline{\lambda}_{P_E}}{2} r_e^2 - \sqrt{\gamma_L \underline{\lambda}_{P_E}}\lambda_{L,\max} r_e \ . \tag{4.168}$$

which can be fulfilled if $k_N(r_e)>0$, $k_L(r_e)>0$, which in turn requires

$$\gamma_N > \frac{4\theta_{N,\max}^2}{r_e^2 \underline{\lambda}_{P_E}} =: \underline{\gamma}_N^{(0)} \tag{4.169}$$

$$\gamma_L > \frac{4\lambda_{L,\max}^2}{r_e^2 \underline{\lambda}_{P_E}} =: \underline{\gamma}_L^{(0)} \tag{4.170}$$

and, by substituting V_N,V_L in (4.167), (4.168), we obtain further conditions for the σ-modification parameters and K_0.

$$\theta_{x,\max}^2\sigma_x - 2k_N(r_e,\gamma_N)\sigma_N + \lambda_{L,\max}^2\sigma_L + 2K_0 < -2V_D \tag{4.171}$$

$$\theta_{x,\max}^2\sigma_x + \lambda_{N,\max}^2\sigma_N - 2k_L(r_e,\gamma_L)\sigma_L + 2K_0 < -2V_D \tag{4.172}$$

Summing up so far, if r_e, γ_x,, γ_N, γ_L simultaneously fulfill (4.162), (4.165), (4.169), (4.170), then $k_e(r_e)>0$, $k_x(r_e,\gamma_x)>0$, $k_N(r_e,\gamma_N)>0$ and $k_L(r_e,\gamma_L)>0$ which is necessary that the system of inequalities

$$\begin{array}{rrrrcr}
\theta_{x,\max}^2\cdot\sigma_x & +\theta_{N,\max}^2\cdot\sigma_N & +\lambda_{L,\max}^2\cdot\sigma_L & +2K_0 & < & 2k_e(r_e) \\
-2k_x(r_e,\gamma_x)\cdot\sigma_x & +\theta_{N,\max}^2\cdot\sigma_N & +\lambda_{L,\max}^2\cdot\sigma_L & +2K_0 & < & -2V_D \\
\theta_{x,\max}^2\cdot\sigma_x & -2k_N(r_e,\gamma_N)\cdot\sigma_N & +\lambda_{L,\max}^2\cdot\sigma_L & +2K_0 & < & -2V_D \\
\theta_{x,\max}^2\cdot\sigma_x & +\theta_{N,\max}^2\cdot\sigma_N & -2k_L(r_e,\gamma_L)\cdot\sigma_L & +2K_0 & < & -2V_D
\end{array} \tag{4.173}$$

is solved for some σ_x, σ_N, σ_L, $K_0>0$. However, $k_e(r_e)$ $k_i(r_e,\gamma_i)>0$, $i=x,N,L$, is only a necessary condition. For sufficiency, we potentially need, that $k_e(r_e)$ $k_i(r_e,\gamma_i)$ adopt positive values away from zero. However, if (4.173) has a solution for some k_e^*, k_i^*, then it has also a solution for $k_e \geq k_e^*, k_i \geq k_i^*$, which can be easily verified, if we rewrite (4.173).

$$\begin{array}{rrrrcr}
\theta_{x,\max}^2\cdot\sigma_x & +\theta_{N,\max}^2\cdot\sigma_N & +\lambda_{L,\max}^2\cdot\sigma_L & +2K_0 & < & 2k_e^* + 2(k_e - k_e^*) \\
-2k_x^*\cdot\sigma_x & +\theta_{N,\max}^2\cdot\sigma_N & +\lambda_{L,\max}^2\cdot\sigma_L & +2K_0 & < & -2V_D + 2(k_x - k_x^*)\sigma_x \\
\theta_{x,\max}^2\cdot\sigma_x & -2k_N^*\cdot\sigma_N & +\lambda_{L,\max}^2\cdot\sigma_L & +2K_0 & < & -2V_D + 2(k_N - k_N^*)\sigma_N \\
\theta_{x,\max}^2\cdot\sigma_x & +\theta_{N,\max}^2\cdot\sigma_N & -2k_L^*\cdot\sigma_L & +2K_0 & < & -2V_D + 2(k_L - k_L^*)\sigma_L
\end{array}$$

The left hand side and the first expressions of the right hand side solve (4.173) and since the second expressions of the right hand side are non-negative, the claim is obviously true. Consequently, with $k_e^*, k_i^* > 0$, we obtain increased lower bounds on r_e and γ_x, γ_N, γ_L, from (4.161), (4.164), (4.167), (4.168)

$$\underline{e}_e = 2\frac{\overline{\lambda}_{P_E}}{\underline{\lambda}_{Q_E}}\sqrt{\frac{\overline{\lambda}_{P_E}}{\underline{\lambda}_{P_E}}}D + \sqrt{\left(2\frac{\overline{\lambda}_{P_E}}{\underline{\lambda}_{Q_E}}\sqrt{\frac{\overline{\lambda}_{P_E}}{\underline{\lambda}_{P_E}}}D\right)^2 + \frac{4\overline{\lambda}_{P_E}k_e^*}{\underline{\lambda}_{Q_E}\underline{\lambda}_{P_E}}} \tag{4.174}$$

$$\underline{\gamma}_x = \frac{1}{r_e^2}\left(\frac{\theta_{x,\max}}{\sqrt{\underline{\lambda}_{P_E}}} + \sqrt{\left(\frac{\theta_{x,\max}}{\sqrt{\underline{\lambda}_{P_E}}}\right)^2 + \frac{2k_x^*}{\underline{\lambda}_{P_E}}}\right)^2 \tag{4.175}$$

$$\underline{\gamma}_N = \frac{1}{r_e^2}\left(\frac{\theta_{N,\max}}{\sqrt{\underline{\lambda}_{P_E}}} + \sqrt{\left(\frac{\theta_{N,\max}}{\sqrt{\underline{\lambda}_{P_E}}}\right)^2 + \frac{2k_N^*}{\underline{\lambda}_{P_E}}}\right)^2 \tag{4.176}$$

$$\underline{\gamma}_L = \frac{1}{r_e^2}\left(\frac{\lambda_{L,\max}}{\sqrt{\underline{\lambda}_{P_E}}} + \sqrt{\left(\frac{\lambda_{L,\max}}{\sqrt{\underline{\lambda}_{P_E}}}\right)^2 + \frac{2k_L^*}{\underline{\lambda}_{P_E}}}\right)^2 \tag{4.177}$$

Finally (4.173) is solved for some σ_x, σ_N, σ_L, $K_0>0$, if

$$r_e \geq \underline{e} \;\;,\;\; \gamma_x \geq \underline{\gamma}_x \;\;,\;\; \gamma_N \geq \underline{\gamma}_N \;\;,\;\; \gamma_L \geq \underline{\gamma}_L \tag{4.178}$$

and the system states are uniformly ultimately bounded. This result is summarized in the following.

Theorem 4.4 MRAC Variant 1 – Ultimate Boundedness

Consider:

- stable **reference dynamics** (4.56)
- **plant dynamics** in Byrnes-Isidori normal form (4.90), subjected to linearizing state feedback (4.125) with pseudo control (4.121), adaptive term (4.122), error feedback (4.54) and reference pseudo control (4.55)
- **parameterizations** (4.107) – (4.110) for state dependent uncertainty, (4.111) or (4.114) and for affine control effectiveness, (4.115) – (4.117) for nonaffine control effectiveness
- **parameter-estimation-errors** $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Theta}}_N$, $\tilde{\boldsymbol{\Lambda}}_L$, defined in (4.124), (4.119), (4.112)
- **update law** (4.131) using
  - symmetric positive definite solution $\mathbf{P}_E$ of Lyapunov equation (4.128) for some symmetric positive definite $\mathbf{Q}_E$
  - positive learning rates γ_x, γ_N, γ_L
  - projection operator with parameters ε_x, ε_N, ε_L,>0, $\theta_{x,max}$, $\theta_{N,max}$, $\lambda_{L,max}>0$ symmetric positive definite weighting matrices $\boldsymbol{\Gamma}_x$, $\boldsymbol{\Gamma}_N$, $\boldsymbol{\Gamma}_L$ and
  - switching σ-modification (4.143) with modification gains σ_x, σ_N, $\sigma_L>0$
- Sets: $\overline{\mathcal{B}_{\bar{\zeta}}} \subset \mathbb{R}^r$, $\mathcal{D}_\eta \subset \mathbb{R}^{n-r}$, $\mathcal{U} \subset \mathbb{R}^{p_N}$ $\mathcal{D} \subset \mathbb{R}^d$ for some $\bar{\zeta}>0$

The **tracking error** (4.126) and $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Theta}}_N$, $\tilde{\boldsymbol{\Lambda}}_L$ are **uniformly ultimately bounded**, if

- assumption A is fulfilled for some θ_x, θ_N, $\lambda_L>0$ and $\theta_{x,max}$, $\theta_{N,max}$, $\lambda_{L,max}$ are defined by (4.133) – (4.135)
- assumption B is fulfilled for some, $D \geq 0$
- $(\boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) \in \mathcal{D}_\eta \times \mathcal{U} \times \mathcal{D}$ for all $t \geq t_0$
- the reference model state is bounded according to (4.139) for some $\bar{\zeta}_R>0$
- r_e simultaneously satisfies $r_e \geq \underline{e}_e$ according to (4.178) and $r_e \leq \bar{\zeta} - \bar{\zeta}_R$ according to (4.151) where $\underline{e}_e$ is defined in (4.174) with k_e^*, k_x^*, k_N^*, k_L^* such that inequalities (4.173) are solved for σ_x, σ_N, σ_L, $K_0>0$
- the learning rates are chosen such that $\gamma_x \geq \underline{\gamma}_x, \gamma_N \geq \underline{\gamma}_N, \gamma_L \geq \underline{\gamma}_L$ according to (4.178) where $\underline{\gamma}_x, \underline{\gamma}_N, \underline{\gamma}_L$ are defined in (4.175) - (4.177)

- initial conditions: $\|\mathbf{e}(t_0)\|_2 \le \delta_e$, $\left\|\tilde{\boldsymbol{\Theta}}_x(t_0)\right\|_{\boldsymbol{\Gamma}_x} \le \delta_x$, $\left\|\tilde{\boldsymbol{\Theta}}_N(t_0)\right\|_{\boldsymbol{\Gamma}_N} \le \delta_N$, $\left\|\tilde{\boldsymbol{\Lambda}}_L^T(t_0)\right\|_{\boldsymbol{\Gamma}_L} \le \delta_L$
  - $\delta_e \in [0, \rho_e]$, $\delta_x \in [0, \rho_x]$, $\delta_N \in [0, \rho_N]$, $\delta_L \in [0, \rho_L]$
  - ρ_e, ρ_x, ρ_N, ρ_L defined in (4.155)

such that

- $\|\mathbf{e}(t)\|_2 \le r_e$ forall $t \ge t_0$ and $\|\mathbf{e}(t)\|_2 \le b_e$ forall $t \ge t_0 + T_e(b_e, \delta_e)$
- $\left\|\tilde{\boldsymbol{\Theta}}_x(t)\right\|_{\boldsymbol{\Gamma}_x} \le \sqrt{\gamma_x \underline{\lambda}_{P_E}}\, r_e$, $\left\|\tilde{\boldsymbol{\Theta}}_N(t)\right\|_{\boldsymbol{\Gamma}_N} \le \sqrt{\gamma_N \underline{\lambda}_{P_E}}\, r_e$, $\left\|\tilde{\boldsymbol{\Lambda}}_L^T(t)\right\|_{\boldsymbol{\Gamma}_L} \le \sqrt{\gamma_L \underline{\lambda}_{P_E}}\, r_e$ for all $t \ge t_0$
- $\left\|\tilde{\boldsymbol{\Theta}}_x(t)\right\|_{\boldsymbol{\Gamma}_x} \le b_x$ for all $t \ge t_0 + T_x(b_x, \delta_x)$, $\left\|\tilde{\boldsymbol{\Theta}}_N(t)\right\|_{\boldsymbol{\Gamma}_N} \le b_N$ for all $t \ge t_0 + T_N(b_N, \delta_N)$
  $\left\|\tilde{\boldsymbol{\Lambda}}_L^T(t)\right\|_{\boldsymbol{\Gamma}_L} \le b_L$ for all $t \ge t_0 + T_L(b_L, \delta_L)$

where

- $b_e = \sqrt{\underline{\lambda}_{P_E}^{-1} \nu}$, $b_x = \sqrt{\gamma_x \nu}$, $b_N = \sqrt{\gamma_N \nu}$, $b_L = \sqrt{\gamma_L \nu}$, $\nu = \overline{\lambda}_{P_E} \mu_e^2 + \gamma_x^{-1} \mu_x^2 + \gamma_N^{-1} \mu_N^2 + \gamma_L^{-1} \mu_L^2$
  $\mu_e, \mu_x, \mu_N, \mu_L$ defined in (4.158), (4.160)
- $T_e(b_e, \delta_e) = \max\left[0, K_0^{-1}\left(\overline{\lambda}_{P_E} \delta_e^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_e^2 - \nu\right)\right]$, $T_x(b_x, \delta_x) = \max\left[0, K_0^{-1}\left(\delta_x^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_e^2 - \nu\right)\right]$
  $T_N(b_N, \delta_N) = \max\left[0, K_0^{-1}\left(\delta_N^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_e^2 - \nu\right)\right]$, $T_L(b_L, \delta_L) = \max\left[0, K_0^{-1}\left(\delta_L^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_e^2 - \nu\right)\right]$

4.3.3 Conclusions

If the conditions of Theorem 4.4 are satisfied, particularly if there are r_e, D such that (4.136) is fulfilled and (4.173) is solved for some σ_x, σ_N, $\sigma_L > 0$, then the system states are uniformly ultimately bounded and it is assured that $||\mathbf{e}||_2 \le r_e$. I.e. the a priori assumptions (4.136), (4.137) are shown to be fulfilled for initial the conditions of Theorem 4.4. Potentially, even more restrictive bounds on the parameter-estimation-errors, as obtained from Theorem 4.4, can be established since the projection operator itself (Appendix D.1) restricts the parameter estimates to

$$\left\|\hat{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} \le \theta_{x,\max}, \quad \left\|\hat{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} \le \theta_{N,\max}, \quad \left\|\hat{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} \le \lambda_{L,\max}$$

and the ideal parameters are bounded by assumption A. Hence, by triangle inequality, we obtain

$$\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} \le \theta_{x,\max} + \theta_x, \quad \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} \le \theta_{N,\max} + \theta_N, \quad \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} \le \lambda_{L,\max} + \lambda_L. \tag{4.179}$$

Remark 1

If (4.173) cannot be solved for some fixed r_e, obviously constants $k_e(r_e)$, $k_x(r_e, \gamma_x)$, $k_N(r_e, \gamma_N)$ and $k_L(r_e, \gamma_L)$ are too small. While $k_x(r_e, \gamma_x)$, $k_N(r_e, \gamma_N)$, $k_L(r_e, \gamma_L)$ can be increased by increasing the learning rates, $k_e(r_e)$ is fixed by the parameter bounds $\theta_{x,max}$, $\theta_{N,max}$, $\lambda_{L,max}$ and V_D (which primarily depends on D), $k_e(r_e)$ can be increased by enlargement of r_e – the radius which defines the bound on the tracking error $\mathbf{e}$. This might be undesirable, since the maximum tracking error is a quality feature for control systems, or even impossible if this leads to $r_e + \zeta_R > \overline{\zeta}$ (refer to equations (4.136), (4.137), (4.139), (4.151)). Then, enlargement of the maximum tracking error r_e is potentially for the cost of a

larger $\overline{\zeta}$ which, in turn, increases D according to assumption B as sketched in Figure 4.12.

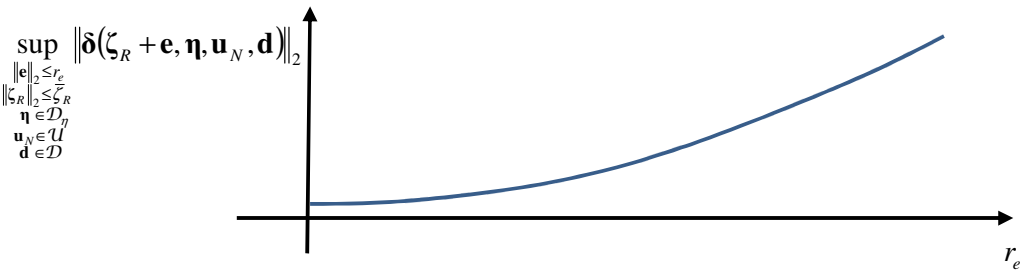


Figure 4.12 Relationship between Tracking Error and Magnitude of Unmatched Uncertainty

If it is not possible to increase the tracking error bound, other measures have to be taken. Referring to inequalities (4.173), smaller parameter bounds $\theta_{x,max}$, $\theta_{N,max}$, $\lambda_{L,max}$ potentially lead to a smaller k_e^* and to smaller $\underline{e}_e$, according to equation (4.174). Referring to section 4.2.2, since the unknown parameters are in fact the deviation between real dynamics and the model, used for the linearizing state feedback, a more accurate model could reduce the parameter bounds. Besides cancellation of model deviations, an adaptive controller should also be able to compensate for damages and failure cases. This however requires higher bounds on the parameters, which depend on the failure cases, the adaptive scheme should compensate for.

Summing up, if we require the adaptive control system to have a minimum quality (w.r.t. the maximum tracking error that can occur) it is a matter of making only such demands on the adaptive scheme – i.e. to which extend should it be able to compensate for model deviations as well as for failure cases – it can fulfill. In descriptive words, if boundedness cannot be shown for the required tracking accuracy, expectations on the adaptive scheme have to be scaled down by reducing $\theta_{x,max}$, $\theta_{N,max}$, $\lambda_{L,max}$, D such that (4.173) has a solution.

Remark 2

Theorem 4.4 assures ultimate boundedness of the tracking error with an ultimate bound $b_e<r_e$ and $||\mathbf{e}||_2 \le r_e$ for all $t \ge t_0$, i.e. it will not leave the valid set where assumption B is valid, if the initial conditions are chosen appropriately. Additionally, the nonlinear controls $\mathbf{u}_N$ and external disturbances $\mathbf{d}$ have to be restricted to their valid domains. As proposed in section 4.1.7 the nonaffine controls are not used for the linearizing state feedback, but they are set to an external value. It is hence no problem to choose the external value such that $\mathbf{u}_N \in \mathcal{U}$. The external disturbances, of course, cannot be prescribed but it is a matter of environmental conditions, under which the system is operated.

The most disputable condition is the restriction of the internal state in (4.90) to $\mathcal{D}_\eta$. The internal dynamics are excited by the external states and, if not input-

normalized, by the controls. For minimum phase systems, it is assured, that the internal dynamics are at least locally stable, but the there is no explicit statement on what “local” means, i.e. how far is the system is allowed to depart from equilibrium such that it is still stable. This remains an open point, but possibly, considerations of input-to state stability (for a definition refer to [Kha02]) for the internal dynamics could provide a solution to this point. However, this is left for future work. Nevertheless, if the system does not have internal states at all, this condition is trivially fulfilled and can be dropped.

Particularly in the flight control algorithms, proposed subsequently, the controlled variables, i.e. the external state, are the body-fixed angular rates. In this case, the internal states are

- velocity V
- angle of attack α
- angle of side-slip β
- bank angle: Φ
- pitch angle Θ

The most unproblematic one is the aircraft attitude, since the attitude is bounded anyway due to its cyclic character. Either a pilot or an additional speed controller could prevent the aircraft from leaving the valid velocity range. More problematic are α and β, since their dynamic bandwidth is in the same region as the controlled variables. However, an additional control cascade for α and β, including a limitation could restrict the variables to the valid domain. Yet, this not a rigorous argumentation and a mathematical proof remains an open item.

4.4 Multi Model Q Modification

A common problem in MRAC architectures, using the standard learning law is poor adaptation performance. Higher learning rates accelerate adaptation to a certain extent but it is known that high gains result in high frequency oscillation and reduced robustness. This fact initiated plenty of publications since beginning of the millennium that are concerned with improvement of adaptation performance. Most of them utilize modifications of the learning law ([Ngu08], [Yuc09], [Yuc091]). Referring to (4.131), the learning laws take the form of a dyadic product, i.e. they are composed by a column vector ($\boldsymbol{\Gamma}_x\boldsymbol{\varphi}$) and a row vector ($\mathbf{e}^T\mathbf{PH}$). This is also known as the rank-1 property of the standard learning law, i.e. the parameter update only takes place into the direction of $\boldsymbol{\Gamma}_x\boldsymbol{\varphi}$. Modifications that should improve performance aim to increase the rank of the update law; a promising one is q-modification ([Vol061], [Vol06] [Yuc092], [Vol09]).

4.4.1 Recapitulation on Single Model q Modification

In classical q-modification, the dynamic equations are utilized to gain a filtered version of the model uncertainty. With external dynamics (4.120), we obtain

$$\begin{aligned}\dot{\boldsymbol{\zeta}}-\mathbf{J}\boldsymbol{\zeta}+\mathbf{H}\left(-\mathbf{v}+\hat{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})+\mathbf{B}_L(\boldsymbol{\zeta},\boldsymbol{\eta})\hat{\boldsymbol{\Lambda}}_L\mathbf{u}\right)\\ =\mathbf{H}\left(\boldsymbol{\Theta}_x^T\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})+\boldsymbol{\Theta}_N^T\boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})+\mathbf{B}_L(\boldsymbol{\zeta},\boldsymbol{\eta})\boldsymbol{\Lambda}_L\mathbf{u}+\boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right)\end{aligned} \tag{4.180}$$

where the left hand side of the equation only contains quantities that either are known or can be measured, while the right hand side contains the unknown parametric uncertainty. Problematic in our case is $\dot{\boldsymbol{\zeta}}$, since sensors for measurements of angular acceleration are rather expensive. Nevertheless if it can be measured or estimated

$$\mathbf{c}_0:=\dot{\boldsymbol{\zeta}}-\mathbf{J}\boldsymbol{\zeta}+\mathbf{H}\left(-\mathbf{v}+\hat{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})+\mathbf{B}_L(\boldsymbol{\zeta},\boldsymbol{\eta})\hat{\boldsymbol{\Lambda}}_L\mathbf{u}\right) \tag{4.181}$$

provides an expression for the unknown uncertainty. If it cannot be measured or estimated, at least a filtered version of (4.181), using a filter described by a strictly proper and stable transfer function, could be gained.

$$G(s)=\frac{b_m s^m+\ldots+b_1 s+b_0}{s^n+a_{n-1}s^{n-1}\ldots+a_1 s+a_0} \tag{4.182}$$

Thereby $m<n$ and then the filtered uncertainty is obtained by

$$\mathbf{c}(s):=G(s)(s\mathbf{I}-\mathbf{J})\boldsymbol{\zeta}(s)+G(s)\mathbf{H}\left[-\mathbf{v}(s)+\left[\hat{\boldsymbol{\Theta}}_N^T\boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right](s)+\left[\mathbf{B}_L(\boldsymbol{\zeta},\boldsymbol{\eta})\hat{\boldsymbol{\Lambda}}_L\mathbf{u}\right](s)\right]. \tag{4.183}$$

Note that $m<n$ implies that $G(s)(s\mathbf{I}\text{-}\mathbf{J})$ is a proper transfer function ($m\leq n$). In order to keep expressions readable, we will, in a slight abuse of notation, switch between time- and frequency domain notation. Before continuing, an additional assumption has to be imposed, namely $\mathbf{B}_L(\boldsymbol{\zeta},\boldsymbol{\eta})$ is split up into a constant matrix part $\mathbf{B}_L$ and a scalar state dependent part $b(\boldsymbol{\zeta},\boldsymbol{\eta})$.

$$\mathbf{B}_L(\boldsymbol{\zeta},\boldsymbol{\eta})=b(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{B}_L \tag{4.184}$$

This special structure is also physically justified for aircraft, since effectiveness of all aerodynamic control surfaces are mainly dependent on dynamic pressure and thus the introduced parameterization describes the physics with $b(\boldsymbol{\zeta},\boldsymbol{\eta})=\rho V^2$ (ρ: air density, V: aerodynamic velocity). If this assumption holds and, in view of the right hand side of (4.180), we obtain

$$\begin{aligned}\mathbf{c}(s)&=G(s)\mathbf{H}\left(\left[\boldsymbol{\Theta}_x^T\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})\right](s)+\left[\boldsymbol{\Theta}_N^T\boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right](s)+\left[b(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{B}_L\boldsymbol{\Lambda}_L\mathbf{u}\right](s)+\left[\boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right](s)\right)\\ &=\mathbf{H}\left(\boldsymbol{\Theta}_x^T G(s)\left[\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})\right](s)+\boldsymbol{\Theta}_N^T\left[G(s)\boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right](s)+\mathbf{B}_L\boldsymbol{\Lambda}_L G(s)\left[b(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{u}\right](s)+\left[\boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right](s)\right)\\ \mathbf{c}(s)&=\mathbf{H}\left(\boldsymbol{\Theta}_x^T\mathbf{q}_x(s)+\boldsymbol{\Theta}_N^T\mathbf{q}_N(s)+\mathbf{B}_L\boldsymbol{\Lambda}_L\mathbf{q}_L(s)+\mathbf{q}_\delta(s)\right)\end{aligned} \tag{4.185}$$

where

$$\begin{aligned}\mathbf{q}_x(s) &= G(s)[\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})](s) \quad & \mathbf{q}_N(s) &= G(s)[\boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})](s)\\ \mathbf{q}_L(s) &= G(s)[b(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{u}](s) \quad & \mathbf{q}_\delta(s) &= G(s)[\boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})](s)\end{aligned} \tag{4.186}$$

are the filtered versions of regressors and unmatched uncertainty. Notice that we have made use of the fact that transfer functions are linear operators such that constants can be posed in front of the transfer function. For a compact notation, we further define

$$\mathbf{q}^T(s) = \begin{pmatrix}\mathbf{q}_x^T(s) & \mathbf{q}_N^T(s) & \mathbf{q}_L^T(s)\end{pmatrix} \text{ and } \mathbf{W}^T = \begin{pmatrix}\boldsymbol{\Theta}_x^T & \boldsymbol{\Theta}_N^T & \mathbf{B}_L\boldsymbol{\Lambda}_L\end{pmatrix} \tag{4.187}$$

and (4.185) also reads as

$$\mathbf{c}(t) = \mathbf{H}\left(\mathbf{W}^T\mathbf{q}(t) + \mathbf{q}_\delta(t)\right) \tag{4.188}$$

On the other hand, an analog of the filtered estimated uncertainty can be computed since the regressors and in turn, its filtered versions are known

$$\hat{\mathbf{c}}(t) = \mathbf{H}\,\hat{\mathbf{W}}^T(t)\mathbf{q}(t) \tag{4.189}$$

where $\hat{\mathbf{W}}(t)$ denotes the estimated compound parameter analogous to (4.187). Using (4.188), (4.189) we obtain the filtered uncertainty estimation error.

$$\tilde{\mathbf{c}}(t) := \hat{\mathbf{c}}(t) - \mathbf{c}(t) = \mathbf{H}\left(\tilde{\mathbf{W}}^T(t)\mathbf{q}(t) - \mathbf{q}_\delta(t)\right) \tag{4.190}$$

where $\tilde{\mathbf{W}}(t) = \hat{\mathbf{W}}(t) - \mathbf{W}$ is the compound parameter-estimation-error. q-modification uses this additional information to improve adaptation performance. Therefore, we try to minimize a quadratic cost function that evaluates the quality of the parameter estimation.

$$\begin{aligned}J(\tilde{\mathbf{W}}) &= \frac{1}{2}\tilde{\mathbf{c}}^T\tilde{\mathbf{c}}\\ &= \frac{1}{2}\mathbf{q}^T\tilde{\mathbf{W}}\mathbf{H}^T\mathbf{H}\tilde{\mathbf{W}}^T\mathbf{q} - \mathbf{q}_\delta^T\mathbf{H}^T\mathbf{H}\tilde{\mathbf{W}}^T\mathbf{q} + \frac{1}{2}\mathbf{q}_\delta^T\mathbf{H}^T\mathbf{H}\mathbf{q}_\delta\\ J(\tilde{\mathbf{W}}) &= \frac{1}{2}\mathbf{q}^T\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T\mathbf{q} - \mathbf{q}_\delta^T\tilde{\mathbf{W}}^T\mathbf{q} + \frac{1}{2}\mathbf{q}_\delta^T\mathbf{q}_\delta\end{aligned} \tag{4.191}$$

Here we have used $\mathbf{H}^T\mathbf{H} = \mathbf{I}$. In order to minimize J we compute the Jacobians w.r.t. $\boldsymbol{\Theta}_x$, $\boldsymbol{\Theta}_N$, $\boldsymbol{\Lambda}_L$.

$$\frac{\partial J(\tilde{\mathbf{W}})}{\partial\hat{\boldsymbol{\Theta}}_x} = \mathbf{q}_x\tilde{\mathbf{c}}^T\mathbf{H} \qquad \frac{\partial J(\tilde{\mathbf{W}})}{\partial\hat{\boldsymbol{\Theta}}_N} = \mathbf{q}_N\tilde{\mathbf{c}}^T\mathbf{H} \qquad \frac{\partial J(\tilde{\mathbf{W}})}{\partial\hat{\boldsymbol{\Lambda}}_L} = \mathbf{B}_L^T\mathbf{H}^T\tilde{\mathbf{c}}\mathbf{q}_L^T \tag{4.192}$$

Auxiliary Calculation

The computation of the Jacobians is done, using tensor notation, i.e. a matrix is a second order tensor, written as $\boldsymbol{\Theta} = \theta_{ij}$, Einstein's notation, i.e. if some index appears twice in an expression means the sum is taken over that index:

$$\theta_{ij} q_j = \sum_j \theta_{ij} q_j$$

Cronecker delta, a second order tensor, such that

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{else} \end{cases}$$

and the fact that a matrix, derived w.r.t. to a matrix yields a forth order tensor of the form

$$\frac{\partial \theta_{ij}}{\partial \theta_{pq}} = \delta_{ip} \delta_{jq} .$$

Then, exemplarily, we compute the first Jacobian.

$$\frac{\partial J}{\partial \hat{\boldsymbol{\Theta}}_x} = \frac{\partial J}{\partial \hat{\theta}_{x,pq}} = \frac{\partial J}{\partial \hat{\theta}_{x,pq}} \left(\frac{1}{2} \tilde{c}_i^T \tilde{c}_i \right) = \tilde{c}_i \frac{\partial \tilde{c}_i}{\partial \hat{\theta}_{x,pq}}$$

Inserting $\tilde{c}_i$ yields:

$$\begin{aligned} \frac{\partial J}{\partial \hat{\boldsymbol{\Theta}}_x} &= \tilde{c}_i \frac{\partial J}{\partial \hat{\theta}_{x,pq}} \left(h_{ij} \left(\left(\hat{\theta}_{x,kj} - \theta_{x,kj} \right) q_{x,k} + \left(\hat{\theta}_{N,kj} - \theta_{N,kj} \right) q_{N,k} + b_{l,jk} \left(\lambda_{L,kl} - \lambda_{L,kl} \right) q_{x,l} - q_{\delta,j} \right) \right) \\ &= \tilde{c}_i h_{ij} \delta_{kp} \delta_{jq} q_{x,k} = \tilde{c}_i h_{iq} q_{x,p} = \mathbf{q}_x \tilde{\mathbf{c}}^T \mathbf{H} \end{aligned}$$

The negative Jacobians point into the directions, the respective parameter should change such that J is minimized. This motivates the incorporation of these terms into the learning laws. Additionally to the σ-modification term (4.143), which introduces robustness, q-modification is added to the learning law (4.131) for improved performance.

$$\begin{aligned} \mathbf{M}_x &= \sigma_x \bar{f}_{\varepsilon_x, \theta_x, \boldsymbol{\Gamma}_x} \left(\hat{\boldsymbol{\Theta}}_x \right) \hat{\boldsymbol{\Theta}}_x + \kappa \boldsymbol{\Gamma}_x \mathbf{q}_x \tilde{\mathbf{c}}^T \mathbf{H} \\ \mathbf{M}_N &= \sigma_N \bar{f}_{\varepsilon_N, \theta_N, \boldsymbol{\Gamma}_N} \left(\hat{\boldsymbol{\Theta}}_N \right) \hat{\boldsymbol{\Theta}}_N + \kappa \boldsymbol{\Gamma}_N \mathbf{q}_N \tilde{\mathbf{c}}^T \mathbf{H} \\ \mathbf{M}_L &= \sigma_L \bar{f}_{\varepsilon_L, \lambda_L, \boldsymbol{\Gamma}_L} \left(\hat{\boldsymbol{\Lambda}}_L^T \right) \hat{\boldsymbol{\Lambda}}_L + \kappa \mathbf{B}_L^T \mathbf{H}^T \tilde{\mathbf{c}} \mathbf{q}_L^T \boldsymbol{\Gamma}_L \end{aligned} \tag{4.193}$$

where κ is the modification gain. The q-modification terms introduces an additional direction to the standard update law, which is aligned with the respective filtered regressor. Hence q-modification renders the parameter update rank 2, if the directions of standard- and q-modification law are linearly independent. Let's take a look on the Lyapunov analysis. Using (4.193), (4.144) becomes

$$\begin{aligned} \dot{V} = -\mathbf{e}^T \mathbf{Q}_E \mathbf{e} \quad & -2\mathbf{e}^T \mathbf{P}_E \mathbf{H} \boldsymbol{\delta} - 2\kappa \operatorname{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T \mathbf{q}_x \tilde{\mathbf{c}}^T \mathbf{H} - \tilde{\boldsymbol{\Theta}}_N^T \mathbf{q}_N \tilde{\mathbf{c}}^T \mathbf{H} - \tilde{\boldsymbol{\Lambda}}_L \mathbf{q}_L \tilde{\mathbf{c}}^T \mathbf{H} \mathbf{B}_L \right] \\ & -2\sigma_x \bar{f}\left(\hat{\boldsymbol{\Theta}}_x \right) \operatorname{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T \boldsymbol{\Gamma}_x^{-1} \hat{\boldsymbol{\Theta}}_x \right] - 2\sigma_N \bar{f}\left(\hat{\boldsymbol{\Theta}}_N \right) \operatorname{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T \boldsymbol{\Gamma}_N^{-1} \hat{\boldsymbol{\Theta}}_N \right] - 2\sigma_L \bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T \right) \operatorname{tr}\left[\tilde{\boldsymbol{\Lambda}}_L \boldsymbol{\Gamma}_L^{-1} \hat{\boldsymbol{\Lambda}}_L^T \right] \end{aligned}$$

where the expression in dashed lines appears due to q-modification. Using (4.113), (4.119), (4.124), (4.190) and cyclic property of trace operator, we further obtain:

$$\dot{V} \leq -\mathbf{e}^T\mathbf{Q}_E\mathbf{e} \quad -2\kappa \mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\mathbf{q}_x\overbrace{\left(\mathbf{q}^T\tilde{\mathbf{W}}-\mathbf{q}_\delta^T\right)\mathbf{H}^T}^{\tilde{\mathbf{c}}^T}\mathbf{H}+\tilde{\boldsymbol{\Theta}}_N^T\mathbf{q}_N\overbrace{\left(\mathbf{q}^T\tilde{\mathbf{W}}-\mathbf{q}_\delta^T\right)\mathbf{H}^T}^{\tilde{\mathbf{c}}^T}\mathbf{H}+\tilde{\boldsymbol{\Lambda}}_L\mathbf{q}_L\overbrace{\left(\mathbf{q}^T\tilde{\mathbf{W}}-\mathbf{q}_\delta^T\right)\mathbf{H}^T}^{\tilde{\mathbf{c}}^T}\mathbf{H}\,\mathbf{B}_L\right]$$

$$-2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta} \quad -2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2+\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\boldsymbol{\Theta}_x\right]\right)-2\sigma_N\bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2+\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\Gamma}_N^{-1}\boldsymbol{\Theta}_N\right]\right)$$

$$-2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2+\mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\boldsymbol{\Lambda}_L^T\right]\right)$$

$$\dot{V} \leq -\mathbf{e}^T\mathbf{Q}_E\mathbf{e} \quad -2\kappa \mathrm{tr}\left[\overbrace{\left(\tilde{\boldsymbol{\Theta}}_x^T\mathbf{q}_x+\tilde{\boldsymbol{\Theta}}_N^T\mathbf{q}_N+\mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\mathbf{q}_L\right)}^{\tilde{\mathbf{W}}^T\mathbf{q}}\mathbf{q}^T\tilde{\mathbf{W}}-\overbrace{\left(\tilde{\boldsymbol{\Theta}}_x^T\mathbf{q}_x+\tilde{\boldsymbol{\Theta}}_N^T\mathbf{q}_N+\mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\mathbf{q}_L\right)}^{\tilde{\mathbf{W}}^T\mathbf{q}}\mathbf{q}_\delta^T\right]$$

$$-2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta} \quad -2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2+\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\boldsymbol{\Theta}_x\right]\right)-2\sigma_N\bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2+\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\Gamma}_N^{-1}\boldsymbol{\Theta}_N\right]\right)$$

$$-2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2+\mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\boldsymbol{\Lambda}_L^T\right]\right)$$

$$\begin{aligned}\dot{V} \leq -\mathbf{e}^T\mathbf{Q}_E\mathbf{e} \quad & -2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta}-2\kappa\mathrm{tr}\left[\tilde{\mathbf{W}}^T\mathbf{q}\,\mathbf{q}^T\tilde{\mathbf{W}}\right]+2\kappa\mathrm{tr}\left[\tilde{\mathbf{W}}^T\mathbf{q}\,\mathbf{q}_\delta^T\right]\\ & -2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2+\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\boldsymbol{\Theta}_x\right]\right)-2\sigma_N\bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2+\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\Gamma}_N^{-1}\boldsymbol{\Theta}_N\right]\right)\\ & -2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2+\mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\boldsymbol{\Lambda}_L^T\right]\right)\end{aligned} \tag{4.194}$$

Using matrix induced norms, bounds on quadratic forms (Theorem B.20), assumption A with equations (4.133) – (4.135), as well as Cauchy-Schwartz inequality, we obtain an upper bound on $\dot{V}$.

$$\begin{aligned}\dot{V} \leq & -\underline{\lambda}_{Q_E}\|\mathbf{e}\|_2^2 \quad +2\bar{\lambda}_{P_E}\|\boldsymbol{\delta}\|_2\|\mathbf{e}\|_2\\ & -2\kappa\left\|\tilde{\mathbf{W}}^T\mathbf{q}\right\|_2^2+2\kappa\left\|\tilde{\mathbf{W}}^T\mathbf{q}\right\|_2\|\mathbf{q}_\delta\|_2\\ & -2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2-\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\theta_{x,\max}\right)-2\sigma_N\bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2-\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\theta_{N,\max}\right)\\ & -2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2-\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\lambda_{L,\max}\right)\end{aligned} \tag{4.195}$$

Remark 1

It reveals that, in case of perfectly matched uncertainty, i.e. $||\mathbf{q}_\delta||=0$, the q-modification term introduces an additional negative definite term $-2\kappa(||\tilde{\mathbf{W}}^T\mathbf{q}||_2)^2$ which indicates that the adaptation is accelerated. This is considered as the key advantage of q-modification. However in presence of unmatched uncertainty, the bound on $\dot{V}$ is even increased by the positive term $2\kappa(||\tilde{\mathbf{W}}^T\mathbf{q}||_2||\mathbf{q}_\delta||_2)$, i.e. adaptation performance could even be deteriorated if the unmatched uncertainty is dominant. Explained physically, the dynamic equation (4.180), only

allows the computation of matched and unmatched uncertainty at once, and there is no possibility to distinguish between matched an unmatched part (refer to (4.190)). Hence, the cost function (4.191) tries to minimize a combination of both; however, adaptation only can compensate for the matched part, which leads to a contradiction.

Remark 2

It is also evident that potentially $\tilde{\mathbf{W}}^T\mathbf{q}=\mathbf{0}$ even if $\tilde{\mathbf{W}}\neq\mathbf{0}$, which is the case, if $\mathbf{q}$ is in the null space of $\tilde{\mathbf{W}}$. I.e., even if parameters have not converged, q-modification does not guarantee an additional negative definite term in $\dot{V}$. Considering the associated expression in (4.194), this is the case, since $\mathbf{q}\mathbf{q}^T$ is a rank one matrix and hence only positive semi-definite. If this matrix is positive definite, we would at minimum have a an additional negative term $-2\kappa\underline{\lambda}_{\mathbf{q}\mathbf{q}^T}\left\|\tilde{\mathbf{W}}\right\|_F^2$, which suffices to show exponential stability in absence of unmatched uncertainties ([Vol06]), while in presence of unmatched uncertainties, exponential convergence of the system states to a bounded set could be shown. The background learning / concurrent learning approach ([Cho10], [Cho09], [Joh04]) utilizes this fact.

4.4.2 Recapitulation on Concurrent Learning

In 2010, Chowdhary developed a modification to adaptive schemes that guarantee convergence of parameters even if the system is not persistently excited ([Cho102]). It has been developed for single hidden layer neural network based adaptation schemes, but also for linearly parameterized uncertainties ([Joh04], [Cho09], [Cho10], [Cho101]). Similarly to q-modification, the concurrent learning approach also utilizes the system equation (4.120) to obtain an expression for the uncertainty. Contrary to q-modification, they are not computed online, using a filter but, by using a stack of recorded data. For past time instances t_i, $i=1,\ldots,N$, $\mathbf{c}_0(t_i)$ is computed according to (4.181). Thereby the state derivative is either measured, or estimated.

$$\mathbf{c}_0(t_i) := \dot{\zeta}_i - \mathbf{J}\zeta_i + \mathbf{H}\left(-\mathbf{v}_i + \hat{\Theta}_{N,i}^T\boldsymbol{\omega}\left(\zeta_i,\boldsymbol{\eta}_i,\mathbf{u}_{N,i},\mathbf{d}_i\right) + \mathbf{B}_L(\zeta_i,\boldsymbol{\eta}_i)\hat{\boldsymbol{\Lambda}}_{L,i}\mathbf{u}_i\right) \tag{4.196}$$

Analogously to (4.189) and (4.292)

$$\hat{\mathbf{c}}_0(t_i) = \mathbf{H}\left(\hat{\boldsymbol{\Theta}}_x^T(t)\boldsymbol{\varphi}(\zeta_i,\boldsymbol{\eta}_i) + \hat{\boldsymbol{\Theta}}_{Ni}^T(t)\boldsymbol{\varphi}\left(\zeta_i,\boldsymbol{\eta}_i,\mathbf{u}_{N,i},\mathbf{d}_i\right) + \mathbf{B}_L\hat{\boldsymbol{\Lambda}}_L(t)b(\zeta_i,\boldsymbol{\eta}_i)\mathbf{u}_i\right) \tag{4.197}$$

$$\begin{aligned}\tilde{\mathbf{c}}_0(t_i) &= \hat{\mathbf{c}}_0(t_i) - \mathbf{c}_0(t_i)\\ &= \tilde{\boldsymbol{\Theta}}_x^T(t)\boldsymbol{\varphi}(\zeta_i,\boldsymbol{\eta}_i) + \tilde{\boldsymbol{\Theta}}_N^T(t)\boldsymbol{\varphi}\left(\zeta_i,\boldsymbol{\eta}_i,\mathbf{u}_{N,i},\mathbf{d}_i\right) + \mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L(t)b(\zeta_i,\boldsymbol{\eta}_i)\mathbf{u}_i - \boldsymbol{\delta}\left(\zeta_i,\boldsymbol{\eta}_i,\mathbf{u}_{N,i},\mathbf{d}_i\right)\end{aligned} \tag{4.198}$$

where we abbreviated $\boldsymbol{\zeta}(t_i)=\boldsymbol{\zeta}_i,\ldots$ for a compact notation. Note that $\hat{\mathbf{c}}_0(t_i)$ is computed with the current parameter estimates $\boldsymbol{\Theta}_x(t),\ldots$ while the regressors $\boldsymbol{\varphi}(\boldsymbol{\zeta}_i,\boldsymbol{\eta}_i),\ldots$ are taken at the respective recorded time instant t_i. It can hence be interpreted as estimated uncertainty, if the plant is in a state, it was at time t_i, but evaluated with the current

parameter estimates. With definition of the compound parameter $\mathbf{W}$ in (4.187), and the compound regressor

$$\boldsymbol{\sigma}^T(\zeta,\boldsymbol{\eta},\mathbf{u}_N\mathbf{d}) = \begin{pmatrix}\boldsymbol{\varphi}^T(\zeta,\boldsymbol{\eta}) & \boldsymbol{\omega}^T(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\boldsymbol{\delta}) & b(\zeta,\boldsymbol{\eta})\mathbf{u}^T\end{pmatrix} \tag{4.199}$$

we arrive at

$$\tilde{\mathbf{c}}_0(t_i) = \mathbf{H}\left(\tilde{\mathbf{W}}^T\boldsymbol{\sigma}_i - \boldsymbol{\delta}_i\right). \tag{4.200}$$

If we add the Jacobi of a cost function, similarly to (4.191), to the update law (4.131), we obtain

$$\begin{aligned}\mathbf{M}_x &= \sigma_x \bar{f}_{\varepsilon_x,\theta_x,\boldsymbol{\Gamma}_x}\left(\hat{\boldsymbol{\Theta}}_x\right)\hat{\boldsymbol{\Theta}}_x + \kappa\boldsymbol{\Gamma}_x\sum_{i=1}^{N}\boldsymbol{\varphi}(t_i)\tilde{\mathbf{c}}_0^T(t_i)\mathbf{H}\\ \mathbf{M}_N &= \sigma_N \bar{f}_{\varepsilon_N,\theta_N,\boldsymbol{\Gamma}_N}\left(\hat{\boldsymbol{\Theta}}_N\right)\hat{\boldsymbol{\Theta}}_N + \kappa\boldsymbol{\Gamma}_N\sum_{i=1}^{N}\boldsymbol{\omega}(t_i)\tilde{\mathbf{c}}_0^T(t_i)\mathbf{H}\\ \mathbf{M}_L &= \sigma_L \bar{f}_{\varepsilon_L,\lambda_L,\boldsymbol{\Gamma}_L}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\hat{\boldsymbol{\Lambda}}_L + \kappa\sum_{i=1}^{N}\mathbf{B}_L^T\mathbf{H}^T\tilde{\mathbf{c}}_0(t_i)b(t_i)\mathbf{u}^T(t_i)\boldsymbol{\Gamma}_L\end{aligned} \tag{4.201}$$

where we have also applied the switching σ-modification term for robustification. If we define matrices, where the regressors, associated with respective time instances are stacked into their columns, i.e.

$$\boldsymbol{\Phi} = \begin{bmatrix}\boldsymbol{\varphi}(t_1) & \cdots & \boldsymbol{\varphi}(t_N)\end{bmatrix},\ \boldsymbol{\Omega} = \begin{bmatrix}\boldsymbol{\omega}(t_1) & \cdots & \boldsymbol{\omega}(t_N)\end{bmatrix},\ \mathbf{U} = \begin{bmatrix}b(t_1)\mathbf{u}(t_1) & \cdots & b(t_N)\mathbf{u}(t_N)\end{bmatrix} \tag{4.202}$$

and a matrix whose columns contain $\tilde{\mathbf{c}}_0(t_i)$

$$\tilde{\mathbf{C}}_0 = \begin{bmatrix}\tilde{\mathbf{c}}_0(t_1) & \cdots & \tilde{\mathbf{c}}_0(t_N)\end{bmatrix} \tag{4.203}$$

(4.201) can also be written in compact matrix-vector notation.

$$\begin{aligned}\mathbf{M}_x &= \sigma_x \overline{f}_{\varepsilon_x,\theta_x,\boldsymbol{\Gamma}_x}\left(\hat{\boldsymbol{\Theta}}_x\right)\hat{\boldsymbol{\Theta}}_x + \kappa\boldsymbol{\Gamma}_x\boldsymbol{\Phi}\tilde{\mathbf{C}}_0\mathbf{H}\\ \mathbf{M}_N &= \sigma_N \overline{f}_{\varepsilon_N,\theta_N,\boldsymbol{\Gamma}_N}\left(\hat{\boldsymbol{\Theta}}_N\right)\hat{\boldsymbol{\Theta}}_N + \kappa\boldsymbol{\Gamma}_N\boldsymbol{\Omega}\tilde{\mathbf{C}}_0\mathbf{H}\\ \mathbf{M}_L &= \sigma_L \overline{f}_{\varepsilon_L,\lambda_L,\boldsymbol{\Gamma}_L}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\hat{\boldsymbol{\Lambda}}_L + \kappa\mathbf{B}_L^T\mathbf{H}^T\tilde{\mathbf{C}}_0\mathbf{U}^T\boldsymbol{\Gamma}_L\end{aligned} \tag{4.204}$$

Comparing to q-modification (4.193), concurrent learning seems to be very similar. The filtered regressor vector is replaced by a matrix, whose columns contain the unfiltered recorded regressors $\boldsymbol{\Phi}$,… and the filtered estimation error is replaced by a matrix whose columns contain the unfiltered recorded estimation errors $\tilde{\mathbf{C}}_0$. Lyapunov analysis continues along similar. Equation (4.194) becomes

$$\begin{aligned}\dot{V} \le -\mathbf{e}^T\mathbf{Q}_E\mathbf{e} \quad & -2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta} - 2\kappa\mathrm{tr}\left[\tilde{\mathbf{W}}^T\boldsymbol{\Sigma}\boldsymbol{\Sigma}^T\tilde{\mathbf{W}}\right] + 2\kappa\mathrm{tr}\left[\tilde{\mathbf{W}}^T\boldsymbol{\Sigma}\,\boldsymbol{\Delta}^T\right]\\ & -2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 + \mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\boldsymbol{\Theta}_x\right]\right) - 2\sigma_N\bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 + \mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\Gamma}_N^{-1}\boldsymbol{\Theta}_N\right]\right)\\ & -2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 + \mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\boldsymbol{\Lambda}_L^T\right]\right)\end{aligned} \tag{4.205}$$

where

$$\boldsymbol{\Sigma}^T = \begin{bmatrix} \boldsymbol{\Phi}^T & \boldsymbol{\Omega}^T & \mathbf{U}^T \end{bmatrix} \tag{4.206}$$

$$\boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\delta}(t_1) & \cdots & \boldsymbol{\delta}(t_N) \end{bmatrix}. \tag{4.207}$$

Considering (4.205), the quadratic term of $\tilde{\mathbf{W}}$ has changed, compared to q-modification, such that $\Sigma\Sigma^T$ is no longer subjected to a rank-1 condition. Actually, if the history stack contains more recorded time instances than the total number of regressors, i.e. $N \geq s_x + s_N + p$ and if the time instances t_i are chosen such that $\boldsymbol{\Sigma}$ contains at least s_x+x_N+p linearly independent columns, then $\Sigma\Sigma^T$ is positive definite, whose smallest eigenvalue equals $\underline{\lambda}_\Sigma = \underline{\sigma}_\Sigma^2 > 0$. Then, by Theorem B.21

$$\operatorname{tr}\left(\tilde{\mathbf{W}}^T \boldsymbol{\Sigma}\boldsymbol{\Sigma}^T \tilde{\mathbf{W}}\right) \geq \underline{\sigma}_\Sigma^2 \left\|\tilde{\mathbf{W}}\right\|_F^2$$

i.e. contrary to q-modification, in this case the trace expression does not vanish, if $\tilde{\mathbf{W}} \neq \mathbf{0}$ and is bounded from below by a quadratic expression. An upper bound on $\dot{V}$ is obtained similarly to the q-modification case.

$$\begin{aligned} \dot{V} \leq & -\underline{\lambda}_{Q_E} \|\mathbf{e}\|_2^2 + 2\overline{\lambda}_{P_E} \|\boldsymbol{\delta}\|_2 \|\mathbf{e}\|_2 \\ & -2\kappa\underline{\sigma}_\Sigma^2 \left\|\tilde{\mathbf{W}}^T\right\|_F^2 + 2\kappa\overline{\sigma}_\Sigma \left\|\tilde{\mathbf{W}}^T\right\|_F \|\boldsymbol{\Delta}\|_F \\ & -2\sigma_x \bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 + \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} \theta_{x,\max}\right) - 2\sigma_N \bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 + \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} \theta_{N,\max}\right) \\ & -2\sigma_L \bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 + \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} \lambda_{L,\max}\right) \end{aligned} \tag{4.208}$$

Remark 3

It is proved in various publications that, in absence of unmatched uncertainties, concurrent learning assures even exponential convergence of the tracking error and parameter-estimation-errors to zero ([Cho10], [Cho101]). Actually, it effects an additional negative definite term in $\dot{V}$, depending on the parameter-estimation-errors, which indicates an accelerated convergence. However, if unmatched uncertainties are present, merely ultimate boundedness of the system states can be concluded. Nevertheless, concurrent learning assures boundedness of the system states in presence of unmatched uncertainties without any further modification and can hence serve as a robustness modification ([Cho102]).

A drawback of this approach is a lack of flexibility in case of a sudden damage. The history stack contains data from some fixed time instances in the past. Although algorithms have been developed, that replace old records by new ones, it takes some time until the whole history stack is renewed. Then, if configuration of the plant changes, e.g. due to damage, the old history stack

does not fit the new configuration and hence, concurrent learning will drive the parameters into the wrong direction. Particularly for unmanned aircraft, fast recover of control performance in failure cases is of special importance.

4.4.3 The MMQ Approach

Both, q-modification and concurrent learning are quite similar and both have advantages as well as disadvantages. Motivation for MMQ modification is the combination of the advantages of both approaches. While concurrent learning provides a full rank matrix $\boldsymbol{\Sigma}$ of recorded regressors which even guarantees exponential stability in absence of unmatched uncertainty, the filtered regressor in q-modification also responds to changes in the plant configuration within the bandwidth of the filter, the regressors are sent through. The basic idea is the usage of multiple filters $G_i(s)$, $i=1,\ldots,N$ according to (4.182) to obtain multiple filtered uncertainty estimation errors according to (4.190).

$$\tilde{\mathbf{c}}_i(t) = \mathbf{H}\left(\tilde{\mathbf{W}}^T(t)\mathbf{q}_i(t) - \mathbf{q}_{\delta,i}(t)\right) \tag{4.209}$$

where

$$\begin{aligned} \mathbf{q}_{x,i}(s) &= G_i(s)[\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})](s) \quad & \mathbf{q}_{N,i}(s) &= G_i(s)[\boldsymbol{\omega}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})](s) \\ \mathbf{q}_{L,i}(s) &= G_i(s)[b(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{u}](s) \quad & \mathbf{q}_{\delta,i}(s) &= G_i(s)[\boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})](s) \end{aligned} \tag{4.210}$$

$$\mathbf{q}_i^T(s) = \begin{pmatrix} \mathbf{q}_{x,i}^T(s) & \mathbf{q}_{N,i}^T(s) & \mathbf{q}_{L,i}^T(s) \end{pmatrix}. \tag{4.211}$$

The gradient-based terms (4.192) of the N filtered expressions are added to (4.193)

$$\begin{aligned} \mathbf{M}_x &= \sigma_x \bar{f}_{\varepsilon_x,\theta_x,\boldsymbol{\Gamma}_x}\left(\hat{\boldsymbol{\Theta}}_x\right)\hat{\boldsymbol{\Theta}}_x + \kappa\boldsymbol{\Gamma}_x\textstyle\sum_{i=1}^{N}\left(\mathbf{q}_{x,i}\tilde{\mathbf{c}}_i^T\right)\mathbf{H} \\ \mathbf{M}_N &= \sigma_N \bar{f}_{\varepsilon_N,\theta_N,\boldsymbol{\Gamma}_N}\left(\hat{\boldsymbol{\Theta}}_N\right)\hat{\boldsymbol{\Theta}}_N + \kappa\boldsymbol{\Gamma}_N\textstyle\sum_{i=1}^{N}\left(\mathbf{q}_{N,i}\tilde{\mathbf{c}}_i^T\right)\mathbf{H} \\ \mathbf{M}_L &= \sigma_L \bar{f}_{\varepsilon_L,\lambda_L,\boldsymbol{\Gamma}_L}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\hat{\boldsymbol{\Lambda}}_L + \kappa\mathbf{B}_L^T\mathbf{H}^T\textstyle\sum_{i=1}^{N}\left(\tilde{\mathbf{c}}_i\mathbf{q}_{L,i}^T\right)\boldsymbol{\Gamma}_L \end{aligned} \tag{4.212}$$

with modification gain $\kappa>0$. If we stack filtered regressor and uncertainty estimation error into matrices

$$\begin{aligned} \mathbf{Q}_x(t) &= \begin{bmatrix}\mathbf{q}_{x,1}(t) & \cdots & \mathbf{q}_{x,N}(t)\end{bmatrix}, \ \mathbf{Q}_N(t) = \begin{bmatrix}\mathbf{q}_{N,1}(t) & \cdots & \mathbf{q}_{N,N}(t)\end{bmatrix}, \\ \mathbf{Q}_L(t) &= \begin{bmatrix}\mathbf{q}_{L,1}(t) & \cdots & \mathbf{q}_{L,N}(t)\end{bmatrix}, \ \mathbf{Q}_\delta(t) = \begin{bmatrix}\mathbf{q}_{\delta,1}(t) & \cdots & \mathbf{q}_{\delta,N}(t)\end{bmatrix} \end{aligned} \tag{4.213}$$

$$\tilde{\mathbf{C}}(t) = \begin{bmatrix}\tilde{\mathbf{c}}_1(t) & \cdots & \tilde{\mathbf{c}}_N(t)\end{bmatrix} \tag{4.214}$$

(4.212) can also be written in compact matrix-vector notation.

$$\begin{aligned} \mathbf{M}_x &= \sigma_x \bar{f}_{\varepsilon_x,\theta_x,\boldsymbol{\Gamma}_x}\left(\hat{\boldsymbol{\Theta}}_x\right)\hat{\boldsymbol{\Theta}}_x + \kappa\boldsymbol{\Gamma}_x\mathbf{Q}_x\tilde{\mathbf{C}}^T\mathbf{H} \\ \mathbf{M}_N &= \sigma_N \bar{f}_{\varepsilon_N,\theta_N,\boldsymbol{\Gamma}_N}\left(\hat{\boldsymbol{\Theta}}_N\right)\hat{\boldsymbol{\Theta}}_N + \kappa\boldsymbol{\Gamma}_N\mathbf{Q}_N\tilde{\mathbf{C}}^T\mathbf{H} \\ \mathbf{M}_L &= \sigma_L \bar{f}_{\varepsilon_L,\lambda_L,\boldsymbol{\Gamma}_L}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\hat{\boldsymbol{\Lambda}}_L + \kappa\mathbf{B}_L^T\mathbf{H}^T\tilde{\mathbf{C}}\mathbf{Q}_L^T\boldsymbol{\Gamma}_L \end{aligned} \tag{4.215}$$

Also

$$\tilde{\mathbf{C}}(t) = \mathbf{H}\left(\tilde{\mathbf{W}}^T(t)\mathbf{Q}(t) - \mathbf{Q}_\delta(t)\right) \tag{4.216}$$

where

$$\mathbf{Q}^T = \begin{bmatrix} \mathbf{Q}_x^T & \mathbf{Q}_N^T & \mathbf{Q}_L^T \end{bmatrix}. \tag{4.217}$$

With these definitions, it is straightforward to see that now (4.194) takes the following form.

$$\begin{aligned} \dot{V} \leq & -\mathbf{e}^T\mathbf{Q}_E\mathbf{e} \quad -2\mathbf{e}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta} - 2\kappa\,\mathrm{tr}\left[\tilde{\mathbf{W}}^T\mathbf{Q}\,\mathbf{Q}^T\tilde{\mathbf{W}}\right] + 2\kappa\,\mathrm{tr}\left[\tilde{\mathbf{W}}^T\mathbf{Q}\mathbf{Q}_\delta^T\right] \\ & -2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 + \mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\boldsymbol{\Theta}_x\right]\right) - 2\sigma_N\bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 + \mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\Gamma}_N^{-1}\boldsymbol{\Theta}_N\right]\right) \\ & -2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 + \mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\boldsymbol{\Lambda}_L^T\right]\right) \end{aligned} \tag{4.218}$$

Comparing this result with the Lyapunov function derivative for concurrent learning (4.205), it reveals that it both are almost equal, except for the difference that the matrix of recorded regressors $\boldsymbol{\Sigma}$ is replaced by the matrix of filtered regressors $\mathbf{Q}$.

In concurrent learning, the history stack is constructed such that $\boldsymbol{\Sigma}$ has full rank, but the history stack is fixed then, as explained above. The MMQ modification overcomes this drawback, however, for the prize that full rank condition cannot be guaranteed. Nevertheless, the more filters are used, the higher the chance that $\mathbf{Q}$ has a large number of linearly independent columns. Hence MMQ modification provides potential to overcome the rank-1 condition of the standard update law and even to have a full rank in $\mathbf{Q}$ in case of sufficient excitation, effecting a “boost” in parameter adaptation. An upper for $\dot{V}$ is obtained analogously to (4.195).

$$\begin{aligned} \dot{V} \leq & -\underline{\lambda}_{Q_E}\|\mathbf{e}\|_2^2 \quad + 2\overline{\lambda}_{P_E}\|\boldsymbol{\delta}\|_2\|\mathbf{e}\|_2 \\ & -2\kappa\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F^2 \quad + 2\kappa\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F\|\mathbf{Q}_\delta\|_F \\ & -2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 + \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\theta_{x,\max}\right) - 2\sigma_N\bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 + \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\theta_{N,\max}\right) \\ & -2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 + \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\lambda_{L,\max}\right) \end{aligned} \tag{4.219}$$

In order to progress, we have to take a closer look at the filtered unmatched uncertainty. If assumption B holds, an upper bound on $\mathbf{Q}_\delta$ can be found, using induced transfer function norms as explained in the following.

Bound on Q_δ – Preliminary Consideration

From assumption B, $||\boldsymbol{\delta}_2|| \leq D$ for all $t \geq t_0$, but for (4.219) we need also a bound on $||\mathbf{Q}_\delta||_F$, whose columns contain the unmatched uncertainty, which is filtered by an LTI system,

described by transfer function (4.182). Transfer function induced norms provide a measure to obtain a relationship between the $\mathcal{L}_\infty$ norms of input and output signal of an LTI system.

In order to establish this relationship, the $\mathcal{L}_\infty$-norm of a time domain signal is defined first. For a scalar signal $u:[0,\infty)\rightarrow\mathbb{R}$ we have

$$\|u(t)\|_{\mathcal{L}_\infty} = \sup_{t\geq 0} |u(t)|. \tag{4.220}$$

I.e. the $\mathcal{L}_\infty$-norm is the least upper bound of all absolute values of a signal, taken at any time instant. A least upper bound for a signal, obtained from the output of an LTI system $y(s)=G(s)u(s)$ is given by the induced transfer function $\mathcal{L}_1$-norm ([Hov10]).

$$\|G(s)\|_{\mathcal{L}_1} = \sup_{\|u\|_{\mathcal{L}_\infty}>0}\left(\frac{\|y(t)\|_{\mathcal{L}_\infty}}{\|u(t)\|_{\mathcal{L}_\infty}}\right)$$

Due to this definition, also the following inequality holds.

$$\|y(t)\|_{\mathcal{L}_\infty} \leq \|G(s)\|_{\mathcal{L}_1}\|u(t)\|_{\mathcal{L}_\infty} \tag{4.221}$$

In order to compute the induced transfer function $\mathcal{L}_1$-norm, let $g(t)$ be the impulse response of $G(s)$, then

$$\|G(s)\|_{\mathcal{L}_1} = \int_0^\infty |g(t)|dt$$

i.e. it equals the area enclosed by the absolute value of the impulse response and the time axis. For a vector time domain signal $\mathbf{u}:[0,\infty)\rightarrow\mathbb{R}^n$, where each component is sent through the same LTI system, i.e. $y_i(s)=G(s)u_i(s)$, or equivalently

$$\mathbf{y}(s)=G(s)\mathbf{u}(s)$$

a bound on the output signal of the form

$$\sup_{t\geq 0}\left(\|\mathbf{y}(t)\|_2\right) \leq K \sup_{t\geq 0}\|\mathbf{u}(t)\|_2 \tag{4.222}$$

can be derived for some constant K, using transfer function induced $\mathcal{L}_1$-norm.

Interim consideration 1

In the subsequent derivations, we use the fact that

$$\sup_{t\geq 0}\left(|u(t)|^2\right)=\left(\sup_{t\geq 0}|u(t)|\right)^2. \tag{4.223}$$

Expressed in words, if D is the least upper bound of $|u(t)|$ then D^2 is the least upper bound of $|u(t)|^2$. For the subsequent derivation, it is important to notice that D is a least

upper bound, if and only if $|u(t)|\leq D$ for all $t\geq 0$ and, additionally, for every $\varepsilon>0$, there is some t_0 such that

$$|u(t_0)| > D-\varepsilon. \tag{4.224}$$

It is easily verified that $|u(t)|^2\leq D^2$ for all $t\geq 0$ and it is left to show that, for every $\bar{\varepsilon}>0$, there is some t_0 such that $|u(t_0)|^2>D^2-\bar{\varepsilon}$. Therefore, we use (4.224) and obtain

$$|u(t_0)|^2 > (D-\varepsilon)^2 = D^2-\varepsilon(2D+\varepsilon).$$

Choosing $\bar{\varepsilon}=\varepsilon(2D+\varepsilon)$ proves the claim.

Interim Consideration 2

For some time domain signals $u_1(t)$, $u_2(t)$, the following identity holds.

$$\sup_{t\geq 0}\left[\max\left(|u_1(t)|,|u_2(t)|\right)\right]=\max\left[\sup_{t\geq 0}|u_1(t)|,\sup_{t\geq 0}|u_2(t)|\right] \tag{4.225}$$

i.e. $\max$ and $\sup$ operators are allowed to be exchanged. Let

$$\sup_{t\geq 0}|u_1(t)|=D_1 \text{ and } \sup_{t\geq 0}|u_2(t)|=D_2,\ D=\max(D_1,D_2)$$

then we have for the right hand side of (4.225):

$$\max\left[\sup_{t\geq 0}|u_1(t)|,\sup_{t\geq 0}|u_2(t)|\right]=D \tag{4.226}$$

Moreover, on the one hand

$$|u_1(t)|\leq D_1 \text{ and } |u_2(t)|\leq D_2 \text{ for all } t\geq 0. \tag{4.227}$$

On the other hand, since D_1 and D_2 are the respective **least** upper bounds for u_1 and u_2, there are some time instances t_1,t_2 for any $\varepsilon>0$ such that

$$|u_1(t_1)|>D_1-\varepsilon \text{ and } |u_2(t_2)|>D_2-\varepsilon. \tag{4.228}$$

Inequalities (4.227) imply

$$\max\left(|u_1(t)|,|u_2(t)|\right)\leq\max(D_1,D_2)=D \text{ for all } t\geq 0$$

and in turn

$$\sup_{t\geq 0}\left[\max\left(|u_1(t)|,|u_2(t)|\right)\right]\leq D. \tag{4.229}$$

whereat inequalities (4.228) imply

$$\max\left(|u_1(t_1)|,|u_2(t_1)|\right)>D_1-\varepsilon \text{ and } \max\left(|u_1(t_2)|,|u_2(t_2)|\right)>D_2-\varepsilon.$$

Since the least upper bound (supremum) is certainly equal or greater than the respective values $|u_1(t_1)|$ and $|u_2(t_2)|$ and since there are such values $D_1-\varepsilon$ and $D_2-\varepsilon$ for any $\varepsilon>0$ we have

$$\sup_{t\geq 0}\left[\max\left(|u_1(t)|,|u_2(t)|\right)\right]\geq D_1 \text{ and } \sup_{t\geq 0}\left[\max\left(|u_1(t)|,|u_2(t)|\right)\right]\geq D_2$$

and consequently

$$\sup_{t\geq 0}\left[\max\left(|u_1(t)|,|u_2(t)|\right)\right]\geq \max(D_1,D_2)=D \tag{4.230}$$

Finally, (4.226), (4.229) and (4.230) establish equation (4.225).

Now we continue to establish result (4.222). By (4.223), we have

$$\left(\sup_{t\geq 0}\|\mathbf{y}(t)\|_2\right)^2=\sup_{t\geq 0}\|\mathbf{y}(t)\|_2^2=\sup_{t\geq 0}\sum_{i=1}^{n}|y_i(t)|^2$$

With help of induced transfer function $\mathcal{L}_1$-norm, we obtain

$$\left(\sup_{t\geq 0}\|\mathbf{y}(t)\|_2\right)^2\leq \sup_{t\geq 0}\sum_{i=1}^{n}\left(\|G(s)\|_{\mathcal{L}_1}\|u_i(t)\|_{\mathcal{L}_\infty}\right)^2$$

and since its argument is a constant, the supremum operator can be dropped and the transfer function norm can be set in front of the sum.

$$\left(\sup_{t\geq 0}\|\mathbf{y}(t)\|_2\right)^2\leq \|G(s)\|_{\mathcal{L}_1}^2\sum_{i=1}^{n}\left(\|u_i(t)\|_{\mathcal{L}_\infty}\right)^2=\|G(s)\|_{\mathcal{L}_1}^2\sum_{i=1}^{n}\left(\sup_{t\geq 0}|u_i(t)|\right)^2$$

Since $\|u_i(t)\|\geq 0$, the upper bound is further enlarged by

$$\left(\sup_{t\geq 0}\|\mathbf{y}(t)\|_2\right)^2\leq \|G(s)\|_{\mathcal{L}_1}^2\left(\sum_{i=1}^{n}\sup_{t\geq 0}|u_i(t)|\right)^2.$$

The sum, in turn, is bounded from above by

$$\sum_{i=1}^{n}\sup_{t\geq 0}|u_i(t)|\leq n\max_{i=1,\ldots,n}\left(\sup_{t\geq 0}|u_i(t)|\right)=n\sup_{t\geq 0}\left(\max_{i=1,\ldots,n}|u_i(t)|\right)\leq n\sup_{t\geq 0}\sqrt{\sum_{i=1}^{n}|u_i(t)|^2}=n\sup_{t\geq 0}\|\mathbf{u}(t)\|_2$$

where we exchanged $\max$ and $\sup$ operators according to equation (4.225). Using this bound, we obtain

$$\left(\sup_{t\geq 0}\|\mathbf{y}(t)\|_2\right)^2\leq \|G(s)\|_{\mathcal{L}_1}^2\left(\sum_{i=1}^{n}\sup_{t\geq 0}|u_i(t)|\right)^2\leq \|G(s)\|_{\mathcal{L}_1}^2 n^2\left(\sup_{t\geq 0}\|\mathbf{u}(t)\|_2\right)^2$$

Finally, taking square roots on both sides, we obtain (4.222).

$$\sup_{t\geq 0}\|\mathbf{y}(t)\|_2\leq n\|G(s)\|_{\mathcal{L}_1}\sup_{t\geq 0}\|\mathbf{u}(t)\|_2 \tag{4.231}$$

Bound on Q_δ

The columns of $\mathbf{Q}_\delta$ equal the unmatched uncertainty, sent through different LTI filters.

$$\mathbf{q}_{\delta,i}(s) = G_i(s)\boldsymbol{\delta}(s)$$

In order to find a bound on $\mathbf{Q}_\delta$, let us assume that the induced transfer function $\mathcal{L}_1$-norm of all filters is bounded by some $L>0$.

Assumption C*: Boundedness of $\mathcal{L}_1$-norm of MMQ Filters*

$$\left\|G_i(s)\right\|_{\mathcal{L}_1} \le L \text{ for } i = 1,\dots,N \tag{4.232}$$

Further, it is known, that the Frobenius norm can be expressed in terms of trace operator, which is equivalent to the sum of scalar products of its columns. Using assumptions A, B and result (4.231) we obtain

$$\left\|\mathbf{Q}_\delta\right\|_F = \sqrt{\mathrm{tr}\left(\mathbf{Q}_\delta^T\mathbf{Q}_\delta\right)} = \sqrt{\sum_{i=1}^N \mathbf{q}_{\delta,i}^T\mathbf{q}_{\delta,i}} \le \sqrt{\sum_{i=1}^N \left\|\mathbf{q}_{\delta,i}\right\|_2^2} \le \sqrt{\sum_{i=1}^N m^2\left\|G_i(s)\right\|_{\mathcal{L}_1}^2\left\|\boldsymbol{\delta}\right\|_2^2}$$
$$\left\|\mathbf{Q}_\delta\right\|_F \le mLD\sqrt{N}\,. \tag{4.233}$$

Lyapunov Stability Analysis

Using assumption B and inequality (4.233) an upper bound on (4.218) is given by:

$$\begin{aligned}\dot V \le -\underline{\lambda}_{Q_E}\left\|\mathbf{e}\right\|_2^2 \quad & +2\overline{\lambda}_{P_E}D\left\|\mathbf{e}\right\|_2 - 2\kappa\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F^2 + 2\kappa mLD\sqrt{N}\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F \\ & -2\sigma_x\bar f\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 - \theta_{x,\max}\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2\right) - 2\sigma_N\bar f\left(\hat{\boldsymbol{\Theta}}_N\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 - \theta_{N,\max}\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\right) \\ & -2\sigma_L\bar f\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 - \lambda_{L,\max}\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\right)\end{aligned} \tag{4.234}$$

As in section 4.3, the upper bound is divided into the partial functions (4.147), where an additional partial function appears due to MMQ modification

$$h_q\left(\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F\right) = -2\kappa\left(\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F^2 - mLD\sqrt{N}\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F\right) \tag{4.235}$$

which is a concave parabola and hence adopts some maximum.

$$h_{q,\max} = \kappa\frac{(mLD)^2 N}{2} \tag{4.236}$$

Now, four functions are constructed, each of which depends on only one part of the whole state vector and whose negative values are an upper bound on $\dot V$. It is obtained, by taking the respective partial function of (4.147) and using the respective worst-case bounds (4.148) and (4.236) of the other parts.

$$\begin{aligned}
\gamma_e\left(\|\mathbf{e}\|_2\right) &= \underline{\lambda}_{Q_E}\|\mathbf{e}\|_2^2 - 2\overline{\lambda}_{P_E} D\|\mathbf{e}\|_2 - \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_N\theta_{N,\max}^2 + \sigma_L\lambda_{L,\max}^2 + \kappa(mLD)^2 N\right)\\
\gamma_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\right) &= 2\sigma_x \bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left[\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 - \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\theta_{x,\max}\right] - \frac{\left(\overline{\lambda}_{P_E} D\right)^2}{\underline{\lambda}_{Q_E}} - \frac{1}{2}\left(\sigma_N\theta_{N,\max}^2 + \sigma_L\lambda_{L,\max}^2 + \kappa(mLD)^2 N\right)\\
\gamma_N\left(\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\right) &= 2\sigma_N \bar{f}\left(\hat{\boldsymbol{\Theta}}_N\right)\left[\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 - \left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}\theta_{N,\max}\right] - \frac{\left(\overline{\lambda}_{P_E} D\right)^2}{\underline{\lambda}_{Q_E}} - \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_L\lambda_{L,\max}^2 + \kappa(mLD)^2 N\right)\\
\gamma_L\left(\left\|\tilde{\boldsymbol{\Lambda}}_L\right\|_{\boldsymbol{\Gamma}_L}\right) &= 2\sigma_L \bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left[\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 - \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\lambda_{L,\max}\right] - \frac{\left(\overline{\lambda}_{P_E} D\right)^2}{\underline{\lambda}_{Q_E}} - \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_N\theta_{N,\max}^2 + \kappa(mLD)^2 N\right)
\end{aligned} \tag{4.237}$$

From now on, the stability analysis follows similar lines as in section 4.3, with the single difference that there is an additional term $\kappa(mLD)^2$ due to MMQ modification. Equation (4.150) holds with γ_i defined in (4.237). Further, r_e, r_x, r_N, r_L, defined in (4.151), (4.152), and class $\mathcal{K}$ functions α_i, β_i, defined in (4.153), providing a lower and upper bound on the Lyapunov function candidate remain unchanged and with it $\underline{u}$, ρ_e, ρ_x, ρ_N and ρ_L, defined in (4.154), (4.155). If condition (4.156) holds, (4.237) are convex parabolas

$$\begin{aligned}
\gamma_e &= \underline{\lambda}_{Q_E}\|\mathbf{e}\|_2^2 - 2\overline{\lambda}_{P_E} D\|\mathbf{e}\|_2 - V_e - V_Q & \gamma_x &= 2\sigma_x\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 - 2\sigma_x\theta_x\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} - V_x - V_D - V_Q\\
\gamma_N &= 2\sigma_N\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N}^2 - 2\sigma_N\theta_N\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} - V_N - V_D - V_Q & \gamma_L &= 2\sigma_L\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 - 2\sigma_L\lambda_L\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} - V_L - V_D - V_Q
\end{aligned} \tag{4.238}$$

with V_e, V_x, V_N, V_L, V_D, according to (4.158). Comparing to (4.157), an additional positive constant

$$V_Q = \frac{1}{2}\kappa(mLD)^2 N \tag{4.239}$$

appears in (4.238) due to MMQ modification. Continuing along arguments of section 4.3, (4.238) are rendered positive, if $\|\mathbf{e}\|_2 \geq \mu_e$, $\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x} \geq \mu_x$, $\left\|\tilde{\boldsymbol{\Theta}}_N\right\|_{\boldsymbol{\Gamma}_N} \geq \mu_N$, $\left\|\tilde{\boldsymbol{\Lambda}}_L\right\|_{\boldsymbol{\Gamma}_L} \geq \mu_L$, where

$$\begin{aligned}
\mu_e &= \frac{\overline{\lambda}_{P_E} D}{\underline{\lambda}_{Q_E}} + \sqrt{\left(\frac{\overline{\lambda}_{P_E} D}{\underline{\lambda}_{Q_E}}\right)^2 + \frac{V_e + V_Q + K_0}{\underline{\lambda}_{Q_E}}} & \mu_x &= \frac{\theta_{x,\max}}{2} + \sqrt{\left(\frac{\theta_{x,\max}}{2}\right)^2 + \frac{V_D + V_x + V_Q + K_0}{2\sigma_x}}\\
\mu_N &= \frac{\theta_{N,\max}}{2} + \sqrt{\left(\frac{\theta_{N,\max}}{2}\right)^2 + \frac{V_D + V_N + V_Q + K_0}{2\sigma_N}} & \mu_L &= \frac{\lambda_{L,\max}}{2} + \sqrt{\left(\frac{\lambda_{L,\max}}{2}\right)^2 + \frac{V_D + V_L + V_Q + K_0}{2\sigma_L}}
\end{aligned} \tag{4.240}$$

for some $K_0>0$ and γ_i are class $\mathcal{K}$ functions on $\overline{\mathcal{B}}_r \backslash \mathcal{B}_\mu$ (according to Corollary 3.2). Moreover, for ultimate boundedness, we need $\mu_e<\rho_e$ which yields the following inequality.

$$\frac{\overline{\lambda}_{P_E} D}{\underline{\lambda}_{Q_E}} + \sqrt{\left(\frac{\overline{\lambda}_{P_E} D}{\underline{\lambda}_{Q_E}}\right)^2 + \frac{V_e + V_Q + K_0}{\underline{\lambda}_{Q_E}}} < \sqrt{\frac{\underline{\lambda}_{P_E}}{\overline{\lambda}_{P_E}}}\frac{r_e}{2}$$

After taking squares, we obtain.

$$V_e + V_Q + K_0 < k_e(r_e) \tag{4.241}$$

where $k_e(r_e)$ remains unchanged w.r.t. the case without MMQ modification and is defined in (4.161). Since σ_x, σ_N, σ_L, κ, $K_0>0$ and with it $V_e+V_Q+K_0>0$ can be chosen arbitrarily small, a necessary and sufficient condition for the existence of a solution to (4.241) is $k_e(r_e)>0$which imposes a minimum value on r_e, given by (4.162). Analogously, the remaining conditions on ultimate boundedness, $\mu_x<\rho_x$, $\mu_N<\rho_N$, $\mu_L<\rho_L$, yield the following inequalities

$$\frac{V_D+V_x+V_Q+K_0}{\sigma_x} < k_x(r_e,\gamma_x)\,,\ \frac{V_D+V_N+V_Q+K_0}{\sigma_N} < k_N(r_e,\gamma_N)\,,\ \frac{V_D+V_L+V_Q+K_0}{\sigma_L} < k_L(r_e,\gamma_L) \tag{4.242}$$

where $k_x(r_e,\gamma_x)$, $k_N(r_e,\gamma_N)$, $k_L(r_e,\gamma_L)$ also remain unchanged w.r.t. the case without MMQ modification and are defined in (4.164), (4.167), (4.168). Each left hand side of inequalities (4.242) is positive but arbitrarily small for sufficiently large σ_x, σ_N, σ_L. Hence necessary and sufficient condition for the existence of a solution σ_x, σ_N, σ_L, κ, $K_0>0$ to a single inequality of (4.242) is given by $k_x(r_e,\gamma_x)>0$, $k_N(r_e,\gamma_N)>0$, $k_L(r_e,\gamma_L)>0$ which impose minimum values for the learning rates, given by (4.165), (4.169), (4.170).

Summing up, all conditions on r_e, γ_x, γ_N and γ_L together, namely (4.162), (4.165), (4.169), (4.170), are a necessary condition for the existence of a solution to the following system of inequalities

$$\begin{array}{rrrrrcl}
\theta_{x,\max}^2\cdot\sigma_x & +\theta_{N,\max}^2\cdot\sigma_N & +\lambda_{L,\max}^2\cdot\sigma_L & +2K_0 & +k_Q\kappa & < & 2k_e(r_e) \\
-2k_x(r_e,\gamma_x)\cdot\sigma_x & +\theta_{N,\max}^2\cdot\sigma_N & +\lambda_{L,\max}^2\cdot\sigma_L & +2K_0 & +k_Q\kappa & < & -2V_D \\
\theta_{x,\max}^2\cdot\sigma_x & -2k_N(r_e,\gamma_N)\cdot\sigma_N & +\lambda_{L,\max}^2\cdot\sigma_L & +2K_0 & +k_Q\kappa & < & -2V_D \\
\theta_{x,\max}^2\cdot\sigma_x & +\theta_{N,\max}^2\cdot\sigma_N & -2k_L(r_e,\gamma_L)\cdot\sigma_L & +2K_0 & +k_Q\kappa & < & -2V_D
\end{array} \tag{4.243}$$

which is obtained by inserting definitions (4.158), (4.239) into inequalities (4.241), (4.242). Thereby, we have defined

$$k_Q = (mLD)^2 N\,. \tag{4.244}$$

For sufficiency, however, we potentially need that $k_e(r_e)$, $k_i(r_e,\gamma_i)$, $i=x,N,L$, are positive and bounded away from zero. Therefore, assume that inequalities (4.243) have a solution for some $k_e^*,k_i^*>0$. Then, they also have a solution for $k_e\geq k_e^*,k_i\geq k_i^*$, which can be easily verified by similar arguments as used in section 4.3.

Moreover, for given $k_e^*,k_i^*>0$, we obtain the lower bounds on r_e, γ_x, γ_N and γ_L as in section 4.3 , given by equations (4.174) – (4.177). If r simultaneously satisfies (4.151), such that (4.243) has a solution, then the system states are uniformly ultimately bounded according to Corollary 3.2.

Corollary 4.2 MRAC Variant 1, MMQ Modification – Ultimate Boundedness

Consider:

- stable **reference dynamics** (4.56)
- **plant dynamics** in Byrnes-Isidori normal form (4.90), subjected to linearizing state feedback (4.125) with pseudo control (4.121), adaptive term (4.122), error feedback (4.54) and reference pseudo control (4.55)
- **parameterizations** (4.107) – (4.110) for state dependent uncertainty, (4.111) or (4.114) for affine control effectiveness together with (4.184) and (4.115) – (4.117) for nonaffine control effectiveness
- **parameter-estimation-errors** $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Theta}}_N$, $\tilde{\boldsymbol{\Lambda}}_L$, defined in (4.124), (4.119), (4.112)
- **update law** (4.131) using
  - symmetric positive definite solution $\mathbf{P}_E$ of Lyapunov equation (4.128) for some symmetric positive definite $\mathbf{Q}_E$
  - positive learning rates γ_x, γ_N, γ_L
  - projection operator with parameters ε_x, ε_N, ε_L,>0, $\theta_{x,max}$, $\theta_{N,max}$, $\lambda_{L,max}>0$ and symmetric positive definite weighting matrices $\boldsymbol{\Gamma}_x$, $\boldsymbol{\Gamma}_N$, $\boldsymbol{\Gamma}_L$
  - switching σ-modification with modification gains σ_x, σ_N, σ_L >0 and MMQ modification with modification gain $\kappa>0$ and N filters according to (4.215)
- Sets: $\overline{\mathcal{B}_{\bar{\zeta}}} \subset \mathbb{R}^r$, $\mathcal{D}_\eta \subset \mathbb{R}^{n-r}$, $\mathcal{U} \subset \mathbb{R}^{p_N}$ $\mathcal{D} \subset \mathbb{R}^d$ for some $\bar{\zeta} > 0$

The **tracking error** (4.126) and $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Theta}}_N$, $\tilde{\boldsymbol{\Lambda}}_L$ are **uniformly ultimately bounded**, if

- assumption A is fulfilled for some θ_x, θ_N, λ_L >0 and $\theta_{x,max}$, $\theta_{N,max}$, $\lambda_{L,max}$ are defined by (4.133) – (4.135)
- assumption B is fulfilled for some $D \geq 0$
- $(\boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) \in \mathcal{D}_\eta \times \mathcal{U} \times \mathcal{D}$ for all $t \geq t_0$
- assumption C on the $\mathcal{L}_1$ bound of the MMQ filters is fulfilled for some $L>0$
- the reference model state is bounded according to (4.139) for some $\bar{\zeta}_R > 0$
- r_e simultaneously satisfies $r_e > \underline{e}_e$ according to (4.178) and $r_e \leq \bar{\zeta} - \bar{\zeta}_R$ according to (4.151) where $\underline{e}_e$ is defined in (4.174) with k_e^*, k_x^*, k_N^*, k_L^* such that inequalities (4.173) are solved for σ_x, σ_N, σ_L,κ, $K_0>0$
- the learning rates are chosen such that $\gamma_x \geq \underline{\gamma}_x, \gamma_N \geq \underline{\gamma}_N, \gamma_L \geq \underline{\gamma}_L$ according to (4.178) where $\underline{\gamma}_x, \underline{\gamma}_N, \underline{\gamma}_L$ are defined in (4.175) - (4.177).
- initial conditions: $\|\mathbf{e}(t_0)\|_2 \leq \delta_e$, $\|\tilde{\boldsymbol{\Theta}}_x(t_0)\|_{\boldsymbol{\Gamma}_x} \leq \delta_x$, $\|\tilde{\boldsymbol{\Theta}}_N(t_0)\|_{\boldsymbol{\Gamma}_N} \leq \delta_N$, $\|\tilde{\boldsymbol{\Lambda}}_L(t_0)\|_{\boldsymbol{\Gamma}_L} \leq \delta_L$
  - $\delta_e \in [0, \rho_e]$, $\delta_x \in [0, \rho_x]$, $\delta_N \in [0, \rho_N]$, $\delta_L \in [0, \rho_L]$
  - ρ_e, ρ_x, ρ_N, ρ_L defined in (4.155)

such that

- $\|\mathbf{e}(t)\|_2 \leq r_e$ forall $t \geq t_0$ and $\|\mathbf{e}(t)\|_2 \leq b_e$ forall $t \geq t_0 + T_e(b_e, \delta_e)$
- $\|\tilde{\boldsymbol{\Theta}}_x(t)\|_{\boldsymbol{\Gamma}_x} \leq \sqrt{\gamma_x \underline{\lambda}_{P_E}} r_e$, $\|\tilde{\boldsymbol{\Theta}}_N(t)\|_{\boldsymbol{\Gamma}_N} \leq \sqrt{\gamma_N \underline{\lambda}_{P_E}} r_e$, $\|\tilde{\boldsymbol{\Lambda}}_L^T(t)\|_{\boldsymbol{\Gamma}_L} \leq \sqrt{\gamma_L \underline{\lambda}_{P_E}} r_e$ forall $t \geq t_0$
- $\|\tilde{\boldsymbol{\Theta}}_x(t)\|_{\boldsymbol{\Gamma}_x} \leq b_x$ forall $t \geq t_0 + T_x(b_x, \delta_x)$, $\|\tilde{\boldsymbol{\Theta}}_N(t)\|_{\boldsymbol{\Gamma}_N} \leq b_N$ forall $t \geq t_0 + T_N(b_N, \delta_N)$
  $\|\tilde{\boldsymbol{\Lambda}}_L^T(t)\|_{\boldsymbol{\Gamma}_L} \leq b_L$ forall $t \geq t_0 + T_L(b_L, \delta_L)$

where

- $b_e = \sqrt{\underline{\lambda}_{P_E}^{-1} \nu}$, $b_x = \sqrt{\gamma_x \nu}$, $b_N = \sqrt{\gamma_N \nu}$, $b_L = \sqrt{\gamma_L \nu}$, $\nu = \bar{\lambda}_{P_E} \mu_e^2 + \gamma_x^{-1} \mu_x^2 + \gamma_N^{-1} \mu_N^2 + \gamma_L^{-1} \mu_L^2$
  μ_e, μ_x, μ_N, μ_L defined in (4.158), (4.239), (4.240)
- $T_e(b_e, \delta_e) = \max\left[0, K_0^{-1}\left(\bar{\lambda}_{P_E} \delta_e^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_e^2 - \nu\right)\right]$, $T_x(b_x, \delta_x) = \max\left[0, K_0^{-1}\left(\delta_x^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_e^2 - \nu\right)\right]$
  $T_N(b_N, \delta_N) = \max\left[0, K_0^{-1}\left(\delta_N^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_e^2 - \nu\right)\right]$, $T_L(b_L, \delta_L) = \max\left[0, K_0^{-1}\left(\delta_L^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_e^2 - \nu\right)\right]$

Remark 4

As explained in remarks 1 and 3 of this section, MMQ modification introduces an additional negative semi definite term $-2\kappa(/|\tilde{\mathbf{W}}^T\mathbf{Q}/|_F)^2$ to the upper bound on $\dot{V}$ – equation (4.234) – if unmatched uncertainties are absent, leading to a faster convergence of the parameter-estimation-errors. In case of sufficient excitation, such that the matrix of filtered regressors $\mathbf{Q}$ has full rank, even exponential convergence of $\tilde{\mathbf{W}}$ can be shown by analogous arguments as used for concurrent learning ([Cho102]). However, in presence of unmatched uncertainty, the terms that occur due to MMQ modification in the upper bound on $\dot{V}$ are

$$-2\kappa\left(\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F^2 + mLD\sqrt{N}\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F\right)$$

i.e. there is a negative quadratic term and a positive linear term. For small values of $||\cdot||_F$, the linear term is dominating, which increases the upper bound on $\dot{V}$. Hence, it is desirable to minimize the linear term. Usually m, the number of components of the unmatched uncertainty vector is fixed given by the system and cannot be varied. Also, the uncertainty bound D is a fixed value. Evidently, the dominance of the linear term can be influenced by the control system designer by a proper choice of filters, comprising a small $\mathcal{L}_1$-norm bound L and the number of employed filters N. Fortunately, N merely enters the linear term as square root.

4.5 Nonlinear-in-Control Design

The control system, derived in section 4.3 does not utilize the nonaffine controls for feedback linearization of the system, but their values are specified separately and their influence on the dynamics is cancelled by feedback linearization using the affine controls (second part of Figure 4.7). However, if the inverse of the nonlinear map $\mathbf{g}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$, that describes the effect of the nonaffine controls onto the dynamics, is not known (refer to equation (4.91)), the algorithm introduced in this section provides a method to obtain the desired value for $\mathbf{u}_N$ in an online gradient minimization approach.

In recent publications, this concept has been as introduced as “nonlinear-in-control” (NIC) design, using singular perturbation theory ([Lav07b], [Lav07a], [Hov06]). The algorithms, presented there have to be tailored for our purpose, however, initially the basic results of these publications are stated.

4.5.1 Singular Perturbation and Gradient Systems

NIC design uses the so-called singular perturbation (SP) theory. Although a detailed description can be found in [Kha02], the main ideas are briefly introduced. SP considers systems of the form

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(t,\mathbf{x},\mathbf{u},\varepsilon)\\ \varepsilon\dot{\mathbf{u}} &= \mathbf{g}(t,\mathbf{x},\mathbf{u},\varepsilon)\end{aligned} \tag{4.245}$$

where the first equation is referred to the *slow dynamics* while the second represents the *fast dynamics*. Such a division of a system into slow and fast dynamics often occurs in real physical systems, e.g. referring to aircraft, the slow dynamics could be associated with the rigid body states, while the actuator state are the fast dynamics. ε thereby is a small positive parameter, that determines the speed of the fast dynamics compared to the slow dynamics and for $\varepsilon=0$ the fast dynamics degenerate to an algebraic equation which corresponds to an infinite fast dynamics such that the equilibrium state $\mathbf{u}$ – the solution of the algebraic equation – is reached instantaneously.

In order for an equilibrium to exist, it has to be assumed that $\mathbf{g}(t,\mathbf{x},\mathbf{u},0)$ has an isolated root for each fixed t and $\mathbf{x}$, such that $\mathbf{u}=\mathbf{h}(t,\mathbf{x})$ implies that $\mathbf{g}(t,\mathbf{x},\ \mathbf{h}(t,\mathbf{x}),0)=0$. If we insert the isolated root into the slow dynamics, and set $\varepsilon=0$, we obtain the so-called *reduced dynamics*

$$\dot{\mathbf{x}}=\mathbf{f}(t,\mathbf{x},\mathbf{h}(t,\mathbf{x}),0) \tag{4.246}$$

which, referring to aircraft, is the dynamics of the rigid body states in case of infinite fast actuators. For stability analysis within the SP framework, the fast dynamics are transformed into the fast timescale τ which is related to the slow timescale by $t=\varepsilon\tau$. Applying the chain rule

$$\dot{\mathbf{u}}=\frac{d\mathbf{u}}{dt}=\frac{1}{\varepsilon}\cdot\frac{d\mathbf{u}}{d\tau}$$

and setting $\varepsilon=0$ yields the *boundary layer dynamics*

$$\frac{d\mathbf{u}}{d\tau}=\mathbf{g}(t,\mathbf{x},\mathbf{u},0) \tag{4.247}$$

which could be interpreted as actuator dynamics in the fast timescale, such fast that the rigid body states $\mathbf{x}$ as well as the normal time scale t are frozen.

The central theorem in singular perturbation theory “Thikonov’s Theorem” states that – under various continuity conditions on the nonlinear maps, involved into the dynamics, are well as its partial derivatives – if the boundary layer dynamics as well as the reduced dynamics are exponentially stable, then there is some $\varepsilon^*>0$ such that the states of the full system (4.245) are bounded and in a vicinity of the states of reduced and boundary layer system, for all $\varepsilon<\varepsilon^*$. Unfortunately the theorem does not provide an explicit value ε^*, but just states that it exists. This is due to the general nonlinear structure of $\mathbf{f}$ and $\mathbf{g}$. As will be shown in section 4.6, if we use the NIC design within the MRAC controller, the dynamics adopt a special form

$$\dot{\mathbf{e}} = \mathbf{A}\mathbf{e} - \mathbf{H}\left[\mathbf{g}(t,\mathbf{e},\mathbf{u}) - \mathbf{w}^*(t,\mathbf{e})\right] \tag{4.248}$$

$$\varepsilon\dot{\mathbf{u}} = -\mathbf{J}_g^T(t,\mathbf{e},\mathbf{u})\left[\mathbf{g}(t,\mathbf{e},\mathbf{u}) - \mathbf{w}^*(t,\mathbf{e})\right] \tag{4.249}$$

where $\mathbf{e}\in\mathbb{R}^r$ is the tracking error, $\mathbf{e}\in\mathbb{R}^r$, $\mathbf{A}\in\mathbb{R}^{r\times r}$ is the system matrix with eigenvalues in the open l.h.c.p., $\mathbf{H}$ is defined in (4.16), $\mathbf{w}=\mathbf{g}(t,\mathbf{e},\mathbf{u})\in\mathbb{R}^m$ is the nonlinear map that describes the effect of the nonaffine controls $\mathbf{u}\in\mathbb{R}^{p_N}$ onto the rigid body dynamics, $\mathbf{J}_g$ is the Jacobian of $\mathbf{g}$ w.r.t. $\mathbf{u}$, $\mathbf{w}^*$ is the desired virtual control, originating either from the linearizing state feedback (equation (4.93)) or is prescribed from extern. The desired virtual control is produced by some ideal $\mathbf{u}^*$ such that

$$\mathbf{g}\left(t,\mathbf{e},\mathbf{u}^*\right) = \mathbf{w}^*. \tag{4.250}$$

Notice that $\mathbf{u}^*$ is different from $\mathbf{u}$ which is the state of differential equation (4.249), but it is desired that (4.249) approximately generates $\mathbf{u}=\mathbf{u}^*$. $\mathbf{g}$ is assumed to be known and the inverse $\mathbf{u}=\mathbf{g}^{-1}(t,\mathbf{e},\mathbf{w})$ exists but is unknown. Existence of a (unique) inverse in turn requires that the number of controls equals the number of nonlinear maps, i.e. $p_N=m$. Additionally, the desired virtual control $\mathbf{w}^*$ has to be achievable by some $\mathbf{u}^*$.

Assumption: *Existence of an Inverse to the Nonlinear Control Map*

- The number of controls and nonlinear maps are equal: $\mathbf{g}(t,\mathbf{e},\mathbf{u})\in\mathbb{R}^m$, $\mathbf{u}\in\mathbb{R}^m$
- the inverse $\mathbf{u}=\mathbf{g}^{-1}(t,\mathbf{e},\mathbf{w})$ exists (uniquely) but is unknown
- there is always a $\mathbf{u}^*$ such that $\mathbf{g}\left(t,\mathbf{e},\mathbf{u}^*\right)=\mathbf{w}^*(t,\mathbf{e})$

Remark

As stated before, Thikonov's theorem does not provide an explicit value ε^* that guarantees stability of the NIC algorithm (4.248), (4.249). Particularly in aircraft applications, an explicit value is very important, since certification authorities require guaranties that the implemented controller operates the aircraft safely. If an explicit value is not known, a controller has to be implemented based on a guess for ε^*, for which reason there is no guaranty on stability of the system. In the subsequent derivations in section 4.6, we will utilize the special structure (4.248), (4.249) to obtain a specific value for ε^*.

In order to approximate $\mathbf{u}\approx\mathbf{u}^*$ such that $\mathbf{g}(t,\mathbf{e},\mathbf{u})\approx\mathbf{w}^*$, the fast dynamics should "tune" $\mathbf{u}$ and they should be such fast, that they could be neglected w.r.t. to the slow dynamics. An online tuning system of the form (4.249) is also referred to a *gradient system*. It inherits the role of the fast dynamics within the SP framework.

As Thikonov's theorem requires that the boundary layer system is exponentially stable, the next section will investigate, under which conditions on $\mathbf{g}$, the boundary layer system is exponentially stable.

4.5.2 Exponential Stability for Autonomous Gradient Systems

For stability analysis, the fast dynamics (4.249) are appropriately transformed such that the equilibrium is shifted to the origin. Defining the error state $\tilde{\mathbf{u}} = \mathbf{u} - \mathbf{u}^*$, the fast dynamics read as

$$\varepsilon \dot{\tilde{\mathbf{u}}} = -\mathbf{J}_g^T(t,\mathbf{e})\left[\mathbf{g}\left(t,\mathbf{e},\mathbf{u}^* + \tilde{\mathbf{u}}\right) - \mathbf{w}^*(t,\mathbf{e})\right] - \varepsilon \frac{d\mathbf{u}^*(t,\mathbf{e})}{dt} \tag{4.251}$$

and consequently, we obtain the boundary layer dynamics by switching to the fast timescale and setting $\varepsilon = 0$.

$$\frac{d\tilde{\mathbf{u}}}{d\tau} = -\mathbf{J}_g^T(t,\mathbf{e})\left[\mathbf{g}\left(t,\mathbf{e},\tilde{\mathbf{u}} + \mathbf{u}^*\right) - \mathbf{w}^*(t,\mathbf{e})\right] \tag{4.252}$$

where t and $\mathbf{e}$ are now considered as frozen parameters while the dynamics of $\mathbf{u}$ are described in context of the fast timescale τ. The boundary layer dynamics could thus be considered as autonomous system, which depends on the parameters t and $\mathbf{e}$. Before continuing with the derivations, we will introduce a reduced notation for readability:

- $\mathbf{x} := \tilde{\mathbf{u}}$
- $\mathbf{f}(\mathbf{x}) := \mathbf{g}\left(t,\mathbf{e},\mathbf{x}+\mathbf{u}^*\right) - \mathbf{w}^*(t,\mathbf{e})$
- $\mathbf{J}_f(\mathbf{x}) := \dfrac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{J}_g\left(t,\ \mathbf{e},\ \mathbf{x}+\mathbf{u}^*\right)$

It is tacitly assumed that $\mathbf{f}(\mathbf{x})$ depends on parameters t and $\mathbf{e}$ and the subsequent conditions hold uniformly in the parameters. Note that $\mathbf{f}(\mathbf{0})=\mathbf{0}$, which is a consequence of (4.250). The dynamics of the boundary layer system can be derived from the gradient of a cost function as explained in the following. Therefore, we define the quadratic cost function

$$V(\mathbf{x}) = \frac{1}{2}\mathbf{f}^T(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x})$$

whose gradient is

$$\nabla V(\mathbf{x}) = \left(\frac{\partial V(\mathbf{x})}{\partial \mathbf{x}}\right)^T = \left(\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}\right)^T \cdot \mathbf{f}(\mathbf{x}) = \mathbf{J}_f^T(\mathbf{x})\mathbf{f}(\mathbf{x})$$

If V should be minimized in an online model, it is nearby to set the right hand side of the differential equation equal to the negative gradient of V.

$$\dot{\mathbf{x}} = -\mathbf{J}_f^T(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) \tag{4.253}$$

The question that arises is, under which conditions on $\mathbf{f}$ the equilibrium of such gradient systems is (asymptotically or even exponentially) stable, such that V is actually minimized by (4.253). Theorem 4.5 in the following provides sufficient conditions for

exponential stability of the gradient system. It additionally assures that the nonlinear mapping is a C^1 diffeomorphism, i.e. it is continuously differentiable and the inverse of $\mathbf{f}(\mathbf{x})$ exists uniquely and is continuously differentiable, too. This fact is necessary for NIC design, where it is preliminary that the inverse exists, even if there is no analytical expression available. Notice that the theorem explicitly provides the exponential rate of convergence as well as a region of admissible initial values.

Theorem 4.5 Exponential Stability of Gradient Systems

For an open convex set $\mathcal{X} \subset \mathbb{R}^m$, containing the origin such that the closed ball $\overline{\mathcal{B}}_{r_0} \subset \mathcal{X}$ for some $r_0 > 0$, let a continuously differentiable map $\mathbf{f}(\mathcal{X}) = \mathcal{Y} \subset \mathbb{R}^m$ with $\mathbf{f}(\mathbf{0}) = \mathbf{0}$. For $\mathbf{x}, \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$, let

$$\mathbf{J}_f(\mathbf{x}) = \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}},\ \mathbf{J}_{f,s}(\mathbf{x}) = \frac{1}{2}\left[\mathbf{J}_f(\mathbf{x}) + \mathbf{J}_f^T(\mathbf{x})\right],\ \mathbf{P}(\mathbf{x}_1, \mathbf{x}_2) = \int_0^1 \mathbf{J}_f\left[\mathbf{x}_1 + s(\mathbf{x}_2 - \mathbf{x}_1)\right] ds,$$

$\underline{\sigma}_J$: minimum singular value of $\mathbf{J}_f(\mathbf{x})$, $\lambda_i\left[\mathbf{J}_{f,s}(\mathbf{x})\right]$: i^{th} eigenvalue of $\mathbf{J}_{f,s}(\mathbf{x})$

$\underline{\sigma}_P, \overline{\sigma}_P$: minimum/ maximum singular value of $\mathbf{P}(\mathbf{x}_1, \mathbf{x}_2)$.

Then, $\mathbf{f}$ is a C^1 diffeomorphism on $\mathcal{Y}$, if

1. for some $k_0 > 0$, $\underline{\sigma}_J \geq k_0$ and
2. for some $k_1 > 0$, either
   a. $\underline{\sigma}_P \geq k_1$ or
   b. $\mathbf{J}_{f,s}(\mathbf{x})$ positive or negative definite such that either

   $\lambda_i\left[\mathbf{J}_{f,s}(\mathbf{x})\right] \geq k_1$ or $\lambda_i\left[\mathbf{J}_{f,s}(\mathbf{x})\right] \leq -k_1$ respectively for $i = 1, \ldots, m$. If additionally,
3. for some $k_2 \geq k_1$: $\overline{\sigma}_P \leq k_2$,

then the origin of the gradient system

$$\dot{\mathbf{x}}(t) = -\mathbf{J}_f^T(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}), \qquad \mathbf{x}(t) = \mathbf{x}_0$$

is the *unique exponentially stable equilibrium* such that

$$\|\mathbf{x}(t)\|_2 \leq \|\mathbf{x}(t_0)\|_2 \cdot \frac{k_2}{k_1} \cdot \exp\left[-\frac{1}{2}\left(\frac{k_0 k_1}{k_2}\right)^2 (t - t_0)\right] \quad \text{if} \quad \|\mathbf{x}(t_0)\|_2 < \frac{k_1}{k_2} r_0$$

Proof

The first condition implies that the Jacobian is nonsingular, i.e.

$$\det\left[\mathbf{J}_f(\mathbf{x})\right] \neq 0 \tag{4.254}$$

for $\mathbf{x} \in \mathcal{X}$. This fact together with condition 2 of the theorem implies that f is a C^1 diffeomorphism on $\mathcal{Y}$ by Corollary B.1. Lemma B.4 in connection with condition 2, implies

$$\|\mathbf{f}(\mathbf{x})\|_2 \geq k_1 \cdot \|\mathbf{x}\|_2, \tag{4.255}$$

hence $\mathbf{x}=\mathbf{0}$ is the unique equilibrium of the system and it is left to show that it is exponentially stable. Therefore, consider the Lyapunov function candidate

$$V(\mathbf{x}) = \frac{1}{2}\mathbf{f}^T(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) = \frac{1}{2}\|\mathbf{f}(\mathbf{x})\|_2^2 \tag{4.256}$$

which is a continuously differentiable positive definite function since $\mathbf{f}$ is continuously differentiable and $\mathbf{f}(\mathbf{x}) \neq \mathbf{0}$ if $\mathbf{x} \neq \mathbf{0}$. The time derivative along the trajectories of the gradient system is

$$\dot{V}(\mathbf{x}) = -\mathbf{f}^T(\mathbf{x})\mathbf{J}_f(\mathbf{x})\mathbf{J}_f^T(\mathbf{x})\mathbf{f}(\mathbf{x}). \tag{4.257}$$

From condition 1, the matrix $\mathbf{Q}(\mathbf{x}) = \mathbf{J}_f(\mathbf{x})\mathbf{J}_f^T(\mathbf{x})$ is symmetric and positive definite and its smallest eigenvalue equals $\underline{\lambda}_Q = \underline{\sigma}_J^2$. By use of condition 1 in connection with Theorem B.20 and inequality (4.255) we arrive at

$$\dot{V}(\mathbf{x}) \leq -\underline{\sigma}_J^2 \|\mathbf{f}(\mathbf{x})\|_2^2 \leq -(k_0 k_1)^2 \cdot \|\mathbf{x}\|_2^2. \tag{4.258}$$

Also, the Lyapunov function candidate itself can be bounded in a similar way, using result (4.255) for the lower bound and conditions 3 of the theorem as well as mean value theorem (B.19) within Appendix B.4 in connection with Theorem B.18 for the upper bound.

$$\frac{k_1^2}{2}\|\mathbf{x}\|_2^2 \leq V(\mathbf{x}) \leq \frac{k_2^2}{2}\|\mathbf{x}\|_2^2 \tag{4.259}$$

Then all pre-conditions of Theorem C.2 (Lyapunov exponential stability) are fulfilled and hence the equilibrium is exponentially stable such that

$$\|\mathbf{x}(t)\|_2 \leq \|\mathbf{x}(t_0)\| \cdot \frac{k_2}{k_1} \cdot \exp\left[-\frac{1}{2}\left(\frac{k_0 k_1}{k_2}\right)^2 (t - t_0)\right] \quad \text{if} \quad \|\mathbf{x}(t_0)\|_2 < \frac{k_1}{k_2} r_0 \tag{4.260}$$

□

Normalization of Gradient Systems

Considering (4.260), the exponential convergence rate is, among others, dependent on k_0, the lower bound on the singular value of the map Jacobian. Since this lower bound has to hold uniformly for the whole flight conditions, it is potentially quite small, leading to a low convergence rate. By a small modification of the gradient system, the effect of a small k_0 is cancelled out which increases the convergence rate of the gradient

system. Therefore, assume that the singular value decomposition of the map Jacobian is given by

$$\mathbf{J}_f(\mathbf{x}) = \mathbf{V}(\mathbf{x})\boldsymbol{\Sigma}(\mathbf{x})\mathbf{U}^T(\mathbf{x}).$$

In the gradient system (4.253), we replace the Jacobian by

$$\dot{\mathbf{x}} = -\mathbf{U}(\mathbf{x})\boldsymbol{\Sigma}^{-1}(\mathbf{x})\mathbf{U}^T(\mathbf{x})\mathbf{f}(\mathbf{x}) \tag{4.261}$$

which modifies the Lyapunov function derivative (4.257)

$$\begin{aligned}\dot{V}(\mathbf{x}) &= -\mathbf{f}^T(\mathbf{x})\mathbf{J}_f(\mathbf{x})\mathbf{U}(\mathbf{x})\boldsymbol{\Sigma}^{-1}(\mathbf{x})\mathbf{U}^T(\mathbf{x})\mathbf{f}(\mathbf{x}) \\ &= -\mathbf{f}^T(\mathbf{x})\mathbf{V}(\mathbf{x})\boldsymbol{\Sigma}(\mathbf{x})\underbrace{\mathbf{U}^T(\mathbf{x})\mathbf{U}(\mathbf{x})}_{\mathbf{I}}\boldsymbol{\Sigma}^{-1}(\mathbf{x})\mathbf{U}^T(\mathbf{x})\mathbf{f}(\mathbf{x}) \\ &= -\mathbf{f}^T(\mathbf{x})\mathbf{V}(\mathbf{x})\underbrace{\boldsymbol{\Sigma}(\mathbf{x})\boldsymbol{\Sigma}^{-1}}_{\mathbf{I}}(\mathbf{x})\mathbf{U}^T(\mathbf{x})\mathbf{f}(\mathbf{x}) \\ \dot{V}(\mathbf{x}) &= -\mathbf{f}^T(\mathbf{x})\mathbf{V}(\mathbf{x})\mathbf{U}^T(\mathbf{x})\mathbf{f}(\mathbf{x})\end{aligned}$$

Now, since the $\mathbf{V}(\mathbf{x})$ and $\mathbf{U}(\mathbf{x})$ are orthonormal matrices, their singular values equal to 1 and hence, together with inequality (4.255) and Theorem B.20, we have

$$\dot{V} \le k_1^2 \|\mathbf{x}\|_2^2 \tag{4.262}$$

This equals to (4.258) with k_0=1 and hence, with bounds (4.359) on the Lyapunov function derivative, we conclude exponential stability according to Theorem C.2. independent of k_0.

$$\|\mathbf{x}(t)\|_2 \le \|\mathbf{x}(t_0)\| \cdot \frac{k_2}{k_1} \cdot \exp\left[-\frac{1}{2}\left(\frac{k_1}{k_2}\right)^2 (t-t_0)\right] \quad \text{if} \quad \|\mathbf{x}(t_0)\|_2 < \frac{k_1}{k_2} r_0 . \tag{4.263}$$

4.6 Variant 2: Nonaffine Controls used for Linearizing State Feedback

In this section nonaffine thrust vector controls are included into the linearizing state feedback as proposed in section 4.1.7, while the nonaffine controls are tuned by the algorithm , introduced in section 4.5. A comprehensive stability analysis of the closed loop adaptive system, including the NIC algorithm is presented, where switching σ as well as MMQ modification is incorporated into the algorithm. Contrary to variant 1, the nonaffine control map

$$\hat{\mathbf{w}} = \hat{\mathbf{g}}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) \tag{4.264}$$

is not adaptively estimated. In order to account for model uncertainties, the map $\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$, which is used for feedback linearization, is assumed to deviate from the real map by some $\boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$, which takes the role of the unmatched uncertainty.

$$\mathbf{w} = \mathbf{g}(\zeta, \eta, \mathbf{u}_N, \mathbf{d}) = \hat{\mathbf{g}}(\zeta, \eta, \mathbf{u}_N, \mathbf{d}) + \delta(\zeta, \eta, \mathbf{u}_N, \mathbf{d}). \tag{4.265}$$

Hence, external dynamics (4.92) become

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{a}_x(\zeta, \eta) + \begin{bmatrix}\mathbf{B}_L(\zeta, \eta) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u} \\ \hat{\mathbf{g}}(\zeta, \eta, \mathbf{u}_N, \mathbf{d})\end{pmatrix} + \delta(\zeta, \eta, \mathbf{u}_N, \mathbf{d})\right]. \tag{4.266}$$

Internal dynamics are of the form (4.90) and with neglect of actuators, the linearizing state feedback (4.93) is

$$\begin{pmatrix}\mathbf{u} \\ \mathbf{w}^*\end{pmatrix} = \begin{bmatrix}\hat{\mathbf{B}}^T(\zeta, \eta) \\ \mathbf{I}\end{bmatrix}\left[\hat{\mathbf{B}}(\zeta, \eta)\hat{\mathbf{B}}^T(\zeta, \eta) + \mathbf{I}\right]^{-1}\left[\mathbf{v} - \hat{\mathbf{a}}_x(\zeta, \eta)\right] \tag{4.267}$$

where the estimated decoupling matrix is parameterized according to (4.111) or (4.114). Notice that, in compliance with notation of section 4.5, now the desired virtual control is marked with an "*" and, according to (4.250), there is $\mathbf{u}_N^*$ such that

$$\mathbf{w}^* = \hat{\mathbf{g}}(\zeta, \eta, \mathbf{u}_N^*, \mathbf{d}). \tag{4.268}$$

The inverse w.r.t. $\mathbf{u}_N$

$$\mathbf{u}_N = \hat{\mathbf{g}}^{-1}(\zeta, \eta, \mathbf{w}, \mathbf{d}) \tag{4.269}$$

is assumed to exist, however, unknown. Hence also $\mathbf{u}_N^*$ is unknown. On sufficient conditions for existence of a unique inverse, refer to Theorem 4.5.

Recall that we want to employ the algorithm, proposed in section 4.5. Therefore we need to distinguish between the ideal virtual control $\mathbf{w}^*$ and the actual virtual control, which is assumed to be achieved by the current $\mathbf{u}_N$ of the NIC algorithm according to equation (4.264). We obtain the following external dynamics, if we apply the linearizing state feedback (4.267) to (4.266).

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{a}_x(\zeta, \eta) + \begin{bmatrix}\mathbf{B}(\zeta, \eta) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u} \\ \hat{\mathbf{g}}(\zeta, \eta, \mathbf{u}_N, \mathbf{d})\end{pmatrix} + \delta(\zeta, \eta, \mathbf{u}_N, \mathbf{d})\right]$$

$$\dot{\zeta} = \mathbf{J}\zeta + \underbrace{\mathbf{H}\left[\hat{\mathbf{a}}_x(\zeta, \eta) + \begin{bmatrix}\hat{\mathbf{B}}(\zeta, \eta) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u} \\ \mathbf{w}^*\end{pmatrix}\right]}_{+} \underbrace{- \mathbf{H}\left[\hat{\mathbf{a}}_x(\zeta, \eta) + \begin{bmatrix}\hat{\mathbf{B}}(\zeta, \eta) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u} \\ \mathbf{w}^*\end{pmatrix}\right]}_{-}$$
$$+ \mathbf{H}\left[\mathbf{a}_x(\zeta, \eta) + \begin{bmatrix}\mathbf{B}(\zeta, \eta) & \mathbf{I}\end{bmatrix}\begin{pmatrix}\mathbf{u} \\ \hat{\mathbf{g}}(\zeta, \eta, \mathbf{u}_N, \mathbf{d})\end{pmatrix} + \delta(\zeta, \eta, \mathbf{u}_N, \mathbf{d})\right]$$

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\mathbf{v} + \mathbf{H}\left[-\left(\hat{\mathbf{a}}_x(\zeta, \eta) - \mathbf{a}_x(\zeta, \eta)\right) - \left(\hat{\mathbf{B}}(\zeta, \eta) - \mathbf{B}(\zeta, \eta)\right)\mathbf{u} + \left(\hat{\mathbf{g}}(\zeta, \eta, \mathbf{u}_N, \mathbf{d}) - \mathbf{w}^*\right) + \delta(\zeta, \eta, \mathbf{u}_N, \mathbf{d})\right]$$

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{v} - \tilde{\mathbf{a}}_x(\zeta,\eta) - \tilde{\mathbf{B}}(\zeta,\eta)\mathbf{u} + \left(\hat{\mathbf{g}}(\zeta,\eta,\mathbf{u}_N) - \mathbf{w}^*\right) + \boldsymbol{\delta}(\zeta,\eta,\mathbf{u}_N,\mathbf{d})\right] \tag{4.270}$$

Thereby $\tilde{\mathbf{a}}_x(\zeta,\eta)$ and $\tilde{\mathbf{B}}(\zeta,\eta)$ are defined in (4.95), (4.96). Comparing this result to (4.94), it reveals that, beside the fact that there is no reaction deficit due to actuators and no model deviation of the nonaffine control effectiveness according to the recent assumptions, there is an additional term $\mathbf{g}(\zeta,\eta,\mathbf{u}_N,\mathbf{d})$-$\mathbf{w}^*$ due to the deviation between virtual control, required by the linearizing state feedback, and the assumed actual virtual control. achieved by the NIC algorithm. Using (4.110), (4.113) and (4.184), the external dynamics are

$$\dot{\zeta} = \mathbf{J}\zeta + \mathbf{H}\left[\mathbf{v} + \boldsymbol{\Theta}_x^T\boldsymbol{\varphi}(\zeta,\eta) - b(\zeta,\eta)\mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\mathbf{u} + \left(\hat{\mathbf{g}}(\zeta,\eta,\mathbf{u}_N,\mathbf{d}) - \mathbf{w}^*\right) + \boldsymbol{\delta}(\zeta,\eta,\mathbf{u}_N,\mathbf{d})\right]. \tag{4.271}$$

4.6.1 Reference Model, State Predictor and Error Dynamics

State Predictor and Prediction-Tracking Error

Since an additional term occurs in the external dynamics (4.271) due to the NIC algorithm, it appears to be useful for the subsequent stability analysis to add an according term to the reference dynamics (4.56).

$$\dot{\hat{\zeta}} = \mathbf{J}\hat{\zeta} + \mathbf{H}\left[\hat{\mathbf{v}}_R + \hat{\mathbf{g}}(\zeta,\eta,\mathbf{u}_N,\mathbf{d}) - \mathbf{w}^*\right] \tag{4.272}$$

The additional term renders the system rather a *state predictor* ([Lav07a]) than a reference model for reasons, stated in a few lines. According to (4.55), the reference pseudo control $\mathbf{v}_R$ is changed to a predictor based pseudo control.

$$\hat{\mathbf{v}}_R = -\mathbf{K}^T\hat{\zeta} + \mathbf{A}_0\mathbf{y}_C \tag{4.273}$$

System (4.272) is denoted as a state predictor since, if (4.273) is applied to the pseudo control of the external dynamics (4.270), then (4.272) actually predicts the external states if model deviations are absent.

Notice that the predictor contains feedback from the plant such that we cannot assume a priori that the predictor dynamics are stable. However, boundedness of the predictor states will be shown by investigation of Lyapunov stability of the *prediction-tracking error* between predictor and reference model (4.49)

$$\mathbf{e} = \hat{\zeta} - \zeta_R \tag{4.274}$$

and the fact that the reference dynamics are stable. With (4.272), (4.273), (4.49), and (4.55), the tracking error dynamics read as

$$\begin{aligned}\dot{\mathbf{e}} &= \overbrace{\mathbf{J}\hat{\boldsymbol{\zeta}} + \mathbf{H}\left[-\mathbf{K}^T\hat{\boldsymbol{\zeta}} + \mathbf{A}_0\mathbf{y}_C + \hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{w}^*\right]}^{\dot{\hat{\boldsymbol{\zeta}}}} - \overbrace{\left[\mathbf{J}\boldsymbol{\zeta}_R + \mathbf{H}\left(-\mathbf{K}^T\boldsymbol{\zeta}_R + \mathbf{A}_0\mathbf{y}_C\right)\right]}^{\dot{\boldsymbol{\zeta}}_R} \\ &= \mathbf{J}\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_R\right) - \mathbf{H}\mathbf{K}^T\left(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_R\right) + \mathbf{H}\left[\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{w}^*\right] \\ &= \left(\mathbf{J} - \mathbf{H}\mathbf{K}^T\right)\mathbf{e} + \mathbf{H}\left[\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{w}^*\right] \\ \dot{\mathbf{e}} &= \mathbf{A}_R\mathbf{e} + \mathbf{H}\left[\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{w}^*\right]\end{aligned} \tag{4.275}$$

where $\mathbf{A}_R$=$\mathbf{J}$-$\mathbf{H}\mathbf{K}^T$ is stable.

Prediction Error

Since boundedness of the predictor states is shown separately, using the prediction-tracking error dynamics, the proof of ultimate boundedness for the parameter-estimation-errors and the plant states is shown similarly to section 4.3, however, in this case the *prediction error* is used instead of the tracking error.

$$\hat{\mathbf{e}} = \hat{\boldsymbol{\xi}} - \boldsymbol{\zeta}. \tag{4.276}$$

In order to cancel the state dependent uncertainty in the external dynamics (4.271), we choose for the pseudo control

$$\boldsymbol{\nu} = \hat{\boldsymbol{\nu}}_R + \boldsymbol{\nu}_E + \boldsymbol{\nu}_A. \tag{4.277}$$

Comparing with pseudo control in section 4.3 – equation (4.121) – the reference pseudo control $\boldsymbol{\nu}_R$ is replaced is by the predictor based pseudo control (4.273) while the adaptive term $\boldsymbol{\nu}_A$ remains the same as defined in (4.122) and the error feedback term $\boldsymbol{\nu}_E$ is now computed with the prediction error instead of the tracking error.

$$\boldsymbol{\nu}_E = \mathbf{C}^T\hat{\mathbf{e}} \tag{4.278}$$

Hence, if pseudo control (4.277) is applied to the external dynamics (4.271), we obtain

$$\dot{\boldsymbol{\zeta}} = \mathbf{J}\boldsymbol{\zeta} + \mathbf{H}\left[\hat{\boldsymbol{\nu}}_R + \mathbf{C}^T\hat{\mathbf{e}} - \tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta}) - b(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\mathbf{u} + \left(\mathbf{g}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N) - \mathbf{w}^*\right) + \boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right] \tag{4.279}$$

where $\tilde{\boldsymbol{\Theta}}_x = \hat{\boldsymbol{\Theta}}_x - \boldsymbol{\Theta}_x$. With definition of predictor (4.272), the prediction error dynamics are

$$\begin{aligned}\dot{\hat{\mathbf{e}}} &= \mathbf{J}\hat{\boldsymbol{\xi}} + \mathbf{H}\left[\hat{\boldsymbol{\nu}}_R + \hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{w}^*\right] \\ &\quad - \mathbf{J}\boldsymbol{\zeta} - \mathbf{H}\left[\hat{\boldsymbol{\nu}}_R + \mathbf{C}^T\hat{\mathbf{e}} - \tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta}) - b(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\mathbf{u} + \left(\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{w}^*\right) + \boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right] \\ \dot{\hat{\mathbf{e}}} &= \mathbf{J}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\zeta}\right) - \mathbf{H}\mathbf{C}^T\hat{\mathbf{e}} + \mathbf{H}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta}) + b(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\mathbf{u} - \boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right] \\ \dot{\hat{\mathbf{e}}} &= \mathbf{A}_E\hat{\mathbf{e}} + \mathbf{H}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta}) + b(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\mathbf{u} - \boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]\end{aligned} \tag{4.280}$$

where $\mathbf{A}_E$=$\mathbf{J}$-$\mathbf{H}\mathbf{C}^T$ is stable.

4.6.2 Lyapunov Stability Analysis

Beside the common states, involved into an MRAC system, namely

- prediction error: $\hat{\mathbf{e}}$
- parameter-estimation-errors: $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Lambda}}_L$

due to the NIC design we additionally have

- prediction-tracking error $\mathbf{e}$
- command of the NIC tuning algorithm $\mathbf{u}_N$.

Therefore, the Lyapunov analysis is divided into two parts. In the first part, we will use the prediction error dynamics to design the updates for the parameter estimates and show boundedness of $\hat{\mathbf{e}}$, $\tilde{\boldsymbol{\Theta}}_x$ and $\tilde{\boldsymbol{\Lambda}}_L$. In a second step, the prediction-tracking error dynamics are used to design the online gradient minimization system and to show boundedness of $\mathbf{e}$, $\mathbf{u}_N$. Finally, from boundedness of $\hat{\mathbf{e}}$, $\mathbf{e}$ one can conclude boundedness of predictor $\hat{\zeta}$ and plant ζ, using relations

$$\hat{\zeta} = \zeta_R + \mathbf{e} \tag{4.281}$$

$$\zeta = \zeta_R + \mathbf{e} - \hat{\mathbf{e}} \tag{4.282}$$

and the fact that the reference model states ζ_R are bounded a priori. One might hope that the two parts of the stability analysis are independent of each other, however this is not true. For Lyapunov analysis, we have to impose upper bounds on unmatched uncertainty $\boldsymbol{\delta}(\zeta, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d})$ and the nonlinear map $\mathbf{g}(\zeta, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d})$, which depend on the plant state ζ. These bounds, among others, depend on the set of admissible plant states, let us say $\|\zeta\|_2 \le \overline{\zeta}$ for some $\overline{\zeta} > 0$. Each single part of the stability analysis considers boundedness of $\hat{\mathbf{e}}$ and $\mathbf{e}$ respectively (assuming that the assumptions hold). Hence, the bounds on the error states cannot be set decoupled from each other, but, by (4.282), have to fulfill

$$\|\zeta\|_2 \le \|\zeta_R\|_2 + \|\mathbf{e}\|_2 + \|\hat{\mathbf{e}}\|_2 . \tag{4.283}$$

Further, since the reference model is stable, it satisfies

$$\|\zeta_R(t)\|_2 \le \overline{\zeta}_R \quad \text{for all} \quad t \ge t_0 \tag{4.284}$$

for some $\overline{\zeta}_R > 0$ if the exogenous input satisfies

$$\|\mathbf{y}_c(t)\|_2 \le \bar{y} \quad \text{for all} \quad t \ge t_0 \tag{4.285}$$

for a sufficiently small $\bar{y} > 0$, provided initial conditions are appropriate. Consequently, if both parts prove

$$\|\hat{\mathbf{e}}(t)\|_2 \le r_{\hat{e}} \qquad \|\mathbf{e}(t)\|_2 \le r_e \qquad \text{for all} \quad t \ge t_0 \tag{4.286}$$

for some $r_{\hat{e}}, r_e > 0$ then ζ remains within the admissible set if

$$r_{\hat{e}} + r_e + \bar{\zeta}_R \leq \bar{\zeta} \,. \tag{4.287}$$

Take a look at Figure 4.13 with part 1, the classical MRAC Lyapunov analysis, proving boundedness of prediction error $\hat{\mathbf{e}}$ and parameter-estimation-errors $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Lambda}}_L$ and part 2, proving boundedness of NIC design state $\mathbf{u}_N$ and prediction-tracking error $\mathbf{e}$. In order to fulfill coupling condition (4.287), part 1 requires an upper bound on $\mathbf{e}$ a priori, which is established by part 2. Part 2 in turn requires an upper bound on $\hat{\mathbf{e}}$ a priori, which is established by part 1. Proving crosswise boundedness of one part by assuming boundedness of the respective other part, while and the bounds fulfill (4.287), establishes boundedness of the whole system since the considered error state can only escape from its bounded set if the respective other error state has escaped before and hence both error states restrict each other to their respective sets.

An additional assumption, needed for the first part, is the restriction of the nonaffine controls to their valid set, however, this is assured by part 2.

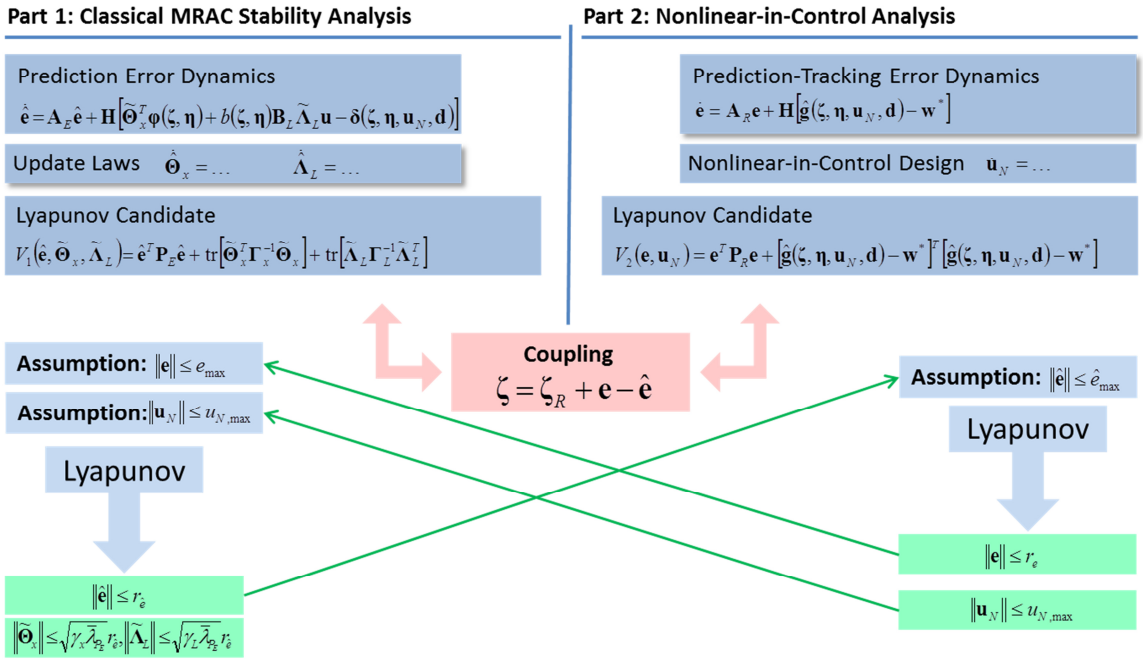


Figure 4.13 Lyapunov Analysis for MRAC with NIC Design

Part 1: Classical MRAC Stability Analysis

For part 1, bounds on the ideal parameters as well as unmatched uncertainty have to be imposed, analogously to assumptions A and B of section 4.3.

Assumption D: *Bounds on ideal parameters and unmatched uncertainty*

$$\left\|\boldsymbol{\Theta}_x\right\|_{\boldsymbol{\Gamma}_x} \leq \theta_x \qquad \left\|\boldsymbol{\Lambda}_L^T\right\|_{\boldsymbol{\Gamma}_L} \leq \lambda_L \tag{4.288}$$

for some $\theta_x, \lambda_L > 0$ and symmetric positive definite matrices $\mathbf{\Gamma}_x, \mathbf{\Gamma}_L$. Additionally

$$\left\|\boldsymbol{\delta}(\zeta, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d})\right\|_2 \leq D \tag{4.289}$$

for some $D \geq 0$ in a set

$$(\zeta, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) \in \overline{\mathcal{B}}_{\bar{\zeta}} \times \mathcal{D}_\eta \times \mathcal{U} \times \mathcal{D}. \tag{4.290}$$

Thereby $\overline{\mathcal{B}}_{\bar{\zeta}} = \left\{\zeta \in \mathbb{R}^r \,\middle|\, \|\zeta\|_2 \leq \bar{\zeta}\right\}$ for some $\bar{\zeta} > 0$, $\mathcal{D}_\eta$ is a set, where the internal states can be restricted, $\mathcal{U}$ is a convex set containing the origin and $\mathcal{D}$ is a set, where $\mathbf{d}$ can be restricted to.

As Lyapunov function candidate, we choose

$$V_1\left(\hat{\mathbf{e}}, \tilde{\mathbf{\Theta}}_x, \tilde{\mathbf{\Lambda}}_L\right) = \hat{\mathbf{e}}^T \mathbf{P}_E \hat{\mathbf{e}} + \operatorname{tr}\left[\tilde{\mathbf{\Theta}}_x^T \gamma_x^{-1} \mathbf{\Gamma}_x^{-1} \tilde{\mathbf{\Theta}}_x\right] + \operatorname{tr}\left[\tilde{\mathbf{\Lambda}}_L \gamma_L^{-1} \mathbf{\Gamma}_L^{-1} \tilde{\mathbf{\Lambda}}_L^T\right] \tag{4.291}$$

where γ_x, and γ_L positive scalars, denoted as learning rates and $\mathbf{\Gamma}_x$, $\mathbf{\Gamma}_L$ are symmetric positive definite matrices of appropriate dimension, which adopt the role of a weighting matrix. $\mathbf{P}_E$ is the symmetric positive definite solution of the Lyapunov equation

$$\mathbf{A}_E^T \mathbf{P}_E + \mathbf{P}_E \mathbf{A}_E = -\mathbf{Q}_E \tag{4.292}$$

for some symmetric positive definite $\mathbf{Q}_E$. The following arguments are very similar those in sections 4.3 and 4.4 and hence we will only state the key results in this case. Taking the time derivative of V_1, we arrive at an expression analogous to (4.130).

$$\begin{aligned}\dot{V}_1 = &-\hat{\mathbf{e}}^T \mathbf{Q}_E \hat{\mathbf{e}} - 2\hat{\mathbf{e}}^T \mathbf{P}_E \mathbf{H} \boldsymbol{\delta}(\zeta, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) \\ &+ 2\operatorname{tr}\left[\tilde{\mathbf{\Theta}}_x^T \mathbf{\Gamma}_x^{-1}\left(\gamma_x^{-1} \dot{\tilde{\mathbf{\Theta}}}_x + \mathbf{\Gamma}_x \boldsymbol{\varphi}(\zeta, \boldsymbol{\eta}) \hat{\mathbf{e}}^T \mathbf{P}_E \mathbf{H}\right)\right] + 2\operatorname{tr}\left[\tilde{\mathbf{\Lambda}}_L \mathbf{\Gamma}_L^{-1}\left(\gamma_x^{-1} \dot{\tilde{\mathbf{\Lambda}}}_L^T + \mathbf{\Gamma}_L b(\zeta, \boldsymbol{\eta}) \mathbf{u} \hat{\mathbf{e}}^T \mathbf{P}_E \mathbf{H} \mathbf{B}_L\right)\right]\end{aligned} \tag{4.293}$$

For update of the parameter estimates we choose

$$\begin{aligned}\dot{\hat{\mathbf{\Theta}}}_x &= \gamma_x \operatorname{Proj}\left[\hat{\mathbf{\Theta}}_x, -\mathbf{\Gamma}_x \boldsymbol{\varphi}(\zeta, \boldsymbol{\eta}) \hat{\mathbf{e}}^T \mathbf{P}_E \mathbf{H} - \mathbf{M}_x\right]_{\varepsilon_x, \theta_{x,\max}, \mathbf{\Gamma}_x} \\ \dot{\hat{\mathbf{\Lambda}}}_L^T &= \gamma_L \operatorname{Proj}\left[\hat{\mathbf{\Lambda}}_L^T, -\mathbf{\Gamma}_L b(\zeta, \boldsymbol{\eta}) \mathbf{u} \hat{\mathbf{e}}^T \mathbf{P}_E \mathbf{H} \mathbf{B}_L - \mathbf{M}_L^T\right]_{\varepsilon_L, \lambda_{L,\max}, \mathbf{\Gamma}_L}\end{aligned} \tag{4.294}$$

with projection operator (Appendix D.1), ε_x, ε_L, $\theta_{x,max}$, $\lambda_{L,max}$ defined in (4.133), (4.135) and modifications $\mathbf{M}_x$, $\mathbf{M}_L$, which will be defined in a few lines. We obtain

$$\dot{V} \leq -\hat{\mathbf{e}}^T \mathbf{Q}_E \hat{\mathbf{e}} - 2\hat{\mathbf{e}}^T \mathbf{P}_E \mathbf{H} \boldsymbol{\delta}(\zeta, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) - 2\operatorname{tr}\left[\tilde{\mathbf{\Theta}}_x^T \mathbf{\Gamma}_x^{-1} \mathbf{M}_x\right] - 2\operatorname{tr}\left[\tilde{\mathbf{\Lambda}}_L \mathbf{\Gamma}_L^{-1} \mathbf{M}_L^T\right]. \tag{4.295}$$

Now we apply switching σ- and MMQ modification

$$\begin{aligned}\mathbf{M}_x &= \sigma_x \bar{f}_{\varepsilon_x, \theta_x, \mathbf{\Gamma}_x}\left(\hat{\mathbf{\Theta}}_x\right) \hat{\mathbf{\Theta}}_x + \kappa \mathbf{\Gamma}_x \frac{\mathbf{Q}_x}{\xi} \tilde{\mathbf{C}}^T \mathbf{H} \\ \mathbf{M}_L &= \sigma_L \bar{f}_{\varepsilon_L, \lambda_L, \mathbf{\Gamma}_L}\left(\hat{\mathbf{\Lambda}}_L^T\right) \hat{\mathbf{\Lambda}}_L + \kappa \mathbf{B}_L^T \mathbf{H}^T \tilde{\mathbf{C}} \frac{\mathbf{Q}_L^T}{\xi} \mathbf{\Gamma}_L\end{aligned} \tag{4.296}$$

where

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_x \\ \mathbf{Q}_L \end{bmatrix}. \tag{4.297}$$

and $\mathbf{Q}_x$, , $\mathbf{Q}_L$ are matrices whose columns are filtered versions of regressors as defined in (4.210), (4.211) and (4.213), σ_x, $\sigma_L > 0$ are the gains of switching σ-modification, $\kappa > 0$ is the gain of MMQ modification, $\bar{f}_{\varepsilon_x,\theta_x,\Gamma_x}(\cdot)$, $\bar{f}_{\varepsilon_L,\lambda_L,\Gamma_L}(\cdot)$ are the switching functions, defined in equations (D.1), (D.12) in Appendix D and

$$\xi = 1 + \|\mathbf{Q}\|_F^2 \tag{4.298}$$

is a normalizing factor. $\tilde{\mathbf{C}}$ is the filtered uncertainty estimation error, which is computed based on external dynamics (4.270). We define, by slight abuse in notation in time and frequency domain,

$$\mathbf{c}_i(s) = G_i(s)(s\mathbf{I} - \mathbf{J})\zeta(s) + G_i(s)\mathbf{H}\left[-\mathbf{v}(s) + \left[b(\zeta,\boldsymbol{\eta})\mathbf{B}_L\hat{\boldsymbol{\Lambda}}_L\mathbf{u}\right](s) - \left[\mathbf{g}(\zeta,\boldsymbol{\eta},\mathbf{u}_N) - \mathbf{w}^*\right](s)\right] \tag{4.299}$$

with strictly proper and stable transfer functions $G_i(s)$, $i=1,\ldots,N$, as defined in (4.182). According to external dynamics (4.270), equation (4.299) equals the filtered true uncertainty

$$\mathbf{c}_i(s) = \mathbf{H}\left(\mathbf{W}^T\mathbf{q}_i(s) + \mathbf{q}_\delta(s)\right) \tag{4.300}$$

where

$$\mathbf{W}^T = \begin{bmatrix} \boldsymbol{\Theta}_x^T & \mathbf{B}_L\boldsymbol{\Lambda}_L \end{bmatrix} \tag{4.301}$$

is the compound ideal parameter and

$$\mathbf{q}_i^T(s) = \begin{pmatrix} \mathbf{q}_{x,i}^T(s) & \mathbf{q}_{L,i}^T(s) \end{pmatrix} \tag{4.302}$$

is the compound filtered regressor, where $\mathbf{q}_{x,i}(s)$, $\mathbf{q}_{L,i}(s)$ are defined according to (4.210). Analogously, the filtered estimated uncertainty is

$$\hat{\mathbf{c}}_i(t) = \mathbf{H}\hat{\mathbf{W}}^T\mathbf{q}_i(t) \tag{4.303}$$

where $\hat{\mathbf{W}}$ is the adaptive estimate for (4.301) and the filtered uncertainty estimation error reads as

$$\tilde{\mathbf{c}}_i(t) := \hat{\mathbf{c}}_i(t) - \mathbf{c}_i(t) = \mathbf{H}\left(\tilde{\mathbf{W}}^T(t)\mathbf{q}_i(t) - \mathbf{q}_{\delta,i}(t)\right). \tag{4.304}$$

If we stack the N vectors into a matrix, we finally get

$$\tilde{\mathbf{C}}(t) := \begin{bmatrix} \tilde{\mathbf{c}}_1(t) & \cdots & \tilde{\mathbf{c}}_N(t) \end{bmatrix} = \mathbf{H}\left(\tilde{\mathbf{W}}^T(t)\mathbf{Q}(t) - \mathbf{Q}_\delta(t)\right). \tag{4.305}$$

where $\mathbf{Q}_\delta$ is defined in (4.213). Inserting modifications (4.296) into (4.295), we obtain.

$$\begin{aligned}\dot{V}_1 \le & -\hat{\mathbf{e}}^T\mathbf{Q}_E\hat{\mathbf{e}} \quad -2\hat{\mathbf{e}}^T\mathbf{P}_E\mathbf{H}\boldsymbol{\delta} - 2\kappa\mathrm{tr}\left[\tilde{\mathbf{W}}^T\frac{\mathbf{Q}}{\xi}\mathbf{Q}^T\tilde{\mathbf{W}}\right] + 2\kappa\mathrm{tr}\left[\tilde{\mathbf{W}}^T\frac{\mathbf{Q}}{\xi}\mathbf{Q}_\delta^T\right] \\ & -2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 + \mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\boldsymbol{\Theta}_x\right]\right) - 2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 + \mathrm{tr}\left[\tilde{\boldsymbol{\Lambda}}_L\boldsymbol{\Gamma}_L^{-1}\boldsymbol{\Lambda}_L^T\right]\right)\end{aligned} \tag{4.306}$$

Then, an upper bound is given by

$$\begin{aligned}\dot{V}_1 \le & -\underline{\lambda}_{Q_E}\left\|\hat{\mathbf{e}}\right\|_2^2 \quad +2\overline{\lambda}_{P_E}D\left\|\hat{\mathbf{e}}\right\|_2 - 2\frac{\kappa}{\xi}\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F^2 \quad +2\frac{\kappa}{\xi}mLD\sqrt{N}\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F \\ & -2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 - \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\theta_{x,\max}\right) - 2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 - \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\lambda_{L,\max}\right)\end{aligned} \tag{4.307}$$

where we have used assumption C in section 4.4.3 and assumption D together with (4.133), (4.135). The upper bound (4.307) is further divided into partial functions, each of which depends on only one variable, according to(4.147) and

$$h_q\left(\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F\right) = -2\frac{\kappa}{\xi}\left(\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F^2 + mLD\sqrt{N}\left\|\tilde{\mathbf{W}}^T\mathbf{Q}\right\|_F\right) \tag{4.308}$$

(4.235). These are all concave parabolas and have upper bounds according to (4.148), (4.236), where we have used $\xi \geq 1$. Moreover, we obtain three functions by taking the negative value of one partial function and the worst case bounds of the respective others, analogously to (4.237).

$$\begin{aligned}\gamma_{\hat{e}}\left(\left\|\hat{\mathbf{e}}\right\|_2\right) &= \underline{\lambda}_{Q_E}\left\|\hat{\mathbf{e}}\right\|_2^2 - 2\overline{\lambda}_{P_E}D\left\|\hat{\mathbf{e}}\right\|_2 - V_{\hat{e}} - V_Q \\ \gamma_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\right) &= 2\sigma_x\bar{f}\left(\hat{\boldsymbol{\Theta}}_x\right)\left[\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2 - \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\theta_{x,\max}\right] - V_D - V_x - V_Q \\ \gamma_L\left(\left\|\tilde{\boldsymbol{\Lambda}}_L\right\|_{\boldsymbol{\Gamma}_L}\right) &= 2\sigma_L\bar{f}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\left[\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 - \left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\lambda_{L,\max}\right] - V_D - V_x - V_Q\end{aligned} \tag{4.309}$$

Thereby, we have defined

$$\begin{aligned}V_{\hat{e}} &= \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_L\lambda_{L,\max}^2\right) \qquad V_x = \frac{1}{2}\sigma_L\lambda_{L,\max}^2 \qquad V_L = \frac{1}{2}\sigma_x\theta_{x,\max}^2 \\ V_D &= \frac{\left(\overline{\lambda}_{P_E}D\right)^2}{\underline{\lambda}_{Q_E}} \qquad\qquad V_Q = \frac{1}{2}\kappa\left(mLD\right)^2N\end{aligned} \tag{4.310}$$

and obtain

$$\dot{V}_1 \le -\max\left[\gamma_{\hat{e}}\left(\left\|\hat{\mathbf{e}}\right\|_2\right), \gamma_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_2\right), \gamma_L\left(\left\|\tilde{\boldsymbol{\Theta}}_L\right\|_2\right)\right]. \tag{4.311}$$

Further, class $\mathcal{K}$ functions α_i, β_i, according to Corollary 3.2, that enclose the Lyapunov candidate (4.291) are given by

$$\alpha_{\hat{e}}\left(\left\|\hat{\mathbf{e}}\right\|_2\right) = \underline{\lambda}_{P_E}\left\|\hat{\mathbf{e}}\right\|_2^2,\ \beta_{\hat{e}}\left(\left\|\hat{\mathbf{e}}\right\|_2\right) = \overline{\lambda}_{P_E}\left\|\hat{\mathbf{e}}\right\|_2^2,\ \alpha_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\right) = \beta_x\left(\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}\right) = \gamma_x^{-1}\left\|\tilde{\boldsymbol{\Theta}}_x\right\|_{\boldsymbol{\Gamma}_x}^2,\ \alpha_L\left(\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}\right) = \beta_L\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L} = \gamma_L^{-1}\left\|\tilde{\boldsymbol{\Lambda}}_L^T\right\|_{\boldsymbol{\Gamma}_L}^2 \tag{4.312}$$

In compliance with the introductory remarks on the procedure of the whole proof (Figure 4.13), we use r_e, $r_{\hat{e}}$ that satisfy (4.287) and assume, a priori that $\|\mathbf{e}\| \le r_e$ and use $r_{\hat{e}}$ according to Corollary 3.2. Moreover $r_x = r_L = \infty$ analogously to (4.152) and hence we obtain

$$\underline{u} = \min\left[\alpha_{\hat{e}}(r_{\hat{e}}), \alpha_x(r_x), \alpha_L(r_L)\right] = \underline{\lambda}_{P_E} r_{\hat{e}}^2 . \tag{4.313}$$

Moreover, the radii that define the sets of admissible initial conditions evaluate to

$$\rho_{\hat{e}} = \beta_{\hat{e}}^{-1}\left(\frac{1}{3}\underline{u}\right) = , \ \rho_x = \beta_x^{-1}\left(\frac{1}{3}\underline{u}\right), \ \rho_L = \beta_L^{-1}\left(\frac{1}{3}\underline{u}\right)$$
$$\rho_{\hat{e}} = \sqrt{\frac{\underline{\lambda}_{P_E}}{3\overline{\lambda}_{P_E}}} r_{\hat{e}} \ , \ \rho_x = \sqrt{\frac{\gamma_x \underline{\lambda}_{P_E}}{3}} r_{\hat{e}} \ , \ \rho_L = \sqrt{\frac{\gamma_L \underline{\lambda}_{P_E}}{3}} r_{\hat{e}} . \tag{4.314}$$

μ_i, according to Corollary 3.2, are obtained by requiring $\gamma_i(\mu_i)=K_0$ for some $K_0>0$.

$$\mu_{\hat{e}} = \frac{\overline{\lambda}_{P_E} D}{\underline{\lambda}_{Q_E}} + \sqrt{\left(\frac{\overline{\lambda}_{P_E} D}{\underline{\lambda}_{Q_E}}\right)^2 + \frac{V_{\hat{e}} + V_Q + K_0}{\underline{\lambda}_{Q_E}}} \quad \mu_x = \frac{\theta_{x,\max}}{2} + \sqrt{\left(\frac{\theta_{x,\max}}{2}\right)^2 + \frac{V_D + V_x + V_Q + K_0}{2\sigma_x}}$$
$$\mu_L = \frac{\lambda_{L,\max}}{2} + \sqrt{\left(\frac{\lambda_{L,\max}}{2}\right)^2 + \frac{V_D + V_L + V_Q + K_0}{2\sigma_L}} \tag{4.315}$$

Finally, conditions on ultimate boundedness $\mu_i < \rho_I$ result in inequalities

$$V_e + V_Q + K_0 < k_{\hat{e}}(r_{\hat{e}}) \ , \ \frac{V_D + V_x + V_Q + K_0}{\sigma_x} < k_x(r_{\hat{e}}, \gamma_x) \ , \ \frac{V_D + V_L + V_Q + K_0}{\sigma_L} < k_L(r_{\hat{e}}, \gamma_L) \tag{4.316}$$

where

$$k_{\hat{e}}(r_{\hat{e}}) := \frac{\underline{\lambda}_{Q_E} \underline{\lambda}_{P_E}}{3\overline{\lambda}_{P_E}} r_{\hat{e}}^2 - \sqrt{\frac{4}{3}\underline{\lambda}_{P_E}\overline{\lambda}_{P_E}} D r_{\hat{e}}$$
$$k_x(r_{\hat{e}}, \gamma_x) := \frac{2}{3}\gamma_x \underline{\lambda}_{P_E} r_{\hat{e}}^2 - \sqrt{\frac{4}{3}\gamma_x \underline{\lambda}_{P_E}} \theta_{x,\max} r_{\hat{e}} \tag{4.317}$$
$$k_L(r_{\hat{e}}, \gamma_L) := \frac{2}{3}\gamma_L \underline{\lambda}_{P_E} r_{\hat{e}}^2 - \sqrt{\frac{2}{3}\gamma_L \underline{\lambda}_{P_E}} \lambda_{L,\max} r_{\hat{e}}$$

and, inserting (4.310) into (4.316) results in a set of inequalities for the gains of switching σ- and MMQ modification

$$\begin{array}{rrrrrr} \theta_{x,\max}^2 \cdot \sigma_x & +\lambda_{L,\max}^2 \cdot \sigma_L & + k_Q \kappa & + 2K_0 & < & 2k_{\hat{e}}(r_{\hat{e}}) \\ -2k_x(r_{\hat{e}}, \gamma_x) \cdot \sigma_x & +\lambda_{L,\max}^2 \cdot \sigma_L & + k_Q \kappa & + 2K_0 & < & -2V_D \\ \theta_{x,\max}^2 \cdot \sigma_x & -2k_L(r_{\hat{e}}, \gamma_L) \cdot \sigma_L & + k_Q \kappa & + 2K_0 & < & -2V_D \end{array} \tag{4.318}$$

where

$$k_Q := (mLD)^2 N . \tag{4.319}$$

Note that, analogously to section 4.3, (4.318) has a solution σ_x, σ_L, κ, $K_0 > 0$ only if $k_{\hat{e}}(r_{\hat{e}}) \geq k_{\hat{e}}^*$, $k_x(r_{\hat{e}}) \geq k_x^*$, $k_L(r_{\hat{e}}) \geq k_L^*$ for some sufficiently large $k_{\hat{e}}^*, k_x^*, k_L^* > 0$. Hence, by solving equations (4.317), we obtain a lower bound on $r_{\hat{e}}$ and learning rates γ_x, γ_L

$$r_{\hat{e}} \geq \underline{\hat{e}} \quad , \quad \gamma_x \geq \underline{\gamma}_x \quad , \quad \gamma_L \geq \underline{\gamma}_L \tag{4.320}$$

where

$$\begin{aligned}
\underline{\hat{e}} &= \frac{\sqrt{3}\overline{\lambda}_{P_E}}{\underline{\lambda}_{Q_E}} \sqrt{\frac{\overline{\lambda}_{P_E}}{\underline{\lambda}_{P_E}}} D + \sqrt{\left(\frac{\sqrt{3}\overline{\lambda}_{P_E}}{\underline{\lambda}_{Q_E}} \sqrt{\frac{\overline{\lambda}_{P_E}}{\underline{\lambda}_{P_E}}} D \right)^2 + \frac{3\overline{\lambda}_{P_E} k_{\hat{e}}^*}{\underline{\lambda}_{Q_E} \underline{\lambda}_{P_E}}} \\
\underline{\gamma}_x &= \frac{1}{r_{\hat{e}}^2} \left(\sqrt{\frac{3}{4\underline{\lambda}_{P_E}}} \theta_{x,\max} + \sqrt{\left(\sqrt{\frac{3}{4\underline{\lambda}_{P_E}}} \theta_{x,\max} \right)^2 + \frac{3k_x^*}{2\underline{\lambda}_{P_E}}} \right)^2 \\
\underline{\gamma}_L &= \frac{1}{r_{\hat{e}}^2} \left(\sqrt{\frac{3}{4\underline{\lambda}_{P_E}}} \lambda_{L,\max} + \sqrt{\left(\sqrt{\frac{3}{4\underline{\lambda}_{P_E}}} \lambda_{L,\max} \right)^2 + \frac{3k_L^*}{2\underline{\lambda}_{P_E}}} \right)^2
\end{aligned} \tag{4.321}$$

and the system states are ultimately bounded if $r_{\hat{e}}$ simultaneously fulfills (4.320) and (4.287) for an assumed r_e and the learning rates are chosen sufficiently large. The result of part 1 is summarized in the following.

Lemma 4.2 MRAC Variant 2 – Part 1

Consider:

- stable **reference dynamics** (4.56) and **state predictor** (4.272)
- **plant dynamics** in Byrnes-Isidori normal form (4.266), subjected to linearizing state feedback (4.267) with pseudo control (4.277), adaptive term (4.122), error feedback (4.278) and predictor based pseudo control (4.273)
- **parameterizations** (4.107) – (4.110) for state dependent uncertainty, (4.111) or (4.114) for affine control effectiveness together with (4.184)
- **parameter-estimation-errors** $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Lambda}}_L$, defined in (4.124), (4.112)
- **update law** (4.294) using
  - symm. pos. def. solution $\mathbf{P}_E$ of (4.292) for some symm. pos. def. $\mathbf{Q}_E$
  - positive learning rates γ_x, γ_L
  - projection operator with parameters ε_x, $\varepsilon_L>0$, $\theta_{x,max}$, $\lambda_{L,max}>0$ and symmetric positive definite weighting matrices $\boldsymbol{\Gamma}_x$, $\boldsymbol{\Gamma}_L$
  - switching □-modification with modification gains $\sigma_x, \sigma_x>0$ and MMQ modification with modification gain $\kappa>0$ and N filters according to (4.296)
- Sets: $\overline{\mathcal{B}_{\bar{\zeta}}} \in \mathbb{R}^r$, $\mathcal{D}_\eta \in \mathbb{R}^{n-r}$, $\mathcal{D} \in \mathbb{R}^d$ for some $\bar{\zeta} > 0$

The **prediction error** (4.276) and $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Lambda}}_L$ are **uniformly ultimately bounded**, if

- assumption D is fulfilled for some $\theta_x, \lambda_L, D>0$ and $\theta_{x,max}, \lambda_{L,max}$ defined by (4.133), (4.135)
- $(\boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) \in \mathcal{D}_\eta \times \mathcal{U} \times \mathcal{D}$ for all $t \geq t_0$
- assumption C is fulfilled for some $L>0$
- the reference model state is bounded according to (4.284) for some $\bar{\zeta}_R > 0$
- $r_{\hat{e}}$ simultaneously satisfies
  - $r_{\hat{e}} \geq \underline{\hat{e}}$, according to (4.320) where $\underline{\hat{e}}$ is defined in (4.321) with $k_{\hat{e}}^*, k_x^*, k_L^*$ such that inequalities (4.318) are solved for $\sigma_x, \sigma_L, \kappa, K_0 > 0$, and
  - $r_{\hat{e}} + r_e + \bar{\zeta}_R \leq \bar{\zeta}$ according to (4.287) – for some $r_e>0$ such that $\|\mathbf{e}\|_2 \leq r_e$ holds for the prediction-tracking error (4.274)
- the learning rates are chosen such that $\gamma_x \geq \underline{\gamma}_x, \gamma_L \geq \underline{\gamma}_L$ according to (4.320) where $\underline{\gamma}_x, \underline{\gamma}_L$ are defined in (4.321)
- initial conditions: $\|\hat{\mathbf{e}}(t_0)\|_2 \leq \delta_{\hat{e}}$, $\|\tilde{\boldsymbol{\Theta}}_x(t_0)\|_{\boldsymbol{\Gamma}_x} \leq \delta_x$, $\|\tilde{\boldsymbol{\Lambda}}_L^T(t_0)\|_{\boldsymbol{\Gamma}_L} \leq \delta_L$
  - $\delta_{\hat{e}} \in [0, \rho_{\hat{e}}]$, $\delta_x \in [0, \rho_x]$, $\delta_L \in [0, \rho_L]$
  - $\rho_{\hat{e}}$, ρ_x, ρ_L defined in (4.314)

such that

- $\|\hat{\mathbf{e}}(t)\|_2 \leq r_{\hat{e}}$ forall $t \geq t_0$ and $\|\hat{\mathbf{e}}(t)\|_2 \leq b_{\hat{e}}$ for all $t \geq t_0 + T_{\hat{e}}(b_{\hat{e}}, \delta_{\hat{e}})$
- $\|\tilde{\boldsymbol{\Theta}}_x(t)\|_{\boldsymbol{\Gamma}_x} \leq \sqrt{\gamma_x \underline{\lambda}_{P_E}} r_{\hat{e}}$, $\|\tilde{\boldsymbol{\Lambda}}_L(t)\|_{\boldsymbol{\Gamma}_L} \leq \sqrt{\gamma_L \underline{\lambda}_{P_E}} r_{\hat{e}}$ for all $t \geq t_0$
- $\|\tilde{\boldsymbol{\Theta}}_x(t)\|_{\boldsymbol{\Gamma}_x} \leq b_x$ for all $t \geq t_0 + T_x(b_x, \delta_x)$, $\|\tilde{\boldsymbol{\Lambda}}_L^T(t)\|_{\boldsymbol{\Gamma}_L} \leq b_L$ for all $t \geq t_0 + T_L(b_L, \delta_L)$

where

- $b_{\hat{e}} = \sqrt{\underline{\lambda}_{P_E}^{-1} \nu_1}$, $b_x = \sqrt{\gamma_x \nu_1}$, $b_L = \sqrt{\gamma_L \nu_1}$, $\nu_1 = \bar{\lambda}_{P_E} \mu_{\hat{e}}^2 + \gamma_x^{-1} \mu_x^2 + \gamma_L^{-1} \mu_L^2$, $\mu_{\hat{e}}$, μ_x, μ_L defined in (4.315)
- $T_{\hat{e}}(b_{\hat{e}}, \delta_{\hat{e}}) = \max\left[0, K_0^{-1}\left(\bar{\lambda}_{P_E} \delta_{\hat{e}}^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_{\hat{e}}^2 - \nu_1\right)\right]$,

  $T_x(b_x, \delta_x) = \max\left[0, K_0^{-1}\left(\delta_x^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_{\hat{e}}^2 - \nu_1\right)\right]$, $T_L(b_L, \delta_L) = \max\left[0, K_0^{-1}\left(\delta_L^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_{\hat{e}}^2 - \nu_1\right)\right]$

Part 2: Nonlinear-in-Control Analysis

In the second part, we will prove boundedness of the prediction-tracking error, subjected to (4.275) and the state of the NIC algorithm. Using results of section 4.5, the NIC algorithm is designed as a gradient system. Therefore recall that we want to minimize

$$\Delta\hat{\mathbf{g}} := \hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N) - \mathbf{w}^* \tag{4.322}$$

by variation of $\mathbf{u}_N$, where

$$\mathbf{w}^* = \left[\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta}) + \mathbf{I}\right]^{-1}\left[\mathbf{v} - \hat{\mathbf{K}}_x^T\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})\right] \tag{4.323}$$

is the desired virtual control, obtained from linearizing state feedback (4.267) and written in terms of parameterization (4.107). According to equation (4.249) in section 4.5, the NIC algorithm is designed as a gradient system of the form

$$\varepsilon\dot{\mathbf{u}}_N = -\mathbf{J}_u^T(\zeta,\boldsymbol{\eta},\mathbf{u}_N)\left[\hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}) - \mathbf{w}^*(\mathbf{v},\zeta,\boldsymbol{\eta})\right] \tag{4.324}$$

where $\varepsilon>0$ is a small parameter, determining the tuning speed of the NIC design and $\mathbf{J}_u(\cdot)$ denotes the map Jacobian w.r.t. $\mathbf{u}_N$. .

It is intuitively clear that we have to require a minimum control authority that is achieved by the nonaffine controls for the NIC algorithm to work properly. Therefore, let $\mathcal{U}$ be the set of achievable nonaffine controls, $\mathbf{u}_N\in\mathcal{U}$. Then, by $\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$, we obtain a set of attainable virtual controls $\hat{\mathbf{w}}$ (equation (4.264)). Moreover, for any fixed $(\boldsymbol{\zeta}_i,\boldsymbol{\eta}_i,\mathbf{d}_i)$, we obtain a different set of virtual controls $\mathcal{V}_i$, attainable for $\mathbf{u}_N\in\mathcal{U}$. Assuming that plant states and external disturbances are restricted to bounded sets, $(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{d})\in\overline{\mathcal{B}_{\bar{\zeta}}}\times\mathcal{D}_\eta\times\mathcal{D}$. Then we obtain a minimum achievable virtual control set, which actually equals the intersection

$$\mathcal{V} = \bigcap_i \mathcal{V}_i$$

for any $(\boldsymbol{\zeta}_i,\boldsymbol{\eta}_i,\mathbf{d}_i)\in\overline{\mathcal{B}_{\bar{\zeta}}}\times\mathcal{D}_\eta\times\mathcal{D}$. In other words, for every $(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{d})\in\overline{\mathcal{B}_{\bar{\zeta}}}\times\mathcal{D}_\eta\times\mathcal{D}$, and every $\hat{\mathbf{w}}\in\mathcal{V}$, there is some $\mathbf{u}_N\in\mathcal{U}$ such that $\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})=\mathbf{w}$. It is further desirable that the origin is contained in $\mathcal{V}$ since we want to be able that the nonaffine controls do not have any effect onto the dynamic and a minimum control authority of the virtual controls means that there is a closed ball $\overline{\mathcal{B}_{\bar{v}}}$ around the origin with a sufficiently small radius $\bar{v}>0$ which is contained in $\mathcal{V}$. Moreover, notice that the NIC algorithm approximates $\mathbf{u}_N$, which implies that there is a deviation between the ideal value $\mathbf{u}_N^*$ and $\mathbf{u}_N$. Therefore, if we have our desired virtual controls restricted to $\overline{\mathcal{B}_{\bar{v}}}$ it is necessary that a ball around $\mathbf{u}_N^*$ with a sufficiently small radius still $\bar{u}$ belongs to $\mathcal{U}$ such that also

$\mathbf{u}_N$ belongs to $\mathcal{U}$ for small deviations. These considerations are summarized in the following.

Assumption E: *Minimum authority of nonaffine controls*

Consider the nonlinear control map (4.264) $\hat{\mathbf{g}}: \overline{\mathcal{B}_{\bar{\zeta}}} \times \mathcal{D}_\eta \times \mathcal{U} \times \mathcal{D} \to \mathbb{R}^m$, where

- $\mathcal{U} \subset \mathbb{R}^m$, $\mathcal{D}_\eta \subset \mathbb{R}^{n-r}$, $\mathcal{D} \subset \mathbb{R}^d$.
- $\overline{\mathcal{B}_{\bar{v}}} \subset \mathbb{R}^m$ (closed ball around the origin with radius $\bar{v} > 0$)
- $\overline{\mathcal{B}_{\bar{\zeta}}} \subset \mathbb{R}^r$ (closed ball around the origin with radius $\bar{\zeta} > 0$),

Then, for every $\hat{\mathbf{w}} \in \overline{\mathcal{B}_{\bar{v}}}$ and $(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{d}) \in \overline{\mathcal{B}_{\bar{\zeta}}} \times \mathcal{D}_\eta \times \mathcal{D}$, there exists $\mathbf{u}_N \in \mathcal{U}$ and a closed ball $\overline{\mathcal{B}_{ru}}(\mathbf{u}_N)$ around $\mathbf{u}_N$ with radius r_u (uniformly for all $\mathbf{u}_N$) such that

- $\hat{\mathbf{g}}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) = \mathbf{w}$
- $\overline{\mathcal{B}_{ru}}(\mathbf{u}_N) \subset \mathcal{U}$

For the NIC algorithm, the inverse of $\hat{\mathbf{g}}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d})$ w.r.t. $\mathbf{u}_N$ in (4.269) has to exists, i.e. a unique $\mathbf{u}_N \in \mathcal{U}$ satisfies (4.264) for every $\mathbf{w} \in \mathcal{V}$ and $(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{d}) \in \overline{\mathcal{B}_{\bar{\zeta}}} \times \mathcal{D}_\eta \times \mathcal{D}$. As explained in the introduction to section 4.5, the gradient system (4.324) together with dynamics of prediction-tracking error (4.275) are in the form a SP model, where (4.324) inherits the role of the fast dynamics. Thikonov's theorem requires exponential stability of the boundary layer dynamics and Theorem 4.5 provides sufficient conditions for existence of a unique inverse and exponential stability of the boundary layer dynamics. Thus, according to Theorem 4.5, the following will be assumed for the nonlinear control map (4.91).

Assumption F: *Sufficient conditions for existence of an inverse of the nonlinear control map and exponential stability of the SP boundary layer system*

Consider the nonlinear control map (4.264). Further, for $(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{d}) \in \overline{\mathcal{B}_{\bar{\zeta}}} \times \mathcal{D}_\eta \times \mathcal{D}$, $\mathbf{u}_N, \mathbf{u}_{N,1}, \mathbf{u}_{N,2} \in \mathcal{U}$, $\mathcal{U}$ convex, let

- $\mathbf{J}_u(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) = \dfrac{\partial \hat{\mathbf{g}}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d})}{\partial \mathbf{u}_N}$
- $\mathbf{J}_{u,s}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) = 1/2\left(\mathbf{J}_u(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}) + \mathbf{J}_u^T(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d})\right)$
- $\mathbf{P}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_{N,1}, \mathbf{u}_{N,2}, \mathbf{d}) = \int_0^1 \mathbf{J}_u(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_{N,1} + s(\mathbf{u}_{N,2} - \mathbf{u}_{N,1}))ds$

The following holds:

1. : $\underline{\sigma}_{J_u} \geq k_0$ for all $\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d}$ and some $k_0 > 0$
2. either
   a. $\underline{\sigma}_P \geq k_1$ or

b. $\mathbf{J}_{u,s}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$ positive or negative definite, such that

$\lambda_i\left[\mathbf{J}_{u,s}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]\geq k_1$ or $\lambda_i\left[\mathbf{J}_{u,s}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right]\leq -k_1$

respectively for all $i=1,\ldots,m$

for all $\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}$ and some $k_1>0$

3. $\bar{\sigma}_P\leq k_2$ for all $\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d}$ and some $k_2\geq k_1$:

Summing up, due to assumptions E and F, we have a minimum authority for the virtual controls, the nonlinear control map is uniquely invertible and the boundary layer system is exponentially stable. Before starting Lyapunov analysis, we transform (4.324) to an error equation, the error in the nonaffine controls is defined as

$$\tilde{\mathbf{u}}_N=\mathbf{u}_N-\mathbf{u}_N^*. \tag{4.325}$$

and we obtain the error dynamics for the nonaffine controls from (4.324).

$$\varepsilon\dot{\tilde{\mathbf{u}}}_N=-\mathbf{J}_u^T(\zeta,\boldsymbol{\eta},\mathbf{u}_N)\left[\mathbf{g}\left(\zeta,\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d}\right)-\mathbf{w}^*\left(\mathbf{v},\zeta,\boldsymbol{\eta}\right)\right]-\varepsilon\dot{\mathbf{u}}_N^* \tag{4.326}$$

A Lyapunov function candidate is given by

$$V_2(\mathbf{e},\tilde{\mathbf{u}}_N)=\mathbf{e}^T\mathbf{P}_R\mathbf{e}+\left[\hat{\mathbf{g}}\left(\zeta,\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d}\right)-\mathbf{w}^*\right]^T\left[\hat{\mathbf{g}}\left(\zeta,\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d}\right)-\mathbf{w}^*\right] \tag{4.327}$$

where $\mathbf{P}_R$ is the symmetric positive definite solution of

$$\mathbf{A}_R^T\mathbf{P}_R+\mathbf{P}_R\mathbf{A}_R=-\mathbf{Q}_R \tag{4.328}$$

for symmetric positive definite $\mathbf{Q}_R$. Note that the second term in (4.327) is indeed positive for $\tilde{\mathbf{u}}_N\neq 0$ if assumption F holds. The time derivative is

$$\dot{V}_2=\dot{\mathbf{e}}^T\mathbf{P}_R\mathbf{e}+\mathbf{e}^T\mathbf{P}_R\dot{\mathbf{e}}+2\left[\hat{\mathbf{g}}\left(\zeta,\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d}\right)-\mathbf{w}^*\right]^T\frac{d}{dt}\left[\hat{\mathbf{g}}\left(\zeta,\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d}\right)-\mathbf{w}^*\right].$$

Before continuing, take a closer look the time derivative of the nonlinear control map. It depends on the external, internal plant states, the nonaffine controls and the external disturbances. If we assume constant disturbances, we obtain

$$\begin{aligned}&\frac{d}{dt}\left[\hat{\mathbf{g}}\left(\zeta,\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d}\right)-\mathbf{w}^*\right]=\\&\mathbf{J}_\zeta\left(\zeta,\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d}\right)\dot{\zeta}+\mathbf{J}_\eta\left(\zeta,\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d}\right)\dot{\boldsymbol{\eta}}+\mathbf{J}_u\left(\zeta,\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d}\right)\dot{\mathbf{u}}_N-\dot{\mathbf{w}}^*\end{aligned} \tag{4.329}$$

where

$$\mathbf{J}_\zeta(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})=\frac{\partial\hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})}{\partial\zeta}\quad,\quad\mathbf{J}_\eta(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})=\frac{\partial\hat{\mathbf{g}}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})}{\partial\boldsymbol{\eta}} \tag{4.330}$$

denote the Jacobians of the nonlinear control map w.r.t. external and internal states. If we insert (4.275) and (4.324), we obtain the time derivative along the system trajectories.

$$\dot{V}_2 = \overbrace{\left\{\mathbf{A}_R\mathbf{e} + \mathbf{H}\left[\mathbf{g}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]\right\}^T}^{\dot{\mathbf{e}}^T}\mathbf{P}_R\mathbf{e} + \mathbf{e}^T\mathbf{P}_R\overbrace{\left\{\mathbf{A}_R\mathbf{e} + \mathbf{H}\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]\right\}}^{\dot{\mathbf{e}}}$$

$$+ 2\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]^T\left[\mathbf{J}_\zeta\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\dot{\boldsymbol{\zeta}} + \mathbf{J}_\eta\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\dot{\boldsymbol{\eta}} - \dot{\mathbf{w}}^*\right]$$

$$+ 2\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]^T\mathbf{J}_u\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\overbrace{\left\{-\frac{1}{\varepsilon}\mathbf{J}_u^T(\mathbf{u}_N)\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]\right\}}^{\dot{\mathbf{u}}_N}$$

$$\dot{V}_2 = \mathbf{e}^T\overbrace{\left(\mathbf{A}_R^T\mathbf{P}_R + \mathbf{P}_R\mathbf{A}_R\right)}^{-\mathbf{Q}_R}\mathbf{e} + 2\left[\hat{\mathbf{g}}\left(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N^* + \tilde{\mathbf{u}}_N, \mathbf{d}\right) - \mathbf{w}^*\right]^T\mathbf{H}^T\mathbf{P}_R\mathbf{e}$$
$$+ 2\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]^T\left[\mathbf{J}_\zeta\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\dot{\boldsymbol{\zeta}} + \mathbf{J}_\eta\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\dot{\boldsymbol{\eta}} - \dot{\mathbf{w}}^*\right]$$
$$- \frac{2}{\varepsilon}\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]^T\mathbf{J}_u\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\mathbf{J}_u^T(\mathbf{u}_N)\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]$$

$$\dot{V}_2 = -\mathbf{e}^T\mathbf{Q}_R\mathbf{e} + 2p\left(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N^* + \tilde{\mathbf{u}}_N, \mathbf{d}, \dot{\boldsymbol{\zeta}}, \dot{\boldsymbol{\eta}}, \dot{\mathbf{w}}^*\right)$$
$$- \frac{2}{\varepsilon}\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]^T\mathbf{J}_u\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\mathbf{J}_u^T(\mathbf{u}_N)\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right] \tag{4.331}$$

Notice that we have dropped the dependencies on $\boldsymbol{\zeta}$, $\boldsymbol{\eta}$, $\mathbf{d}$ for readability and the indefinite part has been abbreviated as

$$p\left(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N^* + \tilde{\mathbf{u}}_N, \mathbf{d}, \dot{\boldsymbol{\zeta}}, \dot{\boldsymbol{\eta}}, \dot{\mathbf{w}}^*\right) = 2\left[\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right]^T\left[\mathbf{H}^T\mathbf{P}_R\mathbf{e} + \mathbf{J}_\zeta\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\dot{\boldsymbol{\zeta}} + \mathbf{J}_\eta\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\dot{\boldsymbol{\eta}} - \dot{\mathbf{w}}^*\right] \tag{4.332}$$

The remaining terms are quadratic and negative definite. Using Theorem B.20 for the first quadratic term, Theorem B.18 and assumption F for the second quadratic term, we obtain an upper bound on (4.331).

$$\dot{V}_2 \leq -\underline{\lambda}_{Q_R}\left\|\mathbf{e}\right\|_2^2 + p\left(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N^* + \tilde{\mathbf{u}}_N, \mathbf{d}, \dot{\boldsymbol{\zeta}}, \dot{\boldsymbol{\eta}}, \dot{\mathbf{w}}^*\right) - \frac{2k_0^2}{\varepsilon}\left\|\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right\|_2^2 \tag{4.333}$$

In a next step, we have to find an upper bound on the indefinite term. At first, by application of Cauchy-Schwartz and triangle inequality, we get

$$p\left(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N^* + \tilde{\mathbf{u}}_N, \mathbf{d}, \dot{\boldsymbol{\zeta}}, \dot{\boldsymbol{\eta}}, \dot{\mathbf{w}}^*\right) \leq$$
$$2\overbrace{\left\|\hat{\mathbf{g}}\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right) - \mathbf{w}^*\right\|_2}^{1}\left(\overbrace{\left\|\mathbf{H}^T\mathbf{P}_R\mathbf{e}\right\|_2}^{2} + \overbrace{\left\|\mathbf{J}_\zeta\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\dot{\boldsymbol{\zeta}}\right\|_2}^{3} + \overbrace{\left\|\mathbf{J}_\eta\left(\mathbf{u}_N^* + \tilde{\mathbf{u}}_N\right)\dot{\boldsymbol{\eta}}\right\|_2}^{4} + \overbrace{\left\|\dot{\mathbf{w}}^*\right\|_2}^{5}\right). \tag{4.334}$$

As computation of an upper bound on (4.334) is quite elaborate, it is shifted to Appendix F.5. However, in order to successfully establish an upper bound on (4.334), we need to impose upper bounds on the nonlinear control map $\hat{\mathbf{g}}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N, \mathbf{d})$ (equation

(4.264)), the scalar control effectiveness $b(\boldsymbol{\zeta},\boldsymbol{\eta})$ (equation (4.184)) and the state dependent regressor $\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})$ (equation (4.107)) as well as on their partial derivatives.

Assumption G: *Upper bound on control map Jacobians of w.r.t. plant states*

Consider the nonlinear control map (4.264). For $(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{d})\in \overline{\mathcal{B}_{\bar{\zeta}}}\times\mathcal{D}_\eta\times\mathcal{D}$, $\mathbf{u}_N\in\mathcal{U}$, $\mathcal{U}$ convex, let

- Jacobi matrix w.r.t. $\boldsymbol{\zeta}$: $\mathbf{J}_\zeta(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})=\dfrac{\partial\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})}{\partial\boldsymbol{\zeta}}$
- Jacobi matrix w.r.t. $\boldsymbol{\eta}$: $\mathbf{J}_\eta(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})=\dfrac{\partial\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})}{\partial\boldsymbol{\eta}}$

The following bounds hold for some $M_\zeta, M_\eta>0$

- $\left\|\mathbf{J}_\zeta(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right\|_2\le M_\zeta$
- $\left\|\mathbf{J}_\eta(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right\|_2\le M_\eta$

Assumption H: *Upper bound on the state dependent regressor*

For $(\boldsymbol{\zeta},\boldsymbol{\eta})\in\overline{\mathcal{B}_{\bar{\zeta}}}\times\mathcal{D}_\eta$, the state dependent regressor $\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})$ and its Jacobians are bounded by some $F,F_\zeta,F_\eta>0$ such that

$$\left\|\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})\right\|_2\le F. \tag{4.335}$$

$$\left\|\frac{\partial\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial\boldsymbol{\zeta}}\right\|_2\le F_\zeta \quad,\quad \left\|\frac{\partial\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right\|_2\le F_\eta. \tag{4.336}$$

Assumption I: *Upper bound on the control effectiveness factor and its Jacobians*

For every $(\boldsymbol{\zeta},\boldsymbol{\eta})\in\overline{\mathcal{B}_{\bar{\zeta}}}\times\mathcal{D}_\eta$, the control effectiveness factor $b(\boldsymbol{\zeta},\boldsymbol{\eta})$ and its Jacobians are bounded by some $\bar{b},\bar{b}_\zeta,\bar{b}_\eta>0$ such that

$$\left|b(\boldsymbol{\zeta},\boldsymbol{\eta})\right|\le\bar{b}. \tag{4.337}$$

$$\left\|\frac{\partial b(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial\boldsymbol{\zeta}}\right\|_2\le\bar{b}_\zeta \quad,\quad \left\|\frac{\partial b(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\right\|_2\le\bar{b}_\eta \tag{4.338}$$

Additionally, we also need to assume bounds on the internal dynamics (4.90).

Assumption J: *Upper bound on* $\dot{\boldsymbol{\eta}}$

Consider the internal dynamics

$$\dot{\boldsymbol{\eta}}(t)=\mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})+\mathbf{P}(\boldsymbol{\zeta},\boldsymbol{\eta})\mathbf{u}.$$

For every $(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{d})\in\overline{\mathcal{B}_{\bar{\zeta}}}\times\mathcal{D}_\eta\times\mathcal{D}$, $\mathbf{u}_N\in\mathcal{U}$ and some $\bar{p},\bar{q}>0$

$$\left\|\mathbf{P}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right\|_2\le\bar{p} \quad,\quad \left\|\mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right\|_2\le\bar{q} \tag{4.339}$$

With results (F.41), (F.42), (F.57), (F.60), (F.93) of Appendix F.5, an upper bound on (4.334) is given by

$$\begin{aligned} &p\left(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d},\dot{\boldsymbol{\zeta}},\dot{\boldsymbol{\eta}},\dot{\mathbf{w}}^*\right)\leq \\ &2k_2\overline{\lambda}_{P_R}\|\mathbf{e}\|_2\|\tilde{\mathbf{u}}_N\|_2 \\ &+2k_2\left(M_\zeta k_2+C_u+C_{\hat{e}u}\|\hat{\mathbf{e}}\|_2\right)\|\tilde{\mathbf{u}}_N\|_2^2 \\ &+2k_2\left[\left(M_\zeta Z_{\hat{e}}+M_\eta Y_{\hat{e}}+C_{\hat{e}}\right)\|\hat{\mathbf{e}}\|_2+C_{\hat{e}^2}\|\hat{\mathbf{e}}\|_2^2+C_{\hat{e}^3}\|\hat{\mathbf{e}}\|_2^3+C_0+M_\zeta Z_0+M_\eta Y_0\right]\|\tilde{\mathbf{u}}_N\|_2. \end{aligned} \tag{4.340}$$

where the constants C_i, Y_i, Z_i, k_4 are defined in equations (F.53), (F.56), (F.59), (F.94), in Appendix F.5. We also have used $\overline{\zeta}$, which defines the valid domain $\overline{\mathcal{B}_{\bar{\zeta}}}$ for the external plant states in assumptions D – I and a bound on the time derivative of the exogenous input $\dot{\mathbf{y}}_C$ in (F.82). Note that the terms with equal power of $\|\mathbf{e}\|_2$ and $\|\tilde{\mathbf{u}}_N\|_2$ have been sorted together while terms that contain $\|\hat{\mathbf{e}}\|_2$ are collected into constants in front of the respective factors, since the current part 2 is concerned with proof of boundedness for $\mathbf{e}$ and $\tilde{\mathbf{u}}_N$ while boundedness of $\hat{\mathbf{e}}$ is proved in part 1 and is assumed a priory for part 2, as explained in the introductory remarks to this section (Figure 4.13). Hence, we assume $\|\hat{\mathbf{e}}\|_2\leq r_{\hat{e}}$, such that inequality (4.287) is satisfied. Moreover, using $r_{\hat{e}}$, we obtain a bound on (4.340).

$$p\left(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N^*+\tilde{\mathbf{u}}_N,\mathbf{d},\dot{\boldsymbol{\zeta}},\dot{\boldsymbol{\eta}},\dot{\mathbf{w}}^*\right)\leq 2k_2\overline{\lambda}_{P_R}\|\mathbf{e}\|_2\|\tilde{\mathbf{u}}_N\|_2+D_{u^2}\|\tilde{\mathbf{u}}_N\|_2^2+D_u\|\tilde{\mathbf{u}}_N\|_2. \tag{4.341}$$

$$D_{u^2}=2k_2\left(M_\zeta k_2+C_u+C_{\hat{e}u}r_{\hat{e}}\right)$$

$$D_u=2k_2\left[\left(M_\zeta Z_{\hat{e}}+M_\eta Y_{\hat{e}}+C_{\hat{e}}\right)r_{\hat{e}}+C_{\hat{e}^2}r_{\hat{e}}^2+C_{\hat{e}^3}r_{\hat{e}}^3+C_0+M_\zeta Z_0+M_\eta Y_0\right] \tag{4.342}$$

Using results (4.341) and (F.41) in the Lyapunov function derivative (4.333) yields

$$\dot{V}_2\leq-\underline{\lambda}_{Q_R}\|\mathbf{e}\|_2^2+2k_2\overline{\lambda}_{P_R}\|\mathbf{e}\|_2\|\tilde{\mathbf{u}}_N\|_2-\left(\frac{2(k_0k_2)^2}{\varepsilon}-D_{u^2}\right)\|\tilde{\mathbf{u}}_N\|_2^2+D_u\|\tilde{\mathbf{u}}_N\|_2. \tag{4.343}$$

As ε can be chosen arbitrarily small, the coefficient in front of $\|\mathbf{u}_N\|_2^2$ is rendered positive by choosing

$$\frac{2(k_0k_2)^2}{\varepsilon}-D_{u^2}=a_1 \tag{4.344}$$

for some $a_1>0$, which implies

$$\varepsilon=\frac{2(k_0k_2)^2}{a_1+D_{u^2}}. \tag{4.345}$$

Then, we obtain for (4.343)

$$\dot{V}_2\leq-\underline{\lambda}_{Q_R}\|\mathbf{e}\|_2^2+2k_2\overline{\lambda}_{P_R}\|\mathbf{e}\|_2\|\tilde{\mathbf{u}}_N\|_2-a_1\|\tilde{\mathbf{u}}_N\|_2^2+D_u\|\tilde{\mathbf{u}}_N\|_2. \tag{4.346}$$

Next, we split up the quadratic term in $||\mathbf{e}||_2$ into two parts, using a parameter $\eta_e>0$.

$$\dot{V}_2 \leq -\frac{\underline{\lambda}_{Q_R}}{1+\eta_e}\|\mathbf{e}\|_2^2 - \frac{\underline{\lambda}_{Q_R}\eta_e}{1+\eta_e}\|\mathbf{e}\|_2^2 + 2k_2\overline{\lambda}_{P_R}\|\mathbf{e}\|_2\|\tilde{\mathbf{u}}_N\|_2 - a_1\|\tilde{\mathbf{u}}_N\|_2^2 + D_u\|\tilde{\mathbf{u}}_N\|_2$$

If we complete squares, we get rid of the mixed term.

$$\dot{V}_2 \leq -\frac{\underline{\lambda}_{Q_R}}{1+\eta_e}\|\mathbf{e}\|_2^2 - a_1\|\tilde{\mathbf{u}}_N\|_2^2 + D_u\|\tilde{\mathbf{u}}_N\|_2$$
$$-\frac{\underline{\lambda}_{Q_R}\eta_e}{1+\eta_e}\left(\|\mathbf{e}\|_2^2 - 2\frac{1+\eta_e}{\underline{\lambda}_{Q_R}\eta_e}k_2\overline{\lambda}_{P_R}\|\mathbf{e}\|_2\|\tilde{\mathbf{u}}_N\|_2 + \left(\frac{1+\eta_e}{\underline{\lambda}_{Q_R}\eta_e}k_2\overline{\lambda}_{P_R}\right)^2\|\tilde{\mathbf{u}}_N\|_2^2 - \left(\frac{1+\eta_e}{\underline{\lambda}_{Q_R}\eta_e}k_2\overline{\lambda}_{P_R}\right)^2\|\tilde{\mathbf{u}}_N\|_2^2\right)$$

$$\dot{V}_2 \leq -\underline{\lambda}_{Q_R}\frac{1}{1+\eta_e}\|\mathbf{e}\|_2^2 - a_1\|\tilde{\mathbf{u}}_N\|_2^2 + D_u\|\tilde{\mathbf{u}}_N\|_2 \overbrace{-\frac{\underline{\lambda}_{Q_R}\eta_e}{1+\eta_e}\left(\|\mathbf{e}\|_2 - \frac{1+\eta_e}{\underline{\lambda}_{Q_R}\eta_e}k_2\overline{\lambda}_{P_R}\|\tilde{\mathbf{u}}_N\|_2\right)^2}^{\leq 0} + \frac{1+\eta_e}{\underline{\lambda}_{Q_R}\eta_e}\left(k_2\overline{\lambda}_{P_R}\right)^2\|\tilde{\mathbf{u}}_N\|_2^2$$

$$\dot{V}_2 \leq -\underline{\lambda}_{Q_R}\frac{1}{1+\eta_e}\|\mathbf{e}\|_2^2 - \left(a_1 - \frac{1+\eta_e}{\underline{\lambda}_{Q_R}\eta_e}\left(k_2\overline{\lambda}_{P_R}\right)^2\right)\|\tilde{\mathbf{u}}_N\|_2^2 + D_u\|\tilde{\mathbf{u}}_N\|_2$$

Further, choosing a_1 such that

$$a_1 - \frac{1+\eta_e}{\underline{\lambda}_{Q_R}\eta_e}\left(k_2\overline{\lambda}_{P_R}\right)^2 = a_2\,, \tag{4.347}$$

for some $a_2>0$, we obtain

$$\dot{V}_2 \leq -\frac{\underline{\lambda}_{Q_R}}{1+b_e}\|\mathbf{e}\|_2^2 - a_2\|\tilde{\mathbf{u}}_N\|_2^2 + D_u\|\tilde{\mathbf{u}}_N\|_2\,. \tag{4.348}$$

Analogously, we split up the quadratic term in $\|\tilde{\mathbf{u}}_N\|$ into two parts, using some $\eta_u>0$.

$$\dot{V}_2 \leq -\frac{\underline{\lambda}_{Q_R}}{1+\eta_e}\|\mathbf{e}\|_2^2 - \frac{a_2\eta_u}{1+b_u}\|\tilde{\mathbf{u}}_N\|_2^2 - \frac{a_2}{1+\eta_u}\|\tilde{\mathbf{u}}_N\|_2^2 + D_u\|\tilde{\mathbf{u}}_N\|_2$$

Completing squares twice, we also get rid of the linear term.

$$\dot{V}_2 \leq -\frac{\underline{\lambda}_{Q_R}}{1+\eta_e}\|\mathbf{e}\|_2^2 - \frac{a_2\eta_u}{1+\eta_u}\|\tilde{\mathbf{u}}_N\|_2^2$$
$$-\frac{a_2}{1+\eta_u}\left(\|\tilde{\mathbf{u}}_N\|_2^2 - \frac{1+\eta_u}{a_2}D_u\|\tilde{\mathbf{u}}_N\|_2 + \left(\frac{1+\eta_u}{a_2}\frac{D_u}{2}\right)^2 - \left(\frac{1+\eta_u}{a_2}\frac{D_u}{2}\right)^2\right)$$

$$\dot{V}_2 \le -\frac{\underline{\lambda}_{Q_R}}{1+\eta_e}\|\mathbf{e}\|_2^2 - \frac{a_2\eta_u}{1+\eta_u}\|\tilde{\mathbf{u}}_N\|_2^2 \overbrace{- \frac{a_2}{1+\eta_u}\left(\|\tilde{\mathbf{u}}_N\|_2^2 - \frac{1+\eta_u}{a_2}\frac{D_u}{2}\right)^2}^{\le 0} + \frac{1+\eta_u}{a_2}\left(\frac{D_u}{2}\right)^2$$

$$\dot{V}_2 \le -\frac{\underline{\lambda}_{Q_R}}{1+\eta_e}\|\mathbf{e}\|_2^2 - \frac{a_2\eta_u}{1+\eta_u}\|\tilde{\mathbf{u}}_N\|_2^2 + \frac{1+\eta_u}{a_2}\left(\frac{D_u}{2}\right)^2 \tag{4.349}$$

By assumption E, the nonaffine controls $\mathbf{u}_N$ belong to $\mathcal{U}$, if the error of the NIC algorithm satisfies

$$\|\tilde{\mathbf{u}}_N\|_2 \le r_u \tag{4.350}$$

and the desired virtual controls satisfy

$$\|\mathbf{w}^*\|_2 \le \bar{v}\,. \tag{4.351}$$

Equation (F.96) in F.5 provides an upper bound on $\mathbf{w}^*$, which depends on the prediction error $\hat{\mathbf{e}}$. However, if we assume that $\|\hat{\mathbf{e}}\|_2 \le r_{\hat{e}}$ holds (Figure 4.13), we obtain

$$\|\mathbf{w}^*\|_2 \le w^* := \left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2\right) r_{\hat{e}} + \|\mathbf{K}\|_2 \bar{\zeta} + \|\mathbf{A}_0\|_2 \bar{y} + \left(\theta_{x,\max}\sqrt{\bar{\lambda}_{\Gamma_x}} + \|\hat{\mathbf{K}}_x\|_2\right) F\,. \tag{4.352}$$

From (4.349) two functions can be constructed which depend on only one of the states $\mathbf{e}$, $\tilde{\mathbf{u}}_N$ respectively and whose negative values establish an upper bound on (4.349).

$$\begin{aligned}\gamma_e\left(\|\mathbf{e}\|_2\right) &= \frac{\underline{\lambda}_{Q_R}}{1+\eta_e}\|\mathbf{e}\|_2^2 - \frac{1+\eta_u}{a_2}\left(\frac{D_u}{2}\right)^2 \\ \gamma_u\left(\|\tilde{\mathbf{u}}_N\|_2\right) &= \frac{a_2\eta_u}{1+\eta_u}\|\tilde{\mathbf{u}}_N\|_2^2 - \frac{1+\eta_u}{a_2}\left(\frac{D_u}{2}\right)^2\end{aligned} \tag{4.353}$$

Hence, we have

$$\dot{V}_2 \le -\max\left[\gamma_e\left(\|\mathbf{e}\|_2, \|\tilde{\mathbf{u}}_N\|_2\right)\right]. \tag{4.354}$$

According to notation of Corollary 3.1, the class $\mathcal{K}$ functions that bound the Lyapunov function candidate (4.327) from below and above, are

$$\begin{aligned}\alpha_e\left(\|\mathbf{e}\|_2\right) &= \underline{\lambda}_{P_R}\|\mathbf{e}\|_2^2 \qquad \beta_e\left(\|\mathbf{e}\|_2\right) = \bar{\lambda}_{P_R}\|\mathbf{e}\|_2^2 \\ \alpha_u\left(\|\tilde{\mathbf{u}}_N\|_2\right) &= k_1\|\tilde{\mathbf{u}}_N\|_2^2 \qquad \beta_u\left(\|\tilde{\mathbf{u}}_N\|_2\right) = k_2\|\tilde{\mathbf{u}}_N\|_2^2\end{aligned} \tag{4.355}$$

where α_u, β_u are obtained from assumption F and Lemma B.4. Then, $r_e>0$ is chosen such that (4.287) is satisfied and, according to Corollary 3.1,

$$\underline{u} = \min\left[\alpha_e(r_e), \alpha_u(r_u)\right] \tag{4.356}$$

and the radii, determining the sets of admissible initial conditions are

$$\rho_e = \sqrt{\frac{\underline{u}}{2\overline{\lambda}_{P_R}}} \quad , \quad \rho_u = \sqrt{\frac{\underline{u}}{2k_2}} \, . \tag{4.357}$$

Then, we need to find μ_e, $\mu_u>0$, such that γ_e, γ_u are class $\mathcal{K}$ functions in the set $\overline{\mathcal{B}_r} \setminus \overline{\mathcal{B}_\mu}$ (according to Corollary 3.1). These are obtained by setting

$$\gamma_e(\mu_e) = K_1 \quad , \quad \gamma_u(\mu_u) = K_1$$

for some small $K_1>0$ and solving for μ_e, μ_u.

$$\mu_e = \sqrt{\frac{1+\eta_e}{\underline{\lambda}_{Q_R}}\left(\frac{1+\eta_u}{a_2}\left(\frac{D_u}{2}\right)^2 + K_1\right)} \quad , \quad \mu_u = \sqrt{\frac{1+\eta_u}{a_2\eta_u}\left(\frac{(1+\eta_u)}{a_2}\left(\frac{D_u}{2}\right)^2 + K_1\right)} \tag{4.358}$$

Sufficient conditions for ultimate boundedness, according to Corollary 3.1, are $\mu_e < \rho_e$ and $\mu_u < \rho_u$, which results in

$$\sqrt{\frac{1+\eta_e}{\underline{\lambda}_{Q_R}}\left(\frac{1+\eta_u}{a_2}\left(\frac{D_u}{2}\right)^2 + K_1\right)} < \sqrt{\frac{\underline{u}}{2\overline{\lambda}_{P_R}}} \quad , \quad \sqrt{\frac{1+\eta_u}{a_2\eta_u}\left(\frac{(1+\eta_u)}{a_2}\left(\frac{D_u}{2}\right)^2 + K_1\right)} < \sqrt{\frac{\underline{u}}{2k_2}} \, . \tag{4.359}$$

Now, we want to fulfill these conditions by choosing a_2 sufficiently large. However, necessary for existence for such a_2, in view of the first condition, is that

$$K_1 < \frac{\underline{u}\underline{\lambda}_{Q_R}}{2\overline{\lambda}_{P_R}(1+\eta_e)} =: \overline{K}_1 \, . \tag{4.360}$$

By choosing ε sufficiently small (according to equations (4.344), (4.345), (4.347)), we obtain some a_2 which fulfills conditions (4.359). The infimum for a_2 is obtained by

$$a_2 > a_2^{(0)}(\eta_e, \eta_u, K_1) := \max\left[a_2^{(e)}(\eta_e, \eta_u, K_1),\, a_2^{(u)}(\eta_u, K_1)\right] \tag{4.361}$$

where

$$\begin{aligned} a_2^{(e)}(\eta_e, \eta_u, K_1) &= \left(\overline{K}_1 - K_1\right)^{-1}\left(\frac{D_u}{2}\right)^2(1+\eta_u) \\ a_2^{(u)}(\eta_u, K_1) &= \frac{1+\eta_u}{\eta_u}\frac{k_2 K_1}{\underline{u}}\left(1+\sqrt{1+2\frac{\underline{u}\eta_u}{k_2 K_1^2}\left(\frac{D_u}{2}\right)^2}\right). \end{aligned} \tag{4.362}$$

Notice that $a_2^{(0)}$ depends on the parameters η_e, η_u, K_1 and hence there is room for optimization. Preferably, we want find the supremum of all ε that stabilize the system. Therefore, insert (4.361) and (4.347) into (4.345)

$$\varepsilon < \varepsilon^{(0)}(\eta_u, \eta_e, K_1) = \frac{2(k_0 k_2)^2}{a_2^{(0)}(\eta_e, \eta_u, K_1) + \frac{1+\eta_e}{\underline{\lambda}_{Q_R}\eta_e}\left(k_2\overline{\lambda}_{P_R}\right)^2 + D_{u^2}} \tag{4.363}$$

and for each $\eta_e,\ \eta_u>0$ and $K_1>0$ that satisfies (4.360), we obtain some $\varepsilon^{(0)}$ such that the system states are ultimately bounded for all $\varepsilon<\varepsilon^{(0)}(\eta_e,\eta_u,K_1)$

Moreover, by numerical optimization, (4.364) can be maximized. Let ε^* denote that maximal $\varepsilon^{(0)}$ and η_e^*,η_u^*,K_1^* the set of optimal parameters, then the system states are ultimately bounded for all

$$\varepsilon<\varepsilon^*=\varepsilon^{(0)}\left(\eta_e^*,\eta_e^*,K_1^*\right). \tag{4.364}$$

Thereby, the optimization problem is summarized as follows:

- maximize (4.363) by variation of η_e,η_u,K_1
- subjected to $\eta_e,\ \eta_u>0,\ 0<K_1<\overline{K}_1$

The results for part 2 are summarized in the following.

Lemma 4.3 MRAC Variant 2 – Part 2

Consider:

- stable **reference dynamics** (4.56) and **state predictor** (4.272)
- **nonaffine control map** $\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$ in (4.264)
- **parameterizations** (4.107) – (4.110), (4.124) for state dependent uncertainty, (4.111) or (4.114) for affine control effectiveness together with (4.184)
- **desired virtual control** $\mathbf{w}^*$ in (4.323) with pseudo control (4.277) consisting of adaptive term (4.122), error feedback (4.278) and predictor based pseudo control (4.273) and the **associated nonaffine control** $\mathbf{u}_N^*$ according to (4.268).
- **update law** (4.294) using
  - symm. pos. def. solution $\mathbf{P}_E$ of (4.292) for some symm. pos. def. $\mathbf{Q}_E$
  - positive learning rates $\gamma_x,\ \gamma_L$
  - projection operator with parameters $\varepsilon_x,\ \varepsilon_L>0,\ \theta_{x,max},\ \theta_{L,max}>0$ and symmetric positive definite weighting matrices $\boldsymbol{\Gamma}_x,\ \boldsymbol{\Gamma}_L$
  - switching □-modification with modification gains $\sigma_x,\ \sigma_L>0$ and MMQ modification with modification gain $\kappa>0$ and N filters according to (4.296)
- symm. pos. def. solution $\mathbf{P}_R$ of (4.328) for some symm. pos. def. $\mathbf{Q}_R$
- **nonlinear-in-control update** (4.324) using **time scale separating parameter** ε
- Sets: $\overline{\mathcal{B}}_{\bar{\zeta}}\in\mathbb{R}^r,\ \mathcal{D}_\eta\in\mathbb{R}^{n-r},\ \mathcal{D}\in\mathbb{R}^d$ for some $\bar{\zeta}>0$

The **prediction-tracking error** $\mathbf{e}$ in (4.275) and **nonaffine control tuning error** $\tilde{\mathbf{u}}_N$ in (4.325) are **uniformly ultimately bounded**, if

- assumption C on the $\mathcal{L}_1$ bound of the MMQ filters is fulfilled for some $L>0$
- assumption D is fulfilled for some $\theta_x,\ \theta_L,\ D>0$ and $\theta_{x,max},\ \theta_{L,max}$ are defined by (4.133), (4.135)
- $(\boldsymbol{\eta},\mathbf{d})\in\mathcal{D}_\eta\times\mathcal{D}$ for all $t\geq t_0$
- assumption E is fulfilled for some minimum nonaffine control effectiveness for $\bar{v}>0$, $r_u>0$
- assumption F on the Jacobian of (4.264) w.r.t. to $\mathbf{u}_N$ is fulfilled for some $k_0,\ k_1>0$ and $k_2\geq k_1$
- assumption G on the Jacobians of (4.91) w.r.t. to external and internal plant states is fulfilled for some $M_\zeta,\ M_\eta>0$
- assumption H on $\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})$ is fulfilled for some $F,\ F_\zeta,\ F_\eta>0$
- assumption I on $b(\boldsymbol{\zeta},\boldsymbol{\eta})$ is fulfilled for some $b,\ b_\zeta,\ b_\eta>0$
- assumption J for bounds on internal dynamics is fulfilled for some $\bar{p},\ \bar{q}>0$
- the reference model state is bounded according to (4.284) for some $\bar{\zeta}_R>0$
- the exogenous input and its time derivative are bounded according to (4.285) and (F.82) by some $\bar{y},\ \bar{y}_d>0$

- $r_e>0$ fulfills (4.287): $r_{\hat{e}} + r_e + \overline{\zeta}_R \le \overline{\zeta}$ – for some $r_{\hat{e}}>0$ such that $\|\hat{\mathbf{e}}\|_2 \le r_{\hat{e}}$ holds for the prediction error (4.276)
- $w^* < \overline{v}$ according to (4.352) and (4.351)
- ε in (4.324) satisfies (4.363), η_e, $\eta_u>0$ $0<K_1<\overline{K}_1$ according to (4.360)
- initial conditions: $\|\mathbf{e}(t_0)\|_2 \le \delta_e$, $\|\tilde{\mathbf{u}}_N(t_0)\|_2 \le \delta_u$

  where $\delta_e \in [0, \rho_e]$, $\delta_u \in [0, \rho_u]$, ρ_e, ρ_u defined in (4.357)

such that

- $\|\mathbf{e}(t)\|_2 \le r_e$ forall $t \ge t_0$ and $\|\mathbf{e}(t)\|_2 \le b_e$ forall $t \ge t_0 + T_e(b_e, \delta_e)$
- $\|\tilde{\mathbf{u}}_N(t)\|_2 \le r_u$ forall $t \ge t_0$ and $\|\tilde{\mathbf{u}}_N(t)\|_2 \le b_u$ forall $t \ge t_0 + T_u(b_u, \delta_u)$

where

- $b_e = \sqrt{\underline{\lambda}_{P_R}^{-1} v_2}$, $b_u = \sqrt{k_1^{-1} v_2}$, $v_2 = \overline{\lambda}_{P_R}\mu_e^2 + k_2\mu_u^2$, μ_e, μ_u defined in (4.358)
- $T_e(b_e, \delta_e) = \max\left[0, (K_1)^{-1}\left(\overline{\lambda}_{P_R}\delta_e^2 + 1/2 \cdot \underline{u} - v_2\right)\right]$, $T_u(b_u, \delta_u) = \max\left[0, (K_1)^{-1}\left(k_2\delta_u^2 + 1/2 \cdot \underline{u} - v_2\right)\right]$

4.6.3 Conclusions

Lemma 4.2 assures boundedness of the prediction error $\|\hat{\mathbf{e}}\|_2 \le r_{\hat{e}}$ where $r_{\hat{e}}$ has a lower bound $\underline{\hat{e}}$, defined by (4.320). Additionally $r_{\hat{e}}$ has to fulfill (4.287) which imposes an upper bound, which has to be large enough such that the lower bound is fulfilled simultaneously. This requires, besides a sufficiently large $\overline{\zeta}$ and sufficiently small $\overline{\zeta}_R$, a sufficiently small r_e. Fortunately, Lemma 4.3 does not require a lower bound on r_e but can be made arbitrarily small by a sufficiently small ε. Hence, if $\underline{\hat{e}}$ in (4.320) satisfies

$$\underline{\hat{e}} + \overline{\zeta}_R < \overline{\zeta} \tag{4.365}$$

then, according to (4.320), we choose $r_{\hat{e}} = \underline{\hat{e}}$ and there is always a $r_e>0$ that fulfills (4.287).

If the bound w^* on the desired virtual control satisfies (4.351), according to assumption E, then $\mathbf{u}_N$ belongs to $\mathcal{U}$, if $\|\tilde{\mathbf{u}}_N\|_2 \le r_u$, which is satisfied for a sufficiently small ε. Summarizing the results so far, the crucial conditions of the complete adaptive system are that

- $\underline{\hat{e}}$ of equation (4.320) satisfies (4.365) and
- w^* of equation (4.352) satisfies (4.351).

If this is the case, there are r_e, $r_u>0$ that guarantee ultimate boundedness of all system states. The result is summarized in the following.

Theorem 4.6 MRAC Variant 2

Consider:

- stable **reference dynamics** (4.56) and **state predictor** (4.272)
- **nonaffine control map** $\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})$ in (4.264)
- **parameterizations** (4.107) – (4.110) for state dependent uncertainty, (4.111) or (4.114) for affine control effectiveness together with (4.184)
- **plant dynamics** in Byrnes-Isidori normal form (4.266), subjected to linearizing state feedback (4.267) with pseudo control (4.277), adaptive term (4.122), error feedback (4.278) and predictor based pseudo control (4.273).
- **parameter-estimation-errors** $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Lambda}}_L$, defined in (4.124), (4.112)
- **update law** (4.294) using
  - symm. pos. def. solution $\mathbf{P}_E$ of (4.292) for some symm. pos. def. $\mathbf{Q}_E$
  - positive learning rates γ_x, γ_L
  - projection operator with parameters $\varepsilon_x,\varepsilon_L>0$, $\theta_{x,max}$, $\theta_{L,max}>0$ and symmetric positive definite weighting matrices $\boldsymbol{\Gamma}_x$, $\boldsymbol{\Gamma}_L$
  - switching σ-modification with modification gains $\sigma_x,\sigma_L>0$ and MMQ modification with modification gain $\kappa>0$ and N filters according to (4.296)
- **nonlinear-in-control update** (4.324) using **time scale separating parameter** ε
- solution $\mathbf{P}_R$ of (4.328) for some positive definite symmetric $\mathbf{Q}_R$,
- Sets: $\overline{\mathcal{B}_{\bar{\zeta}}} \in \mathbb{R}^r$, $\mathcal{D}_\eta \in \mathbb{R}^{n-r}$, $\mathcal{D} \in \mathbb{R}^d$ for some $\bar{\zeta} > 0$

The **prediction-tracking error** $\mathbf{e}$ in (4.275), **prediction error** $\hat{\mathbf{e}}$ in (4.276), **parameter-estimation-errors** $\tilde{\boldsymbol{\Theta}}_x$, $\tilde{\boldsymbol{\Lambda}}_L$ and **NIC tuning error** $\tilde{\mathbf{u}}_N$ in (4.325) are **uniformly ultimately bounded**, if

- assumption C on the $\mathcal{L}_1$ bound of the MMQ filters is fulfilled for some $L>0$
- assumption D is fulfilled for some θ_x, θ_L, $D>0$ and $\theta_{x,max}$, $\theta_{L,max}$ are defined by (4.133), (4.135)
- $(\boldsymbol{\eta},\mathbf{d}) \in \mathcal{D}_\eta \times \mathcal{D}$ for all $t \geq t_0$
- assumption E is fulfilled for some minimum nonaffine control effectiveness for $\bar{\nu} > 0$, $r_u>0$
- assumption F on the Jacobian of (4.264) w.r.t. to $\mathbf{u}_N$ is fulfilled for some k_0, $k_1>0$ and $k_2 \geq k_1$
- assumption G on the Jacobians of (4.91) w.r.t. to external and internal plant states is fulfilled for some $M_\zeta, M_\eta>0$
- assumption H on $\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})$ is fulfilled for some F, F_ζ, $F_\eta>0$
- assumption I on $b(\boldsymbol{\zeta},\boldsymbol{\eta})$ is fulfilled for some b, b_ζ, $b_\eta>0$
- assumption J for bounds on internal dynamics is fulfilled for some $\bar{p}, \bar{q} > 0$
- the reference model state is bounded according to (4.284) for some $\bar{\zeta}_R > 0$
- the exogenous input and its time derivative are bounded according to (4.285) and (F.82) by some $\bar{y}, \bar{y}_d > 0$
- $r_{\hat{e}}, r_e > 0$ satisfy $r_{\hat{e}} + r_e + \bar{\zeta}_R \leq \bar{\zeta}$ according to (4.287) and $r_{\hat{e}}$ additionally satisfies $r_{\hat{e}} \geq \underline{\hat{e}}$ according to (4.320) – where $\underline{\hat{e}}$ is defined in (4.321) and $k_{\hat{e}}^*$, k_x^*, k_L^* such that inequalities (4.318) are solved for σ_x,σ_L,κ, $K_0>0$
- the learning rates are chosen such that $\gamma_x \geq \underline{\gamma}_x, \gamma_L \geq \underline{\gamma}_L$ according to (4.320) where $\underline{\gamma}_x, \underline{\gamma}_L$ are defined is (4.321)
- $w^* < \bar{\nu}$ according to (4.352) and (4.351)
- ε in (4.324) satisfies (4.363), η_e, $\eta_u>0$, $0<K_1<\overline{K}_1$ according to (4.360)
- initial conditions: $\|\mathbf{e}(t_0)\|_2 \leq \delta_e$, $\|\hat{\mathbf{e}}(t_0)\|_2 \leq \delta_{\hat{e}}$, $\|\tilde{\boldsymbol{\Theta}}_x(t_0)\|_{\boldsymbol{\Gamma}_x} \leq \delta_x$, $\|\tilde{\boldsymbol{\Lambda}}_L(t_0)\|_{\boldsymbol{\Gamma}_L} \leq \delta_L$ $\|\tilde{\mathbf{u}}_N(t_0)\|_2 \leq \delta_u$
  - $\delta_e \in [0,\rho_e]$, $\delta_{\hat{e}} \in [0,\rho_{\hat{e}}]$, $\delta_x \in [0,\rho_x]$, $\delta_L \in [0,\rho_L]$, $\delta_u \in [0,\rho_u]$
    ρ_e, $\rho_{\hat{e}}$, ρ_x, ρ_L, ρ_u defined in (4.314), (4.357)

such that

- $\|\mathbf{e}(t)\|_2 \leq r_e$ for all $t \geq t_0$ and $\|\mathbf{e}(t)\|_2 \leq b_e$ for all $t \geq t_0 + T_e(b_e, \delta_e)$
- $\|\hat{\mathbf{e}}(t)\|_2 \leq r_{\hat{e}}$ for all $t \geq t_0$ and $\|\hat{\mathbf{e}}(t)\|_2 \leq b_{\hat{e}}$ for all $t \geq t_0 + T_{\hat{e}}(b_{\hat{e}}, \delta_{\hat{e}})$
- $\left\|\tilde{\mathbf{\Theta}}_x(t)\right\|_{\mathbf{\Gamma}_x}$, $\left\|\tilde{\mathbf{\Lambda}}_L(t)\right\|_{\mathbf{\Gamma}_L} \leq \sqrt{\bar{\lambda}_{P_E}} r_{\hat{e}}$ for all $t \geq t_0$
- $\left\|\tilde{\mathbf{\Theta}}_x(t)\right\|_{\mathbf{\Gamma}_x} \leq b_x$ for all $t \geq t_0 + T_x(b_x, \delta_x)$, $\left\|\tilde{\mathbf{\Lambda}}_L(t)\right\|_{\mathbf{\Gamma}_L} \leq b_L$ for all $t \geq t_0 + T_L(b_L, \delta_L)$
- $\|\tilde{\mathbf{u}}_N(t)\|_2 \leq r_u$ for all $t \geq t_0$ and $\|\tilde{\mathbf{u}}_N(t)\|_2 \leq b_u$ for all $t \geq t_0 + T_u(b_u, \delta_u)$

where

- $b_e = \sqrt{\underline{\lambda}_{P_R}^{-1} \nu_2}$, $b_{\hat{e}} = \sqrt{\underline{\lambda}_{P_E}^{-1} \nu_1}$, $b_x = \sqrt{\gamma_x \nu_1}$, $b_L = \sqrt{\gamma_L \nu_1}$, $b_u = \sqrt{k_1^{-1} \nu_2}$
- $\nu_1 = \bar{\lambda}_{P_E} \mu_{\hat{e}}^2 + \gamma_x^{-1} \mu_x^2 + \gamma_L^{-1} \mu_L^2$, $\nu_2 = \bar{\lambda}_{P_R} \mu_e^2 + k_2 \mu_u^2$
- μ_e , $\mu_{\hat{e}}$, μ_x , μ_L , μ_u , defined in (4.315), (4.358)
- $T_e(b_e, \delta_e) = \max\left[0, (K_1)^{-1}\left(\bar{\lambda}_{P_R} \delta_e^2 + 1/2 \cdot \underline{u} - \nu_2\right)\right]$, $T_{\hat{e}}(b_{\hat{e}}, \delta_{\hat{e}}) = \max\left[0, K_0^{-1}\left(\bar{\lambda}_{P_E} \delta_{\hat{e}}^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_{\hat{e}}^2 - \nu_1\right)\right]$

  $T_x(b_x, \delta_x) = \max\left[0, K_0^{-1}\left(\delta_x^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_{\hat{e}}^2 - \nu_1\right)\right]$, $T_L(b_L, \delta_L) = \max\left[0, K_0^{-1}\left(\delta_L^2 + 3/4 \cdot \underline{\lambda}_{P_E} r_{\hat{e}}^2 - \nu_1\right)\right]$

  $T_u(b_u, \delta_u) = \max\left[0, (K_1)^{-1}\left(k_2 \delta_u^2 + 1/2 \cdot \underline{u} - \nu_2\right)\right]$

4.7 Singular Value Update Algorithm

The absolute scaling for the decoupling matrix in equation (4.111) requires a restriction of the adaptive $\hat{\mathbf{\Lambda}}_L \in \mathbb{R}^{m \times m}$ such that the pseudoinverse of the estimated decoupling matrix $\hat{\mathbf{B}}(\zeta, \mathbf{\eta})$ exists. In adaptive control, this is commonly assured by restricting $\hat{\mathbf{\Lambda}}_L$ to diagonal matrices with positive diagonal entries, although this implies a considerable limitation of the adaptation capabilities as explained in section 4.2.2. Within the MRAC scheme, adaptation of $\hat{\mathbf{\Lambda}}_L$ is accomplished via a differential equation of the form

$$\dot{\hat{\mathbf{\Lambda}}}_L = \mathbf{F}(t) \tag{4.366}$$

where $\mathbf{F}(t)$ is obtained from the update law, e.g. (4.131). The diagonal constraint can be easily achieved by an appropriate choice of initial conditions, setting the off-diagonal elements of $\mathbf{F}(t)$ to zero and limiting the diagonal entries of $\hat{\mathbf{\Lambda}}_L(t)$ to positive values. However, if we drop the diagonal constraint, there no such straightforward way to prevent $\hat{\mathbf{\Lambda}}_L$ from becoming singular, but requires for consideration of its singular value decomposition.

4.7.1 Algorithm Description

The algorithm, introduced in the following, transforms the update (4.366) to an update in terms of its singular value decomposition (SVD), which offers the possibility to limit the singular values away from zero, avoiding singularity. Therefore let

$$\hat{\mathbf{\Lambda}}_L = \mathbf{V} \mathbf{\Sigma} \mathbf{U}^T \tag{4.366}$$

be the SVD, where $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{m \times m}$ are orthonormal matrices, i.e. its columns are pairwise orthogonal and of Euclidean unity length, which is equivalent to the condition

$$\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I} \quad , \quad \mathbf{U}^T\mathbf{U} = \mathbf{U}\mathbf{U}^T = \mathbf{I} \, , \tag{4.367}$$

and

$$\mathbf{\Sigma} = \mathrm{diag}\begin{bmatrix} \overline{\sigma} & \cdots & \underline{\sigma} \end{bmatrix}$$

is a diagonal matrix containing the singular values. The orthonormal columns of $\mathbf{V}$, $\mathbf{U}$ are denoted as $\mathbf{v}_i$ and $\mathbf{u}_i$ respectively. Of course, all terms depend on time, but the dependent variable is dropped for readability. Using product rule, the time derivative of (4.366) is

$$\dot{\hat{\mathbf{\Lambda}}}_L = \dot{\mathbf{V}}\mathbf{\Sigma}\mathbf{U}^T + \mathbf{V}\dot{\mathbf{\Sigma}}\mathbf{U}^T + \mathbf{V}\mathbf{\Sigma}\dot{\mathbf{U}}^T$$

and with (4.366) we obtain

$$\dot{\mathbf{\Sigma}} = \mathbf{Y} - \mathbf{V}^T\dot{\mathbf{V}}\mathbf{\Sigma} - \mathbf{\Sigma}\dot{\mathbf{U}}^T\mathbf{U} \tag{4.368}$$

where we have defined

$$\mathbf{Y} = \mathbf{V}^T\mathbf{F}\mathbf{U} \, . \tag{4.369}$$

Since the columns of both, $\mathbf{V}$ and $\mathbf{U}$ are an orthogonal basis of the whole $\mathbb{R}^m$ respectively, the time derivatives $\dot{\mathbf{V}}$ and $\dot{\mathbf{U}}$ can be parameterized in terms of the former

$$\dot{\mathbf{V}} = \mathbf{V}\mathbf{C}_V \quad , \quad \dot{\mathbf{U}} = \mathbf{U}\mathbf{C}_u \tag{4.370}$$

for some $\mathbf{C}_V, \mathbf{C}_u \in \mathbb{R}^{m \times m}$ whose values will be determined in the following. The time derivatives of $\mathbf{V}$ and $\mathbf{U}$ have to preserve orthogonality constraint (4.367). Therefore

$$\begin{aligned} \mathbf{V}^T(t)\mathbf{V}(t) = \mathbf{I} = const \quad &\Leftrightarrow \quad \frac{d}{dt}\left(\mathbf{V}^T\mathbf{V}\right) = \dot{\mathbf{V}}^T\mathbf{V} + \mathbf{V}^T\dot{\mathbf{V}} = \mathbf{0} \\ \mathbf{U}^T(t)\mathbf{U}(t) = \mathbf{I} = const \quad &\Leftrightarrow \quad \frac{d}{dt}\left(\mathbf{U}^T\mathbf{U}\right) = \dot{\mathbf{U}}^T\mathbf{U} + \mathbf{U}^T\dot{\mathbf{U}} = \mathbf{0} \end{aligned} \tag{4.371}$$

and with (4.367), (4.370), we obtain

$$\begin{aligned} \mathbf{C}_V^T\mathbf{V}^T\mathbf{V} + \mathbf{V}^T\mathbf{V}\mathbf{C}_V = 0 \qquad & \mathbf{C}_U^T\mathbf{U}^T\mathbf{U} + \mathbf{U}^T\mathbf{U}\mathbf{C}_U = 0 \\ \mathbf{C}_V^T = -\mathbf{C}_V \qquad & \mathbf{C}_U^T = -\mathbf{C}_U \end{aligned} \tag{4.372}$$

i.e. the orthogonality constraint is preserved, if $\mathbf{C}_V$ and $\mathbf{C}_U$ are skew symmetric. Hence, for each, $\mathbf{C}_V$ and $\mathbf{C}_U$, we have $1/2\,m(m-1)$ parameters to be determined. Therefore, we insert parameterization (4.370) and result (4.372) into (4.368).

$$\begin{aligned} \dot{\mathbf{\Sigma}} &= \mathbf{Y} - \mathbf{V}^T\mathbf{V}\mathbf{C}_V\mathbf{\Sigma} - \mathbf{\Sigma}\mathbf{C}_U^T\mathbf{U}^T\mathbf{U} \\ \dot{\mathbf{\Sigma}} &= \mathbf{Y} - \mathbf{C}_V\mathbf{\Sigma} - \mathbf{\Sigma}\mathbf{C}_U^T \end{aligned} \tag{4.373}$$

Notice that orthogonal property (4.367) has been used. Now, since $\mathbf{\Sigma}$ is a diagonal matrix by definition, all in all $m(m-1)$ independent parameters $c_{V,rc}$, $c_{U,rc}$ in row r and

column c of $\mathbf{C}_V$ and $\mathbf{C}_U$ are determined uniquely by the constraint that the off-diagonal elements of $\dot{\boldsymbol{\Sigma}}$ are set to zero. Therefore, evaluating elements (r,c) and (c,r) in (4.373), we get

$$\sigma_c c_{V,rc} + \sigma_r c_{U,cr} = y_{rc} \qquad \sigma_r c_{V,cr} + \sigma_c c_{U,rc} = y_{cr}\,.$$

Moreover, from (4.372), we have $c_{V,rc} = -c_{C,cr}$, $c_{U,cr} = -c_{U,rc}$ and hence

$$\sigma_c c_{V,rc} - \sigma_r c_{U,rc} = y_{rc} \qquad -\sigma_r c_{V,rc} + \sigma_c c_{U,rc} = y_{cr}$$

which is conveniently written in matrix-vector notation

$$\mathbf{A}_{rc} \cdot \begin{pmatrix} c_{V,rc} \\ c_{U,rc} \end{pmatrix} = \begin{pmatrix} y_{rc} \\ y_{cr} \end{pmatrix} \tag{4.374}$$

where

$$\mathbf{A}_{rc} = \begin{bmatrix} \sigma_c & -\sigma_r \\ -\sigma_r & \sigma_c \end{bmatrix} \tag{4.375}$$

and, according to (4.369),

$$y_{rc} = \mathbf{v}_r^T \mathbf{F} \mathbf{u}_c \quad , \quad y_{cr} = \mathbf{v}_c^T \mathbf{F} \mathbf{u}_r\,. \tag{4.376}$$

Hence evaluation of elements (r,c) and (c,r) in (4.373) results in a coupled $2x2$ system of equations for the elements $c_{V,rc}$ and $c_{U,rc}$ which has a unique solution if $\sigma_r \neq \sigma_c$. Therefore, all parameters are uniquely determined if all singular values are pairwise different and the singular value decomposition update is given by (4.370), (4.373).

However, if $\sigma_r = \sigma_c = \sigma$, the matrix (4.375) is singular

$$\mathbf{A}_{rc} = \sigma \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{4.377}$$

and (4.374) has a solution only if its right hand side is in the column span of $\mathbf{A}_{rc}$, i.e. $y_{rc} = -y_{cr}$ or, using (4.376),

$$\mathbf{v}_r^T \mathbf{F} \mathbf{u}_c + \mathbf{u}_r^T \mathbf{F}^T \mathbf{v}_c = 0\,. \tag{4.378}$$

Moreover, if there is a group of more than two equal singular values, equation (4.378) has to be fulfilled for all SVs of that group. Assume that the g^{th} of those groups contains N_g SVs, $\sigma_{g_1} = \ldots = \sigma_{g_{N_g}} = \tilde{\sigma}_g$, then (4.378) has to hold for every set $\mathbf{v}_j, \mathbf{v}_k, \mathbf{u}_j, \mathbf{u}_k$, $j,k=1,\ldots,N_g$, $j \neq k$. If we stack the output- and input directions, associated with that group into matrices

$$\mathbf{V}_{\tilde{\sigma}_g} = \begin{bmatrix} \mathbf{v}_{g_1} & \cdots & \mathbf{v}_{g_{N_g}} \end{bmatrix} \quad , \quad \mathbf{U}_{\tilde{\sigma}_g} = \begin{bmatrix} \mathbf{u}_{g_1} & \cdots & \mathbf{u}_{g_{N_g}} \end{bmatrix},$$

equation (4.378) can also be written in matrix-vector notation.

$$\mathbf{V}_{\tilde{\sigma}_g}^T \mathbf{F} \mathbf{U}_{\tilde{\sigma}_g} + \mathbf{U}_{\tilde{\sigma}_g}^T \mathbf{F}^T \mathbf{V}_{\tilde{\sigma}_g} \overset{!}{=} \operatorname{diag}\left[\dot{\sigma}_{g_1} \quad \cdots \quad \dot{\sigma}_{g_{N_g}}\right] \tag{4.379}$$

Thereby the off-diagonal elements are associated with the case $j \neq k$ and should thus evaluate to zero, while the diagonal element are associated with $j=k$ which are allowed to have some arbitrary values.

Since $\mathbf{F}$ is fixed, the only possibility for a solution to (4.379) lies in a suitable choice for output- and input directions $\mathbf{V}_{\tilde{\sigma}_g}$, $\mathbf{U}_{\tilde{\sigma}_g}$. Fortunately, for the case of equal singular values, the input and output directions are not unique as will be explained in the following.

Interim Consideration

We will explain, at first, the non-uniqueness of input/output directions in case of equal singular values intuitively for $2x2$ matrices. It is well-known that SVD decomposes the output and input space of a matrix into orthogonal directions, where each input direction $[\mathbf{u}_1 \quad \mathbf{u}_2]$ is mapped to the associated output direction $[\mathbf{v}_1 \quad \mathbf{v}_2]$ and the length of the input vector is scaled by the associated singular value (Figure 4.14a). Hence a circle in the input space is mapped to an ellipse, whose major and minor semiaxes coincide with the orthogonal output directions. For equal singular values, the ellipse regenerates to a circle (Figure 4.14b) and hence there are no distinct orthogonal directions for the semiaxes.

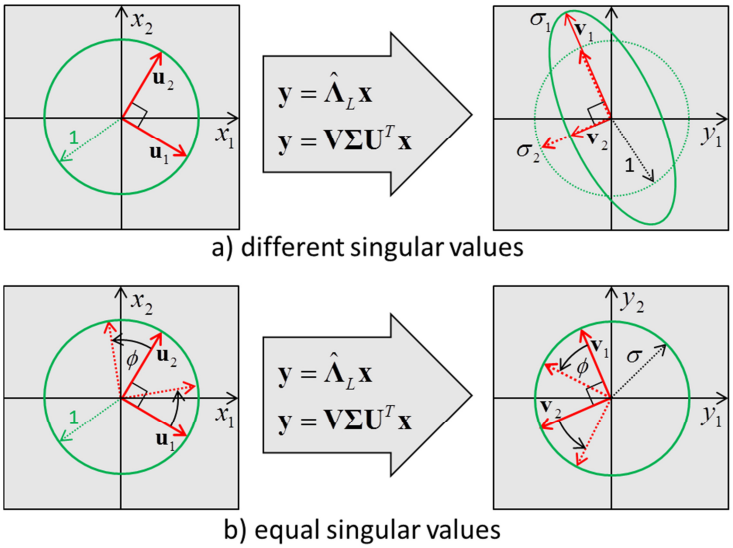


Figure 4.14 Ambiguity of Output- and Input Directions for Equal Singular Values

In fact, if the orthogonal input directions are rotated about some angle ϕ, also the output directions are rotated about the same angle and thus, also the rotated vectors serve as valid input and output directions of the SVD. The new output-/input directions are obtained by post-multiplication of the matrix, containing the original output-/input directions by a rotation matrix

$$\begin{bmatrix}\overline{\mathbf{v}}_1 & \overline{\mathbf{v}}_2\end{bmatrix}=\begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2\end{bmatrix}\begin{bmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{bmatrix}\ ,\ \begin{bmatrix}\overline{\mathbf{u}}_1 & \overline{\mathbf{u}}_2\end{bmatrix}=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix}\begin{bmatrix}\cos\phi & -\sin\phi\\ \sin\phi & \cos\phi\end{bmatrix}$$

which is in fact an orthonormal matrix.

This intuitive result can be generalized to $\mathbb{R}^m$ namely if, in case of equal singular values, the associated output- and input directions are subjected to a number of rotations and/or reflections, then the resulting vectors are still valid output- and input directions. A characteristic of SVD is that the matrices of output- and input directions $\mathbf{V}$, $\mathbf{U}$ are orthonormal. I.e. if we can find another pair of orthonormal matrices $\overline{\mathbf{V}}, \overline{\mathbf{U}}$ that decompose $\hat{\mathbf{\Lambda}}_L$ in the form of (4.366), then

$$\hat{\mathbf{\Lambda}}_L = \overline{\mathbf{V}}\mathbf{\Sigma}\overline{\mathbf{U}}^T \tag{4.380}$$

is also a valid SVD. To that end assume, without loss of generality, that the first N singular values of $\hat{\mathbf{\Lambda}}_L$ are equal, $\sigma_1=\ldots=\sigma_N=\sigma$. The associated output- and input directions are stacked into matrices

$$\mathbf{V}_\sigma=\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_N\end{bmatrix}\ ,\ \mathbf{U}_\sigma=\begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_N\end{bmatrix}, \tag{4.381}$$

as well as the remaining m-N output- and input directions

$$\tilde{\mathbf{V}}=\begin{bmatrix}\mathbf{v}_{N+1} & \cdots & \mathbf{v}_m\end{bmatrix}\ ,\ \tilde{\mathbf{U}}=\begin{bmatrix}\mathbf{u}_{N+1} & \cdots & \mathbf{u}_m\end{bmatrix}. \tag{4.382}$$

Since $\mathbf{v}_1,\ldots,\mathbf{v}_m$ and $\mathbf{u}_1,\ldots,\mathbf{u}_m$ are of unity length and orthogonal to each other, we consequently have

$$\begin{gathered}\mathbf{V}_\sigma^T\mathbf{V}_\sigma=\mathbf{I}\ ,\ \tilde{\mathbf{V}}^T\tilde{\mathbf{V}}=\mathbf{I}\ ,\ \mathbf{U}_\sigma^T\mathbf{U}_\sigma=\mathbf{I}\ ,\ \tilde{\mathbf{U}}^T\tilde{\mathbf{U}}=\mathbf{I}\\ \mathbf{V}_\sigma^T\tilde{\mathbf{V}}=\mathbf{0}\ ,\ \mathbf{U}_\sigma^T\tilde{\mathbf{U}}=\mathbf{0}.\end{gathered} \tag{4.383}$$

With these definitions, the SVD also reads as

$$\hat{\mathbf{\Lambda}}_L=\begin{bmatrix}\mathbf{V}_\sigma & \tilde{\mathbf{V}}\end{bmatrix}\begin{bmatrix}\sigma\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \tilde{\mathbf{\Sigma}}\end{bmatrix}\begin{bmatrix}\mathbf{U}_\sigma^T\\ \tilde{\mathbf{U}}^T\end{bmatrix}=\sigma\mathbf{V}_\sigma\mathbf{U}_\sigma^T+\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{U}}^T \tag{4.384}$$

where

$$\tilde{\mathbf{\Sigma}}=\operatorname{diag}\begin{bmatrix}\sigma_{N+1} & \cdots & \sigma_m\end{bmatrix}. \tag{4.385}$$

Moreover, an orthonormal matrix in $\mathbb{R}^m$ represents a number of rotations and/or reflections ([Tre09]). Therefore, let $\mathbf{Q}\in\mathbb{R}^{N\times N}$ be an orthonormal matrix, i.e. $\mathbf{Q}^T\mathbf{Q}=\mathbf{Q}\mathbf{Q}^T=\mathbf{I}$, and rotate/reflect the columns of $\mathbf{V}_\sigma$, $\mathbf{U}_\sigma$ by $\mathbf{Q}$.

$$\overline{\mathbf{V}}_\sigma=\mathbf{V}_\sigma\mathbf{Q}\ \ ,\ \ \overline{\mathbf{U}}_\sigma=\mathbf{U}_\sigma\mathbf{Q} \tag{4.386}$$

Then, we define modified output- and input direction matrices, which contain the rotated/reflected columns $\overline{\mathbf{v}}_i, \overline{\mathbf{u}}_i$, $i=1,\ldots,N$ of $\overline{\mathbf{V}}_\sigma, \overline{\mathbf{U}}_\sigma$ instead of $\mathbf{v}_1,\ldots,\mathbf{v}_N$, $\mathbf{u}_1,\ldots\mathbf{u}_N$.

$$\overline{\mathbf{V}}=\left[\overline{\mathbf{V}}_\sigma \quad \tilde{\mathbf{V}}\right] \ , \ \overline{\mathbf{U}}=\left[\overline{\mathbf{U}}_\sigma \quad \tilde{\mathbf{U}}\right] \tag{4.387}$$

The columns of $\overline{\mathbf{V}}, \overline{\mathbf{U}}$ are still of unity length and orthogonal to each other. This will be shown exemplarily for $\overline{\mathbf{V}}$ by computing

$$\overline{\mathbf{V}}^T\overline{\mathbf{V}}=\begin{bmatrix}\overline{\mathbf{V}}_\sigma^T\\ \tilde{\mathbf{V}}^T\end{bmatrix}\left[\overline{\mathbf{V}}_\sigma \quad \tilde{\mathbf{V}}\right]=\begin{bmatrix}\mathbf{Q}^T\mathbf{V}_\sigma^T\\ \tilde{\mathbf{V}}^T\end{bmatrix}\left[\mathbf{V}_\sigma\mathbf{Q} \quad \tilde{\mathbf{V}}\right]=\begin{bmatrix}\mathbf{Q}^T\mathbf{V}_\sigma^T\mathbf{V}_\sigma\mathbf{Q} & \mathbf{Q}^T\mathbf{V}_\sigma^T\tilde{\mathbf{V}}\\ \tilde{\mathbf{V}}^T\mathbf{V}_\sigma\mathbf{Q} & \tilde{\mathbf{V}}^T\tilde{\mathbf{V}}\end{bmatrix}=\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}$$

where we have used (4.383) and the orthonormal property of $\mathbf{Q}$.

Further, by (4.384), we get

$$\hat{\mathbf{\Lambda}}_L=\sigma\mathbf{V}_\sigma\overbrace{\mathbf{Q}\mathbf{Q}^T}^{\mathbf{I}}\mathbf{U}_\sigma^T+\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{U}}^T=\sigma\overline{\mathbf{V}}_\sigma\overline{\mathbf{U}}_\sigma^T+\tilde{\mathbf{V}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{U}}^T=\overline{\mathbf{V}}\mathbf{\Sigma}\overline{\mathbf{U}}^T$$

and hence, the modified output- and input directions (4.387) are suitable for an SVD according to (4.380).

Using the recent consideration, the output and input direction matrices in (4.379) are replaced by modified ones, according to (4.386) and we obtain

$$\mathbf{Q}_g^T\mathbf{\Omega}_g\mathbf{Q}_g=\operatorname{diag}\left(\dot{\sigma}_{g_1} \quad \cdots \quad \dot{\sigma}_{g_{N_g}}\right) \tag{4.388}$$

where we have abbreviated

$$\mathbf{\Omega}_g=\mathbf{V}_{\tilde{\sigma}_g}^T\mathbf{F}\mathbf{U}_{\tilde{\sigma}_g}+\mathbf{U}_{\tilde{\sigma}_g}^T\mathbf{F}^T\mathbf{V}_{\tilde{\sigma}_g} \ . \tag{4.389}$$

$\mathbf{\Omega}_g$ is clearly a symmetric matrix and it is well-known, that it has a full system of orthogonal eigenvectors forming a complete basis. It is further diagonalizable by a similarity transformation of the form (4.388) where the columns of $\mathbf{Q}_g$ are the orthogonal eigenvectors of $\mathbf{\Omega}_g$. Therefore, in fact (4.379) can be modified by means orthonormal transformation of the output- and input direction matrices according to (4.386) such that

$$\overline{\mathbf{V}}_{\tilde{\sigma}_g}^T\mathbf{F}\overline{\mathbf{U}}_{\tilde{\sigma}_g}+\overline{\mathbf{U}}_{\tilde{\sigma}_g}^T\mathbf{F}^T\overline{\mathbf{V}}_{\tilde{\sigma}_g}=\operatorname{diag}\left(\dot{\sigma}_{g_1} \quad \cdots \quad \dot{\sigma}_{g_{N_g}}\right). \tag{4.390}$$

Consequently (4.378) is satisfied for every pair of output-/input directions $\overline{\mathbf{v}}_r, \overline{\mathbf{u}}_r$ and $\overline{\mathbf{v}}_c, \overline{\mathbf{u}}_c$ that belong the same singular value and, using columns of $\overline{\mathbf{V}}_{\tilde{\sigma}_g}$, $\overline{\mathbf{U}}_{\tilde{\sigma}_g}$, (4.376) is now

$$\overline{y}_{rc}=\overline{\mathbf{v}}_r^T\mathbf{F}\overline{\mathbf{u}}_c \ , \ \overline{y}_{cr}=\overline{\mathbf{v}}_c^T\mathbf{F}\overline{\mathbf{u}}_r \tag{4.391}$$

and

$$\overline{y}_{rc}=-\overline{y}_{cr}=:b_{rc} \ . \tag{4.392}$$

Now equation (4.374), using matrix (4.377),

$$\tilde{\sigma}_g \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} c_{V,rc} \\ c_{U,rc} \end{pmatrix} = b_{rc} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

has an infinite number of solutions, but a good choice is the one, obtained from the pseudoinverse, minimizing the Euclidean norm.

$$\begin{pmatrix} c_{V,rc} \\ c_{U,rc} \end{pmatrix} = \frac{1}{4\tilde{\sigma}_g} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} b_{rc}$$

$$\begin{pmatrix} c_{V,rc} \\ c_{U,rc} \end{pmatrix} = \frac{b_{rc}}{2\tilde{\sigma}_g} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \tag{4.393}$$

4.7.2 Implementation Aspects

Numerical Issue

The singular values, of course, are functions of time and if two of them approach some common value, the matrix (4.377) gets ill conditioned, leading to large values for $c_{V,rc}$, $c_{U,rc}$. This, in turn, leads to large values for the time derivatives (4.370), (4.373) which potentially causes problems to numeric integration. Assume that singular values only have a difference of a small ε, $\sigma_r=\sigma$, $\sigma_c=\sigma+\varepsilon$, then the inverse of (4.377) is

$$\mathbf{A}_{rc}^{-1} = \frac{-1}{\varepsilon(\varepsilon+2\sigma)} \begin{bmatrix} \sigma & \sigma+\varepsilon \\ \sigma+\varepsilon & \sigma \end{bmatrix}$$

and hence grows unbounded as ε goes to zero. In order to circumvent unbounded time derivatives, a minimum difference $\bar{\varepsilon}$ has to be specified such that singular values are considered equal if

$$|\sigma_r - \sigma_c| \leq \bar{\varepsilon} \tag{4.394}$$

and the solution for $c_{V,rc}$, $c_{U,rc}$ is obtained from (4.393), after a possible rotation/reflection of the associated output- and input directions. This, on the one hand, prevents the SVD time derivatives from growing unbounded for bounded $\mathbf{F}(t)$ but, on the other hand, introduces an artificial error to the algorithm, which deteriorates accuracy.

Real-time Capability

After the matrices $\mathbf{C}_V$ and $\mathbf{C}_U$ have been computed, updates for $\boldsymbol{\Sigma}$, $\mathbf{V}$, $\mathbf{U}$ are obtained for the next time step, using some discrete time integration method for (4.370) and (4.373). These are, overall, $3m^2$ integration steps. Additionally computation of (4.370) requires $2m^3$ multiplications and $2(m-1)m^2$ additions, while computation of (4.373) requires $2m^3$ multiplications and $2(m-1)m^2$ additions for the computation of $\mathbf{Y}$ and extra $2m^2$ additions for summation of the matrices. Overall, we have

- $3m^2$ integration steps
- $4m^3$ multiplications
- $(4m\text{-}2)m^2$ additions

Fortunately, the off-diagonal entries of $\boldsymbol{\Sigma}$ are zero and hence only its m diagonal entries have to be updated. Moreover, due to skew symmetric property of $\mathbf{C}_V$, $\mathbf{C}_U$ and diagonal property of $\boldsymbol{\Sigma}$ the number of computation steps can be further reduced. The steps for computation of (4.370) can be reduced to $2(m-1)m^2$ multiplications and $2(m-2)m^2$ additions. Further, the diagonal elements of $\mathbf{C}_V\boldsymbol{\Sigma}+\boldsymbol{\Sigma}\mathbf{C}_U^T$ vanish and for (4.373), we only have to compute the diagonal entries of $\mathbf{Y}$ which requires $(m+1)m^2$ multiplications and $(m^2-1)m$ additions. Hence, the computational effort can be reduced to

- $(2m+1)m$ integration steps
- $(m\text{-}1)m^2$ multiplications
- $(3m\text{-}4)m^2\text{-}m$ additions

A weak point concerning real-time capability is the fact that an eigenvalue decomposition of $\boldsymbol{\Omega}$ (equation (4.389)) is needed in case of equal singular values. For $N_g=2$ and $N_g=3$, there are analytic solutions available, using quadratic or Cardano's formula for solution of the characteristic polynomial. However for $N_g>3$, the eigenvectors have to be computed numerically using QR algorithm ([Gol85]) which is an iterative algorithm and hence no maximum number of steps (and with it no maximum time effort) for computation of the eigenvectors can be guaranteed.

However, for fixed-wing aircraft application, where the primary – moment producing – controls are elevator, aileron and rudder, the real-time problem could be solved using Cardano's formula.

Numerical Correction for Enforcement of Orthonormality

The orthonormality property is preserved over time if $\mathbf{C}_C, \mathbf{C}_U$ are skew symmetric. However, by discrete time integration, small numerical errors would accumulate over time, destroying orthonormality. It is therefore necessary to add a numerical correction term to $\dot{\mathbf{V}}$ and $\dot{\mathbf{U}}$, which dynamically "pulls" the matrices back to an orthonormal state.

Therefore, a measure is needed for the deviation from orthogonal property. For orthonormal matrices $\mathbf{V}$, $\mathbf{U}$, the expressions

$$\mathbf{N}_V := \mathbf{V}^T\mathbf{V} \quad , \quad \mathbf{N}_U := \mathbf{U}^T\mathbf{U} \tag{4.395}$$

equal the identity matrix according to (4.367). Hence, every deviation from identity is an indication that $\mathbf{V}$ is not orthonormal. Therefore take a closer look on a single element (i,j) of $\mathbf{N}_V$. There we have

$$[\mathbf{N}_V]_{ij} = \mathbf{v}_i^T \mathbf{v}_j .$$

For a perfect orthonormal matrix, we have $\mathbf{v}_i^T \mathbf{v}_j = 1$ for $i = j$ since the columns are of unity length and $\mathbf{v}_i^T \mathbf{v}_i = 0$ for $i \neq j$ since the columns are pairwise orthogonal. Now, if some column i is too long, we have $\mathbf{v}_i^T \mathbf{v}_i > 1$ and $\mathbf{v}_i^T \mathbf{v}_i < 1$ if it is too short. Furthermore, if vector i points into direction of vector j, we have $\mathbf{v}_i^T \mathbf{v}_j > 0$ and if it points in the opposite direction, we have $\mathbf{v}_i^T \mathbf{v}_j < 0$. This motivates a correction of the following form

$$\dot{\mathbf{V}} = \mathbf{V}[\mathbf{C}_V - k_V(\mathbf{N}_V - \mathbf{I})] \quad , \quad \dot{\mathbf{U}} = \mathbf{U}[\mathbf{C}_U - k_U(\mathbf{N}_U - \mathbf{I})] \tag{4.396}$$

for some $k_V, k_U > 0$. Due to the properties discussed recently, the additional expression modifies the time derivatives in a way that too short columns get longer, too long columns get shorter and vectors that are not orthogonal and point into the direction of some other column are rotated such that they become orthogonal again.

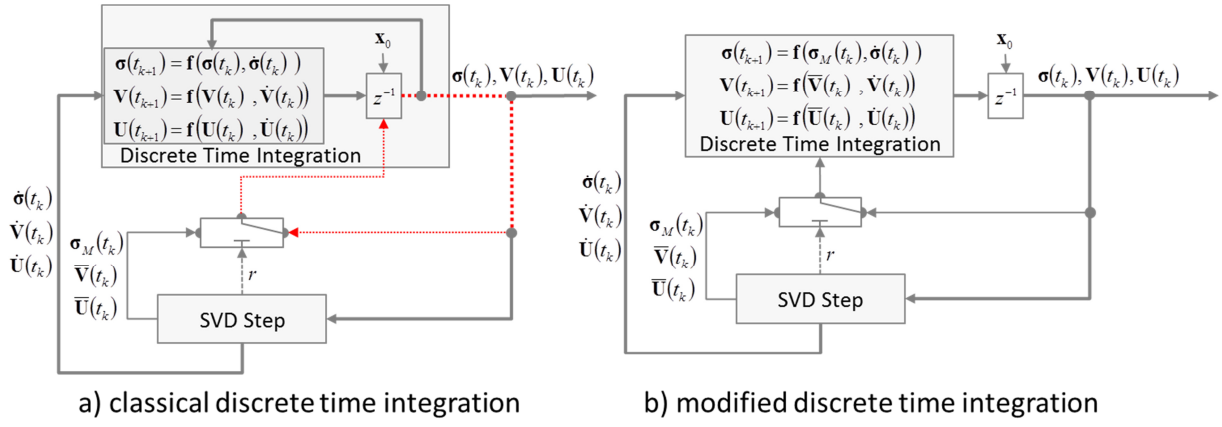
Structural Program Flowchart

For implementation on a computer, the continuous-time algorithm as described above has to be discretized and the algorithm is evaluated only at discrete time instants t_k, $k = 0,1,\cdots$. As the off-diagonal elements of $\boldsymbol{\Sigma}$ are zero anyway, a vector $\boldsymbol{\sigma}^T(t_k) = (\sigma_1(t_k) \;\; \cdots \;\; \sigma_m(t_k))$ containing the SVs instead of the full matrix will be used. Further $\mathbf{V}(t_k) = [\mathbf{v}_1(t_k) \;\; \cdots \;\; \mathbf{v}_m(t_k)]$, the matrix of output directions and $\mathbf{U}(t_k) = [\mathbf{u}_1(t_k) \;\; \cdots \;\; \mathbf{u}_m(t_k)]$, the matrix of input directions are introduced as states. As depicted in Figure 4.15, the whole algorithm is divided into an integration that updates the state and a step function that computes the time derivatives. Notice that discrete time integration methods without direct feed through – such as backward Euler integration – should be used, since an algebraic loop occurs otherwise, which has to be solved iteratively.

It is known from the derivation of the continuous-time algorithm that, in case of equal SVs, the matrices of output/-input directions have to be such that they satisfy condition (4.378). This requires some special mechanism for reset of the states as will be explained in the following. Assume that, there are equal SVs and columns in $\mathbf{V}(t_k)$, $\mathbf{U}(t_k)$ at some time instant t_k that have to be rotated/reflected. Then, the states need to be reset to the transformed output-/input directions at the current time instant t_k.

Classical discrete time integrators (Figure 4.15a) store the latest state internally and use them for computation of the subsequent time instant. This eliminates the possibility to reset the (stored) state in the current time instant, since it results in an

algebraic loop (marked with red dashed lines). The modified algorithm, depicted in Figure 4.15b, does not store the state internally, but receives it from extern. This enables the SVD step function to reset the states immediately, by setting some reset flag r, which routes the rotated output/-input directions $\overline{\mathbf{V}}(t_k)$, $\overline{\mathbf{U}}(t_k)$ of the SVD step function to the integration algorithm that, now, computes the subsequent time step based on the rotated states.



a) classical discrete time integration b) modified discrete time integration

Figure 4.15 SVD Integration

The program structure of the SVD step function is shown in Table 4.1. It receives the SV vector, output-/ input matrices and the desired time derivative $\mathbf{F}(t_k)$ of the current time instant t_k and additional parameters $\bar{\varepsilon}$, $\underline{\sigma}$, on the one hand, the limit distance for a pair of SVs such that they are considered equal according to (4.394) and, on the other hand, a lower bound the SVs are not allowed to underrun. Besides state derivatives and rotated output-/input directions, the SVD step function also provides a modified SV vector $\boldsymbol{\sigma}_M(t_k)$, which is necessary for the limitation of the SVs to their lower bound $\underline{\sigma}$ and a flag r, which triggers a state reset according to Figure 4.15b.

The step function itself contains three main subroutines (Table 4.2, Table 4.4, Table 4.5), the actual computation of the time derivatives and the limitation of the SVs to their lower bound. In order to conduct the rotation/reflection of the output/-input directions in a well-ordered manner, it is suitable to identify groups of equal SVs and sort the SVD, such that equal SVs are grouped together. This is done by subroutine 1. In subroutine 2, on the one hand, the transformation of output-/input directions is applied to those groups that contain at least two SVs and, on the other hand, the SVD grouping of subroutine 1 is reversed. Subroutine 3 computes the matrices $\mathbf{C}_V$, $\mathbf{C}_U$ which enables the computation of the state derivative. Finally, the SVs that underrun $\underline{\sigma}$ during integration are corrected to the lower bound and the associated state derivatives $\dot{\boldsymbol{\sigma}}(t_k)$ are limited to positive values.

Subroutine 1 reorders the SVs $\sigma_i(t_k)$, and the output-/input directions $\mathbf{v}_i(t_k)$, $\mathbf{u}_i(t_k)$ within $\mathbf{V}(t_k)$, $\mathbf{U}(t_k)$ such that equal SVs are grouped together in ascending order. The minimum of the remaining SVs that have not been assigned to a group is determined by subroutine 1a (Table 4.3) in each iteration. Additionally subroutine 1 computes the number of identified groups G and a vector $\mathbf{M}^T=(M_1,\ldots,M_m)$ that contains the accumulated numbers of SVs in each group (the g^{th} element contains the accumulated number of SVs from the first up to the g^{th} group and $M_g=0$ for $g>G$). The sorted SVD is of the following form

$$\boldsymbol{\sigma}_S^T=\operatorname{diag}\left(\boldsymbol{\sigma}_{S,1} \quad \cdots \quad \boldsymbol{\sigma}_{S,G}\right).$$

$$\mathbf{V}_S(t_k)=\left[\mathbf{V}_{\tilde{\sigma}_1}(t_k) \quad \cdots \quad \mathbf{V}_{\tilde{\sigma}_G}(t_k)\right],\ \mathbf{U}_S(t_k)=\left[\mathbf{U}_{\tilde{\sigma}_1}(t_k) \quad \cdots \quad \mathbf{U}_{\tilde{\sigma}_G}(t_k)\right] \tag{4.397}$$

Thereby

$$\boldsymbol{\sigma}_g(t_k)=\tilde{\sigma}_g(t_k)\operatorname{ones}(N_g,1)$$

is a subpart that contains the SVs of group g

$$\mathbf{V}_{\tilde{\sigma}_g}(t_k)=\left[\mathbf{v}_{g_1}(t_k) \quad \cdots \quad \mathbf{v}_{g_{N_g}}(t_k)\right],\ \mathbf{U}_{\tilde{\sigma}_g}(t_k)=\left[\mathbf{u}_{g_1}(t_k) \quad \cdots \quad \mathbf{u}_{g_{N_g}}(t_k)\right] \tag{4.398}$$

are matrices whose columns are stacked with output-/input directions associated with the current group and

$$N_g=\begin{cases}M_1 & \text{for } g=1\\ M_g-M_{g-1} & \text{for } g=2,\ldots,G\end{cases}$$

is the number of SVs in the respective group. Finally the subroutine provides a vector $\mathbf{p}^T=(p_1,\ldots,p_m)$ that contains the indices of the unsorted positions, which is needed if the grouping is to be reversed.

Subroutine 2 rotates/reflects the sorted output-/input directions (4.398) of those groups that contain more than one SV. Therefore,

$$\mathbf{Y}_g=\mathbf{V}_{\tilde{\sigma}_g}^T(t_k)\mathbf{F}(t_k)\mathbf{U}_{\tilde{\sigma}_g}(t)\ ,\ \boldsymbol{\Omega}_g=\mathbf{Y}_g+\mathbf{Y}_g^T$$

is computed for each group according to (4.389) and, according to (4.386), $\mathbf{V}_{\tilde{\sigma}_g}(t_k)$, $\mathbf{U}_{\tilde{\sigma}_g}(t_k)$ are transformed by a matrix $\mathbf{Q}_g$ whose columns are the orthogonal eigenvectors of $\boldsymbol{\Omega}_g$. The transformed output-/input directions are reintegrated into the respective full matrices (4.398) and finally the grouping is reversed using the vector $\mathbf{p}$ of original indices.

Subroutine 3 computes the skew symmetric matrices $\mathbf{C}_V$, $\mathbf{C}_U$. According to (4.369), we compute

$$\overline{\mathbf{Y}}(t_k)=\overline{\mathbf{V}}^T(t_k)\mathbf{F}(t_k)\overline{\mathbf{U}}(t_k)$$

where we have used the rotated output-/input directions of subroutine 2. Since $\mathbf{C}_V$, $\mathbf{C}_U$ are skew symmetric, it is sufficient to compute the lower left half of each matrix. Then, for each row r and column c, it is checked, whether the involved SVs are equal according to (4.394). If this is the case, $c_{V,rc}$, $c_{U,rc}$ are computed according to (4.393), where the mean values of the SVs and the right hand side values b_{rc} are taken to account for potential numerical deviations. In case of different SVs, the associated $c_{V,rc}$, $c_{U,rc}$ are computed by solving (4.374).

Table 4.1 SVD Step Algorithm

SVD Step

Input:

1. $\boldsymbol{\sigma}^T(t_k)=(\sigma_1(t_k) \ \cdots \ \sigma_m(t_k))$: Vector with SVs at current time step t_k
2. $\mathbf{V}(t_k)=[\mathbf{v}_1(t_k) \ \cdots \ \mathbf{v}_m(t_k)]$, $\mathbf{U}(t_k)=[\mathbf{u}_1(t_k) \ \cdots \ \mathbf{u}_m(t_k)]$:
   Matrix of output-/input directions at time step t_k
3. $\mathbf{F}(t_k)$: Desired time derivative at time step t_k
4. $\bar{\varepsilon}$: Minimum distance of SVs to be considered equal
5. $\underline{\sigma}$: Lower limit of SVs

Output:

1. $\dot{\boldsymbol{\sigma}}^T(t_k)=(\dot{\sigma}_1(t_k) \ \cdots \ \dot{\sigma}_m(t_k))$: Time derivative of SVs at time step t_k
2. $\dot{\mathbf{V}}(t_k)=(\dot{\mathbf{v}}_1(t_k) \ \cdots \ \dot{\mathbf{v}}_m(t_k))$, $\dot{\mathbf{U}}(t_k)=(\dot{\mathbf{u}}_1(t_k) \ \cdots \ \dot{\mathbf{u}}_m(t_k))$
   Time derivative of output-/input direction matrix at time step t_k
3. $r=0$ [bool]: Flag, indicating whether reset is necessary
4. $\boldsymbol{\sigma}_M^T(t_k)=(\sigma_{M,1}(t_k) \ \cdots \ \sigma_{M,m}(t_k))$: modified SVs
5. $\bar{\mathbf{V}}(t_k)=(\bar{\mathbf{v}}_1(t_k) \ \cdots \ \bar{\mathbf{v}}_m(t_k))$, $\bar{\mathbf{U}}(t_k)=(\bar{\mathbf{u}}_1(t_k) \ \cdots \ \bar{\mathbf{u}}_m(t_k))$
   Matrix with rotated output-/input directions

Subroutine 1) *Group equal SVs together*

Description: Groups of equal SVs are identified and SVs, output-/input direction matrices are sorted such that equal SVs are grouped together in ascending order.

Input: $\boldsymbol{\sigma}(t_k)$, $\mathbf{V}(t_k)$, $\mathbf{U}(t_k)$

Output:

1. $\boldsymbol{\sigma}_S^T(t_k)=(\sigma_{S,1}(t_k) \ \cdots \ \sigma_{S,m}(t_k))$, $\mathbf{V}_S(t_k)=[\mathbf{v}_{S,1}(t_k) \ \cdots \ \mathbf{v}_{S,m}(t_k)]$, $\mathbf{U}_S(t_k)=[\mathbf{u}_{S,1}(t_k) \ \cdots \ \mathbf{u}_{S,m}(t_k)]$
   Sorted SVs, output-/ input direction matrices
2. G: Numbers of identified groups
3. $\mathbf{M}=(M_1,\ldots,M_m)$: Accumulated numbers of SVs in each group (zero indicates, that group is not assigned)
4. $\mathbf{p}^T=(p_1,\ldots,p_m)$:
   Indices of unsorted positions (enable reverse of grouping)

Subroutine 2) *Rotate/reflect output-/input directions and reverse sorting*

Input: $\mathbf{M}$, $\mathbf{p}$, G , $\boldsymbol{\sigma}_S(t_k)$, $\mathbf{V}_S(t_k)$, $\mathbf{U}_S(t_k)$

Output:

1. $\boldsymbol{\sigma}_M^T(t_k)=(\sigma_{M,1}(t_k) \ \cdots \ \sigma_{M,m}(t_k))$: SVs in original order (limited)
2. $\bar{\mathbf{V}}(t_k)=[\bar{\mathbf{v}}_1(t_k) \ \cdots \ \bar{\mathbf{v}}_m(t_k)]$, $\bar{\mathbf{U}}(t_k)=[\bar{\mathbf{u}}_1(t_k) \ \cdots \ \bar{\mathbf{u}}_m(t_k)]$:
   Rotated/reflected output-/input directions matrices in original order

Y-matrix, computed with rotated output/-input directions: $\bar{\mathbf{Y}}(t_k)=\bar{\mathbf{V}}^T(t_k)\mathbf{F}(t_k)\bar{\mathbf{U}}(t_k)$

Subroutine 3) ***Compute*** $\mathbf{C}_V$, $\mathbf{C}_U$

Input:

1. $\bar{\mathbf{Y}}(t_k)$: Y-matrix, computed with rotated output/-input directions

Output:

1. $\mathbf{C}_V,\mathbf{C}_U\in\mathbb{R}^{m\times m}$, $\mathbf{C}_U$ *: Matrices for computation of time derivatives of output-/input directions*

Time derivatives of i^{th} SV: $\dot{\sigma}_i(t_k)=\bar{\mathbf{v}}_i^T(t_k)\mathbf{F}(t_k)\bar{\mathbf{u}}_i(t_k)$

for $i=1,\ldots,m$

Description: Limit current SV if it has hit lower bound

Current SV hits lower bound $\sigma_i(t_k)\le\underline{\sigma}$

yes

- Limit time derivative of current SV: $\dot{\sigma}_i(t_k)=\max(0,\dot{\sigma}_i(t_k))$
- Set current SV to minimum value $\sigma_{M,i}(t_k)=\underline{\sigma}$
- Trigger "reset" due to limitation of current SV: $r=1$

no

Time derivatives of output-/input directions:

$\dot{\mathbf{V}}(t_k)=\bar{\mathbf{V}}(t_k)(\mathbf{C}_V-k_V\mathbf{N}_V)$, $\dot{\mathbf{U}}(t_k)=\bar{\mathbf{U}}(t_k)(\mathbf{C}_U-k_U\mathbf{N}_U)$, $\mathbf{N}_V=\bar{\mathbf{V}}^T(t_k)\bar{\mathbf{V}}(t_k)$, $\mathbf{N}_U=\bar{\mathbf{U}}^T(t_k)\bar{\mathbf{U}}(t_k)$

Table 4.2 SVD Step – Subroutine 1

Subroutine 1) *Group equal SVs together*
Description: Groups of equal SVs are identified and SV, output-/input direction matrices are sorted such that equal SV are grouped together in ascending order.

Input:
1. $\boldsymbol{\sigma}^T(t_k)=(\sigma_1(t_k) \ \cdots \ \sigma_m(t_k))$: Vector with SVs at current time step t_k
2. $\mathbf{V}(t_k)=[\mathbf{v}_1(t_k) \ \cdots \ \mathbf{v}_m(t_k)]$, $\mathbf{U}(t_k)=[\mathbf{u}_1(t_k) \ \cdots \ \mathbf{u}_m(t_k)]$: Matrix of output-/input directions at time step t_k

Output:
1. $\boldsymbol{\sigma}_S^T(t_k)=(\sigma_{S,1}(t_k) \ \cdots \ \sigma_{S,m}(t_k))$, $\mathbf{V}_S(t_k)=[\mathbf{v}_{S,1}(t_k) \ \cdots \ \mathbf{v}_{S,m}(t_k)]$, $\mathbf{U}_S(t_k)=[\mathbf{u}_{S,1}(t_k) \ \cdots \ \mathbf{u}_{S,m}(t_k)]$: Sorted SVs, output-/ input direction matrices
2. $G=0$: Number of identified groups
3. $\mathbf{M}^T=(M_1,\ldots,M_m)$: Accumulated numbers of SVs in each group (zero indicates, that group is not assigned)
4. $\mathbf{p}^T=(p_1,\ldots,p_m)$: Indices of unsorted positions (enable reverse of sorting)

Variable: $\boldsymbol{\sigma}_{rem}=\text{ones}(1,m)$
Description: Positions of SV that are not yet assigned to some group

Variable: $c=0$
Description: Counter for sorted positions

while $\text{sum}(\boldsymbol{\sigma}_{rem})>0$
Description: Not all singular values are assigned to some group

Increase number of identified groups: $G{+}{+}$

Subroutine 1a) *Find minimum of remaining SVs*
Input: $\boldsymbol{\sigma}_{rem}, \boldsymbol{\sigma}(t_k)$
Output: $\sigma_{\min}$: Minimum of SVs that are not assigned to some group yet

for $i=1,\ldots,m$
Description: Parse remaining SVs whether they belong to the current group.

SV belongs to current group
$\sigma_i(t_k)\le\bar{\varepsilon}+\sigma_{\min}$ & $\sigma_{rem,i}=1$

yes:
- Increase counter: $c{+}{+}$
- Increase accumulated member count of current group $M_G=c$
- Store original position of SV: $p_c=i$
- Mark current SV as assigned to a group $\sigma_{rem,i}=0$
- Assign sorted variables $\mathbf{v}_{S,c}(t_k):=\mathbf{v}_i(t_k)$, $\mathbf{u}_{S,c}(t_k):=\mathbf{u}_i(t_k)$, $\sigma_{S,c}(t_k):=\sigma_i(t_k)$

no:

Table 4.3 SVD Step – Subroutine 1a

Subroutine 1a) *Find minimum of remaining SV*

Input:
1. $\boldsymbol{\sigma}^T(t_k)=(\sigma_1(t_k) \ \cdots \ \sigma_m(t_k))$: Vector with SVs at current time step t_k
2. $\boldsymbol{\sigma}_{rem}=(\sigma_{rem,1} \ \cdots \ \sigma_{rem,m})$: Positions of SV that are not yet assigned to some group

Output:
1. $\sigma_{\min} := 0$: Minimum of SVs that are not assigned to some group yet

for $i=1,\ldots,m$

Current SV smaller than $\sigma_{\min}$ and not assigned to some group
$\sigma_i(t_k)<\sigma_{\min}$ & $\sigma_{rem,i}=1$

yes:
- Assign new minimum SV: $\sigma_{\min}:=\sigma_i(t_k)$

no:

Table 4.4 SVD Step – Subroutine 2

Subroutine 2) *Rotate/reflect output-/input directions and reverse sorting*

Input:

1. $\mathbf{M}$: Accumulated numbers of SV in each group (zero indicates, that group is not assigned)
2. $\mathbf{p}^T=(p_1,\ldots,p_m)$: Indices of unsorted positions (enable reverse of sorting
3. $\boldsymbol{\sigma}_S^T(t_k)=\begin{pmatrix}\sigma_{S,1}(t_k) & \cdots & \sigma_{S,m}(t_k)\end{pmatrix}$, $\mathbf{V}_S(t_k)=\begin{bmatrix}\mathbf{v}_{S,1}(t_k) & \cdots & \mathbf{v}_{S,m}(t_k)\end{bmatrix}$, $\mathbf{U}_S(t_k)=\begin{bmatrix}\mathbf{u}_{S,1}(t_k) & \cdots & \mathbf{u}_{S,m}(t_k)\end{bmatrix}$: Sorted SVs, output- and input direction matrices

Output:

1. $\boldsymbol{\sigma}_M^T(t_k)=\begin{pmatrix}\sigma_{M,1}(t_k) & \cdots & \sigma_{M,m}(t_k)\end{pmatrix}$: Modified SVs in original order (limited)
2. $\overline{\mathbf{V}}(t_k)=\begin{bmatrix}\overline{\mathbf{v}}_1(t_k) & \cdots & \overline{\mathbf{v}}_m(t_k)\end{bmatrix}$, $\overline{\mathbf{U}}(t_k)=\begin{bmatrix}\overline{\mathbf{u}}_1(t_k) & \cdots & \overline{\mathbf{u}}_m(t_k)\end{bmatrix}$: Rotated/reflected output-/input direction matrices in original order

Variable: $\overline{\mathbf{V}}_S(t_k)=\begin{bmatrix}\overline{\mathbf{v}}_{S,1}(t_k) & \cdots & \overline{\mathbf{v}}_{S,m}(t_k)\end{bmatrix}$, $\overline{\mathbf{U}}_S(t_k)=\begin{bmatrix}\overline{\mathbf{u}}_{S,1}(t_k) & \cdots & \overline{\mathbf{u}}_{S,m}(t_k)\end{bmatrix}$

Description: Sorted and rotated output-/input direction matrices

for $g=1,\ldots,G$

Description: Loop over all groups

Current group is first group: $i=1$

Yes	no
Number of SV in current group $N:=M_g$	Number of SV in current group $N:=M_g-M_{g-1}$
Begin index of current group: $N_0:=1$	Begin index of current group: $N_0:=M_{i-1}+1$

Output-/input directions of current group

$\mathbf{V}_{\tilde{\sigma}}:=\begin{bmatrix}\mathbf{v}_{S,N_0}(t_k) & \cdots & \mathbf{v}_{S,N_0+N-1}(t_k)\end{bmatrix}$, $\mathbf{U}_{\tilde{\sigma}}:=\begin{bmatrix}\mathbf{u}_{S,N_0}(t_k) & \cdots & \mathbf{u}_{S,N_0+N-1}(t_k)\end{bmatrix}$

More than 1 SV in current group: $N>1$

yes	no
Interim result: $\mathbf{Y}_{\tilde{\sigma}}:=\mathbf{V}_{\tilde{\sigma}}^T\mathbf{F}(t_k)\mathbf{U}_{\tilde{\sigma}}$, $\boldsymbol{\Omega}_{\tilde{\sigma}}:=\mathbf{Y}_{\tilde{\sigma}}+\mathbf{Y}_{\tilde{\sigma}}^T$	Do not rotate/reflect output-/input directions: $\overline{\mathbf{V}}_{\tilde{\sigma}}=\mathbf{V}_{\tilde{\sigma}}$, $\overline{\mathbf{U}}_{\tilde{\sigma}}=\mathbf{U}_{\tilde{\sigma}}$
Eigenvector matrix of $\boldsymbol{\Omega}_{\tilde{\sigma}}$: $\mathbf{Q}_{\tilde{\sigma}}$	
Rotate/reflect output-/input directions: $\overline{\mathbf{V}}_{\tilde{\sigma}}=\mathbf{V}_{\tilde{\sigma}}\mathbf{Q}_{\tilde{\sigma}}$, $\overline{\mathbf{U}}_{\tilde{\sigma}}=\mathbf{U}_{\tilde{\sigma}}\mathbf{Q}_{\tilde{\sigma}}$	
Trigger "reset" due to rotated output-/input directions: $r=1$	

Reintegrate o-/i directions: $\overline{\mathbf{V}}_{S,[N_0,N_0+N-1]}(t_k):=\overline{\mathbf{V}}_{\tilde{\sigma}}$, $\overline{\mathbf{U}}_{S,[N_0,N_0+N-1]}(t_k):=\overline{\mathbf{U}}_{\tilde{\sigma}}$

for $i=1,\ldots,m$

Description: Reverse sorting

$\sigma_{M,p_i}(t_k)=\sigma_{S,i}(t_k)$, $\overline{\mathbf{v}}_{p_i}(t_k)=\overline{\mathbf{v}}_{S,i}(t_k)$, $\overline{\mathbf{u}}_{p_i}(t_k)=\overline{\mathbf{u}}_{S,i}(t_k)$

Table 4.5 SVD Step – Subroutine 3

Subroutine 3) *Compute* $\mathbf{C}_V$, $\mathbf{C}_U$

Input: $\overline{\mathbf{Y}}(t_k)=\overline{\mathbf{V}}^T(t_k)\mathbf{F}(t_k)\overline{\mathbf{U}}(t_k)$: Y-matrix, computed with rotated output/-input directions

Output: $\mathbf{C}_V,\mathbf{C}_U\in\mathbb{R}^{m\times m}$:

Matrices for computation of time derivatives of output-/input directions

for $r=2,\ldots,m$

Description: Loop over all rows

for $c=1,\ldots,r-1$

Description: Loop over relevant columns of current row

Difference of current SV pair is smaller than limit: $|\sigma_r(t_k)-\sigma_c(t_k)|\le\bar{\varepsilon}$

Yes	no
Mean SV: $\sigma_{rc}=0.5(\sigma_r(t_k)+\sigma_c(t_k))$	System matrix: $\mathbf{A}_{rc}=\begin{bmatrix}\sigma_c(t_k) & -\sigma_r(t_k)\\ -\sigma_r(t_k) & \sigma_c(t_k)\end{bmatrix}$
Mean left hand side: $b_{rc}=0.5\left(\left[\overline{\mathbf{Y}}\right]_{rc}-\left[\overline{\mathbf{Y}}\right]_{cr}\right)$	Compute $\mathbf{C}_V$, $\mathbf{C}_U$: $\begin{pmatrix}c_{V,rc}\\ c_{U,rc}\end{pmatrix}=\mathbf{A}_{rc}^{-1}\begin{pmatrix}[\mathbf{Y}(t_k)]_{rc}\\ [\mathbf{Y}(t_k)]_{cr}\end{pmatrix}$ $\begin{pmatrix}c_{V,cr}\\ c_{U,cr}\end{pmatrix}=-\mathbf{A}_{rc}^{-1}\begin{pmatrix}[\mathbf{Y}(t_k)]_{rc}\\ [\mathbf{Y}(t_k)]_{cr}\end{pmatrix}$
Compute $\mathbf{C}_V$, $\mathbf{C}_U$ using pseudoinverse: $c_{V,rc}=b_{rc}/(2\sigma_{rc})$, $c_{V,cr}=-c_{V,rc}$ $c_{U,rc}=-b_{rc}/(2\sigma_{rc})$, $c_{U,cr}=-c_{U,rc}$	

4.7.3 Open Topics

The SVD update algorithm provides a measure to overcome the diagonal constraint for the adaptive estimate of the control effectiveness matrix while it prevents the matrix from becoming singular. However, the developed algorithm is not necessarily restricted to applications in adaptive control, but could be useful in any case where some matrix, that should not become singular, is updated by a differential equation. Nevertheless, if applied to MRAC schemes, there are still some open topics and weaknesses that have to be discussed.

Commonly, limitation of adaptive parameters in MRAC is implemented via projection operator (Appendix D.1), which allows the restriction of the adaptive parameter to a convex set. The special structure of the projection operator using a convex function allows the incorporation of the projection operator into Lyapunov analysis (particularly due to inequality (D.5) in Appendix D.1). The SVD update algorithm restricts the adaptive parameters to those matrices whose SVs do not underrun some positive limit which is not necessarily a convex set. Hence, one could not hope to obtain a result similar to the projection operator and a proof of stability for the SVD update algorithm applied to an MRAC scheme remains unsolved.

Another issue is a proper choice for $\bar{\varepsilon}$, the minimum distance for two SVs that are considered as different values according to (4.394). On the one hand $\bar{\varepsilon}$ has to be sufficiently small such that the error that is made by setting different SVs to the same value is sufficiently small. On the other hand, the elements of $\mathbf{C}_V$, $\mathbf{C}_U$ grow unbounded, if the SVs approach some common value as explained in section 4.7.2. This implies that also the time derivatives of $\mathbf{V}$ and $\mathbf{U}$ grow unbounded, which causes numerical problems to the discrete time integration. Hence $\bar{\varepsilon}$ has to be chosen sufficiently large. As an illustrative example, recall that the input and output directions have to fulfill the constraint to remain a unity length vector. Hence, the time derivative is perpendicular to its respective direction, which is a direct result of equation (4.371). Figure 4.16 illustrates the situation. Obviously, the vector $\mathbf{v}_i$ remains approximately unity length, if the time derivative is sufficiently small (left picture), while the unit length condition is severely violated, if the time derivative is comparatively large (left picture).

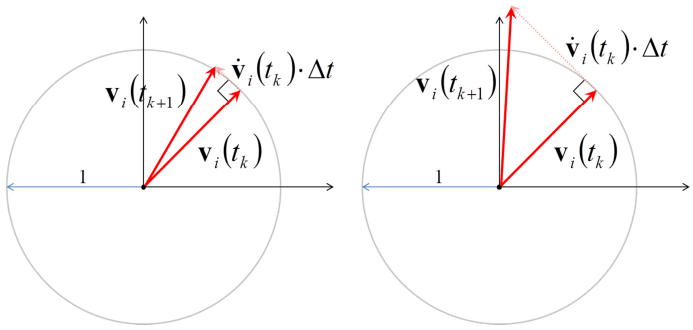


Figure 4.16 Integration of Output-/Input Directions for large Time Derivatives

The decision for a proper bound for $\bar{\varepsilon}$ has to consider this issue namely, for a bounded set of allowable desired time derivatives $\mathbf{F}(t)$, what is the smallest $\bar{\varepsilon}$ such that all time derivatives are sufficiently small. This, of course, at first, requires a definition of a “sufficiently small time derivative”. A first guess is obtained from Figure 4.16. Usually, the algorithm is executed with some constant sample time Δt and the length of the dashed vector in Figure 4.16 is $\left\|\dot{\mathbf{v}}_i(t_k)\right\|_2 \Delta t$. Now, if this vector is considerably smaller than 1, then the length is approximately preserved and hence we require

$$\left\|\dot{\mathbf{v}}_i\right\|_2 << \frac{1}{\Delta t}.$$

Another issue of this algorithm occurs, if two singular have reached equal values At some time instant. In this case, it could occur that these singular values stick together, if their difference in the subsequent time step is still below the limit.

Further parameters that have to be chosen are gains k_V, k_U for dynamic enforcement of orthogonal property (4.396). On the one hand they have to be sufficiently large such that they compensate for numerical errors (the magnitude of numerical error also depends on the choice of $\bar{\varepsilon}$ _ Figure 4.16) and, on the other hand, not too large such that they increase the time derivatives of $\mathbf{V}$, $\mathbf{U}$ too much by themselves, since this might also result in numerical problems.

So far, the SVD update algorithm has been derived for quadratic matrices, but it can also be extended to the rectangular case. This, however, is not considered here and left for future.

Chapter 5

Application and Simulation

In this chapter, the capabilities of the control algorithms, developed in Chapter 4, will be demonstrated. In sections 5.1 and 5.2, the effectiveness of MMQ modification as well as SVD update algorithm is shown for simple linear state space models and in section 5.3 a high fidelity implementation of an MRAC angular rate command system for the FSD Extreme Star is presented. Thereby we will restrict ourselves to variant 1, including MMQ modification (sections 4.3 and 4.4) as the thrust vector controls are not suitable for a use within the linearizing state feedback. This is due to the fact that thrust, produced by the propellers decays with increasing velocity and, there are conditions within the flight envelope, where the propellers do not produce any thrust at all. This in turn implies that moments, induced by the lever arm of the thrust relative to the c.g., are not produced, too. Hence, a minimum control authority over the whole flight enveloped cannot be guaranteed as required by assumption E.

Particularly, the incorporation of redundant control channels to the dynamic inversion framework in section 4.1.7 allows for an airborne system, tolerant w.r.t. to actuator failures, as functionality of failed actuators is resumed by the remaining ones.

5.1 MMQ Modification

Multi model Q modification has been introduced in order to improve adaptation performance. This section shows the capabilities of the approach and its superiority over single model q-modification, by means of a linear approximation of the Extreme Star's lateral dynamics with uncertain control effectiveness.

The states of the lateral dynamics are velocity, roll rate, yaw rate and bank angle, i.e. $\mathbf{x}^T = (\Delta V \quad p \quad r \quad \Phi)$. Notice that velocity state is in fact the deviation from the trimmed value $V_0 = 20\,m/s$, while the trim values of the other states are zero anyway. Lateral controls are aileron and rudder deflection relative to the trim values, i.e. $\mathbf{u}^T = (\Delta\xi \quad \Delta\zeta)$. The linear state space approximation of the lateral motions is

$$\dot{\mathbf{x}} = \mathbf{A}\,\mathbf{x} + \mathbf{B}\,\Lambda\,\mathbf{u}$$

where

$$\mathbf{A}=\begin{bmatrix} -0.2186 & 1.5454 & -20.005 & 9.7696 \\ -3.664 & -19.75 & 3.592 & -1.2392\cdot 10^{-5} \\ 1.1792 & -0.90184 & -1.2777 & 3.569\cdot 10^{-6} \\ -1.5683\cdot 10^{-7} & 1 & 0.02618 & -4.4023\cdot 10^{-10} \end{bmatrix} \quad \mathbf{B}=\begin{bmatrix} 23.653 & 3.6367 \\ -432.74 & 11.947 \\ -13.858 & -30.478 \\ 0 & 0 \end{bmatrix}.$$

$\mathbf{\Lambda}$ is the unknown control effectiveness that has to be estimated adaptively. The output $\mathbf{y}$ to be controlled are bank angle and yaw rate, i.e.

$$\mathbf{y}=\mathbf{C}\mathbf{x} \quad , \quad \mathbf{C}=\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

The control law is

$$\mathbf{u}=\hat{\mathbf{\Lambda}}^{-1}\left(-\mathbf{K}\mathbf{x}+\mathbf{H}\mathbf{r}\right)$$

with $\hat{\mathbf{\Lambda}}$, an adaptive estimate for $\mathbf{\Lambda}$. The feedback and feed forward matrices of the control law are designed, based on decoupling control of Falb/Wolovich (refer to Appendix E), which achieves a dynamically decoupled control of bank angle and yaw rate.

$$\mathbf{K}=\begin{bmatrix} 7.3070 & 36.751 & -8.1288 & -9.1288 \\ -42.015 & 12.88 & 12.807 & 4.1607 \end{bmatrix}\cdot 10^{-3} \quad \mathbf{H}=\begin{bmatrix} -9.1288 & 0.83480 \\ 4.1508 & -32.431 \end{bmatrix}\cdot 10^{-3}$$

The reference dynamics are

$$\dot{\mathbf{x}}_M=\mathbf{A}_M\mathbf{x}_M+\mathbf{B}_M\mathbf{r}$$

where $\mathbf{A}_M$=$\mathbf{A}$-$\mathbf{B}\mathbf{K}$, $\mathbf{B}_M$=$\mathbf{B}_H$. In frequency domain, decoupling of the input/output relationship is specifically visible.

$$\mathbf{G}_M(s)=\mathbf{C}(s\mathbf{I}-\mathbf{A}_M)\mathbf{B}_M=\begin{bmatrix} \dfrac{4.4}{(s+2)^2} & 0 \\ 0 & \dfrac{1}{s+1} \end{bmatrix}$$

Using the above definitions, the closed loop dynamics read as

$$\dot{\mathbf{x}}=\mathbf{A}_M\mathbf{x}+\mathbf{B}_M\mathbf{r}-\mathbf{B}\tilde{\mathbf{\Lambda}}\mathbf{u}\,. \qquad (5.1)$$

Thereby $\tilde{\mathbf{\Lambda}}=\hat{\mathbf{\Lambda}}-\mathbf{\Lambda}$ is the parameter-estimation-error. With definition of the tracking error $\mathbf{e}$=$\mathbf{x}_M$-$\mathbf{x}$, we obtain

$$\dot{\mathbf{e}}=\mathbf{A}_M\mathbf{e}+\mathbf{B}\tilde{\mathbf{\Lambda}}\mathbf{u}\,.$$

The update laws for the parameter estimates are derived from the following Lyapunov function candidate

$$V\left(\mathbf{e},\tilde{\mathbf{\Lambda}}\right)=\mathbf{e}^T\mathbf{P}\mathbf{e}+\mathrm{tr}\left[\tilde{\mathbf{\Lambda}}\mathbf{\Gamma}^{-1}\tilde{\mathbf{\Lambda}}^T\right]$$

where the symmetric positive definite $\mathbf{P}$ solves

$$\mathbf{A}_M^T\mathbf{P}+\mathbf{P}\mathbf{A}_M=-\mathbf{Q}$$

for some symmetric positive definite $\mathbf{Q}$. The resulting update law is

$$\dot{\hat{\mathbf{\Lambda}}}^T=-\mathbf{\Gamma}\left(\mathbf{u}\mathbf{e}^T\mathbf{P}\mathbf{B}+\mathbf{M}^T\right)$$

with symmetric positive definite learning rate $\mathbf{\Gamma}$ and MMQ modification term $\mathbf{M}$ which is designed, based on equation (5.1). We compute

$$\mathbf{c}_i(s)=G_i(s)\left(\dot{\mathbf{x}}(s)-\mathbf{A}\mathbf{x}(s)-\mathbf{B}_M\mathbf{r}(s)+\mathbf{B}\left[\hat{\mathbf{\Lambda}}\mathbf{u}\right](s)\right)$$

with stable linear filters $G_i(s)$ which, equals the filtered version of the true parameters $\mathbf{c}_i(s)=G_i(s)\mathbf{B}\mathbf{\Lambda}\mathbf{u}(s)$. Moreover, the filtered estimated uncertainty is

$$\hat{\mathbf{c}}_i(t)=\mathbf{B}\hat{\mathbf{\Lambda}}(t)\mathbf{q}_i(t)$$

with the filtered regressor $\mathbf{q}_i(s)=G_i(s)\mathbf{u}(s)$. Using N filters, the modification terms is

$$\mathbf{M}=\kappa\mathbf{B}^T\tilde{\mathbf{C}}\mathbf{Q}^T$$

where $\tilde{\mathbf{C}}=\left[\tilde{\mathbf{c}}_1\ \cdots\ \tilde{\mathbf{c}}_N\right]$, $\tilde{\mathbf{c}}_i=\hat{\mathbf{c}}_i-\mathbf{c}_i$, $\mathbf{Q}=\left[\mathbf{q}_1\ \cdots\ \mathbf{q}_N\right]$. lists the implemented controller parameters and Figure 5.1 shows a block diagram of the control system setup.

Table 5.1 Controller Parameters MMQ Modification Example

Lyapunov equation	$\mathbf{P}_E=\begin{bmatrix}0.084781 & 0.1716 & -1.3589 & 0.67354\\ 0.1716 & 8.6537 & -3.341 & 18.093\\ -1.3589 & -3.341 & 43.344 & -12.235\\ 0.67354 & 18.093 & -12.235 & 104.85\end{bmatrix}\cdot 10^{-3}$ $\mathbf{Q}_E=\text{diag}\left[0.04\quad 32.828\quad 32.828\quad 131.31\right]$	
learning rate	$\mathbf{\Gamma}=\text{diag}\left[0.29991\quad 0.02278\right]$	
MMQ modification	number of filters	$N=3$
	filter 1	$G_1=\dfrac{\omega_0}{s+\omega_0}$
	filter 2	$G_2(s)=\dfrac{2\zeta\omega_1 s}{s^2+2\zeta\omega_1+\omega_1^2}$
	filter 3	$G_3(s)=\dfrac{2\zeta\omega_2 s}{s^2+2\zeta\omega_2+\omega_2^2}$
	filter parameter	$\omega_0=61.607\,{}^{rad}/_{s}$, $\omega_1=4.8061\,{}^{rad}/_{s}$, $\omega_2=11.248\,{}^{rad}/_{s}$ $\zeta=0.43807$
	MMQ gain	$\kappa=0.3$

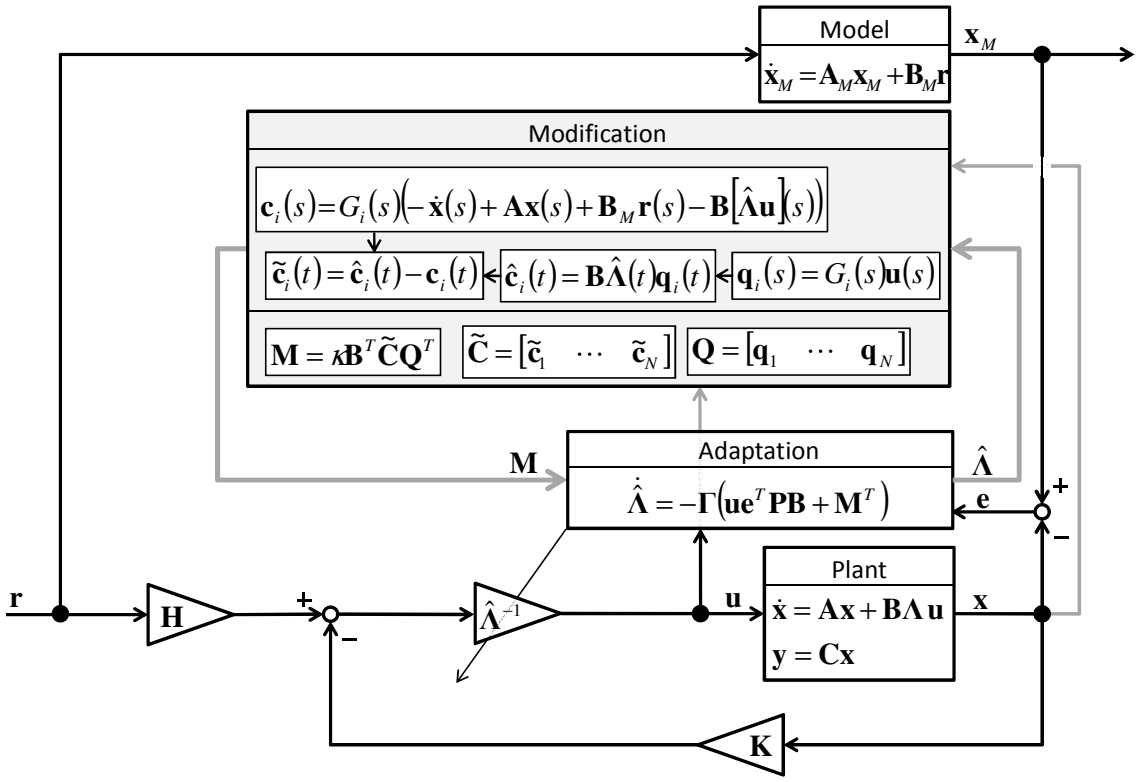


Figure 5.1 MMQ Modification Example

Three simulation runs with a length of $60s$ are presented. The first one shows adaptation performance with the pure Lyapunov based update law without any modification, while the second and third runs show performance with single model q (only filter 1) and MMQ modification (all 3 filters) respectively. In all simulation scenarios, the effectiveness of the aileron reduces to 50% of the nominal value after $5s$, i.e. $\Lambda = \mathrm{diag}[0.5 \quad 1]$. Since the true parameters are known, we are in the fortune situation to know the value of the Lyapunov function and its time derivative. Figure 5.2 shows the Lyapunov function and its derivative, if merely the standard update law is used. The Lyapunov function needs about $20s$ to reduce to zero after degradation of aileron effectiveness. Figure 5.3 shows true and estimated control effectiveness, which needs about $20s$ to adapt to the new situation, too. Notice that some algorithm has to be implemented that prevents $\hat{\Lambda}$ from becoming singular, which is classically done by restricting the estimated control effectiveness to diagonal matrices with positive diagonal entries. In this case, the SVD update algorithm, introduced in section 4.7, has been used to overcome the diagonal property constraint (Figure 5.3). Figure 5.4 shows the plant command and the effective control surface deflection, which differs from the commanded values due to the degraded Λ. Finally, Figure 5.5 shows tracking performance of the lateral states, which confirms a duration of about $20s$ until the system has adapted.

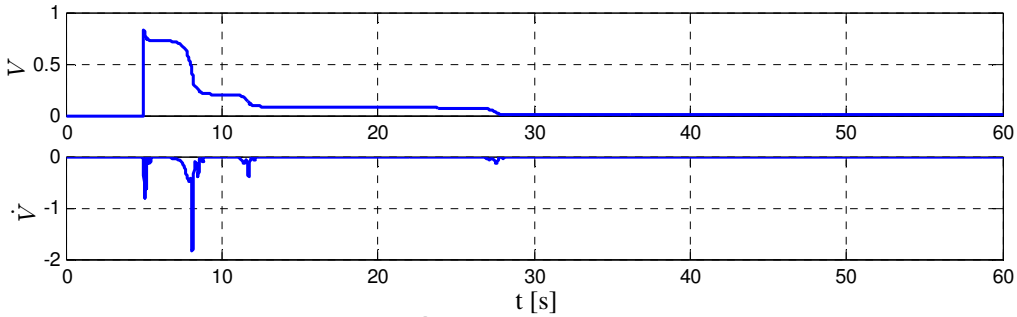


Figure 5.2 Standard Update Law – Lyapunov

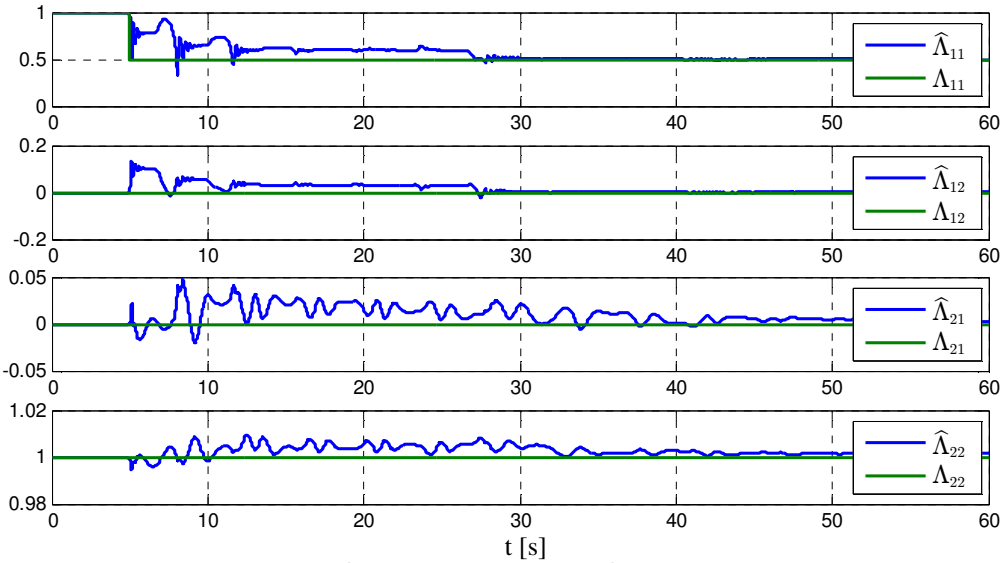


Figure 5.3 Standard Update Law – Control Effectiveness

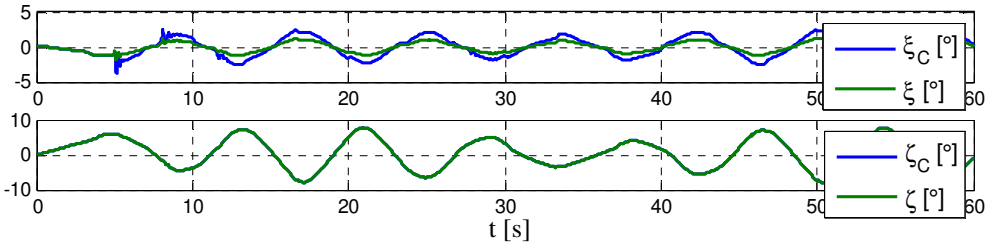


Figure 5.4 Standard Update Law – Actuator Command and Actuator Position

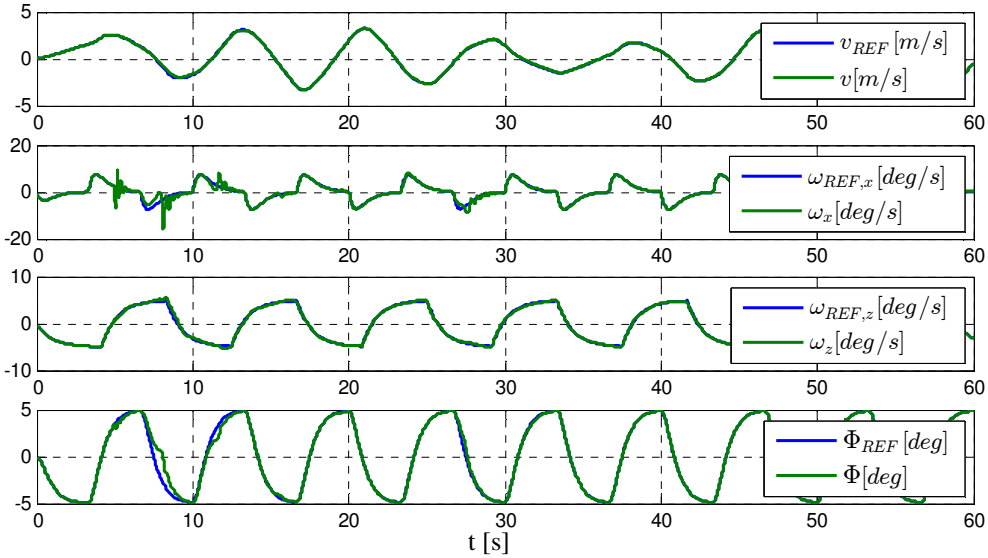


Figure 5.5 Standard Update Law – Tracking Performance

If single model q modification is applied, the time that is needed for adaptation reduces to about $5s$, as can be seen in Figure 5.6, Figure 5.7 and Figure 5.8. But, if MMQ modification is used, adaptation performance is tremendously improved further and is finished after less than $1s$, which is specifically seen in the Lyapunov plot in Figure 5.9. Figure 5.11 also shows that a deterioration of tracking performance is almost not noticeable.

Admittedly, the presented example does not contain any unmatched uncertainty, which makes MMQ modification such effective. However, if unmatched uncertainty is present, then classical q as well as MMQ modification are rather counterproductive as has already been stated in section 4.4.

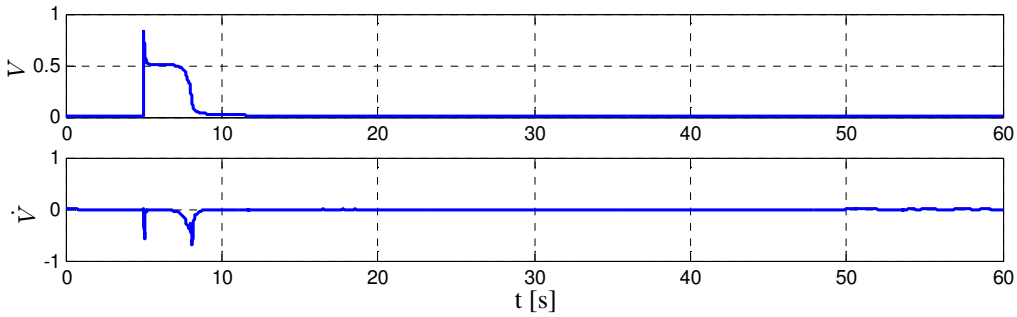


Figure 5.6 Single Model q Modification – Lyapunov

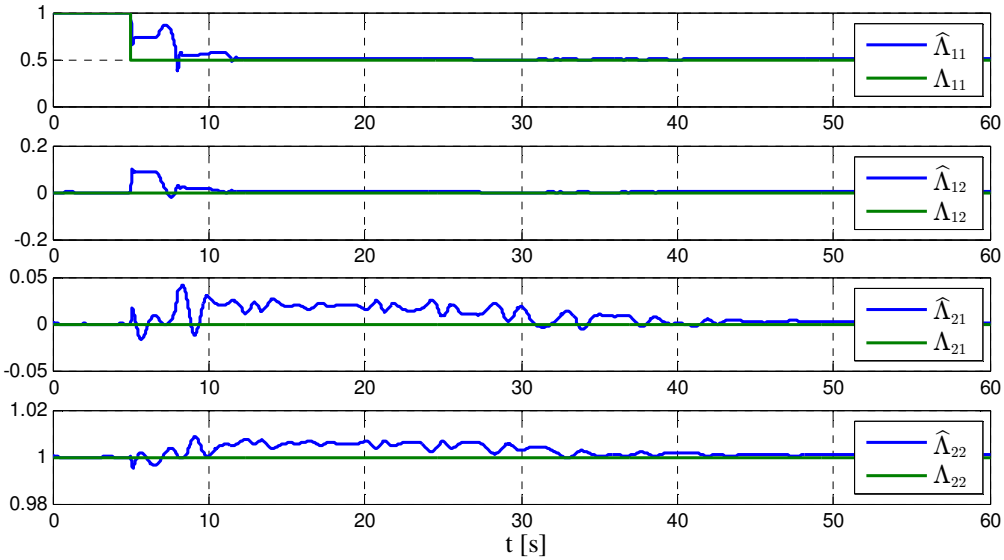


Figure 5.7 Single Model q Modification – Control Effectiveness

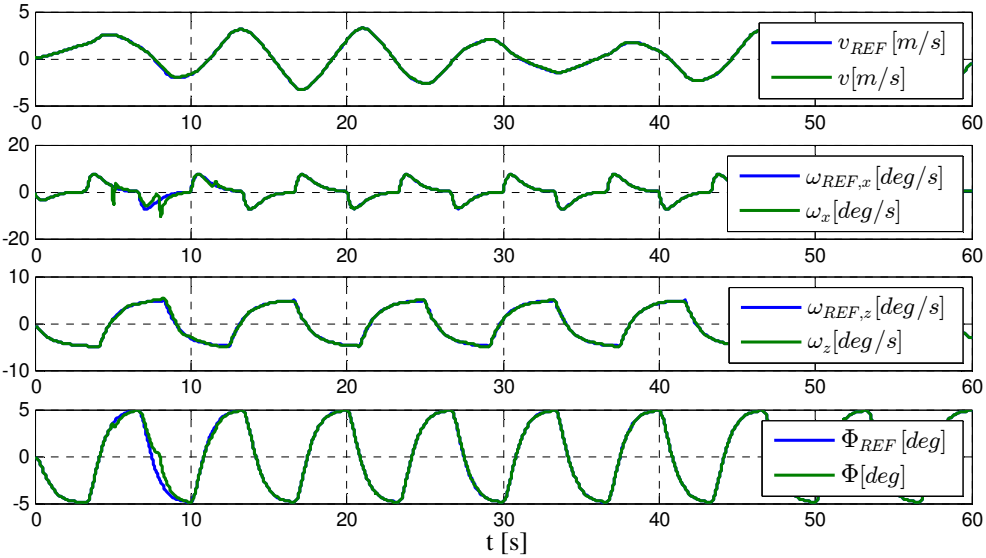


Figure 5.8 Single Model q Modification – Tracking Performance

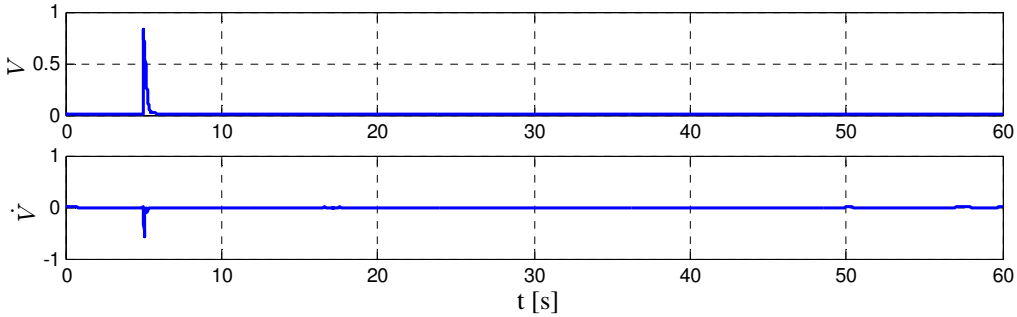


Figure 5.9 MMQ Modification – Lyapunov

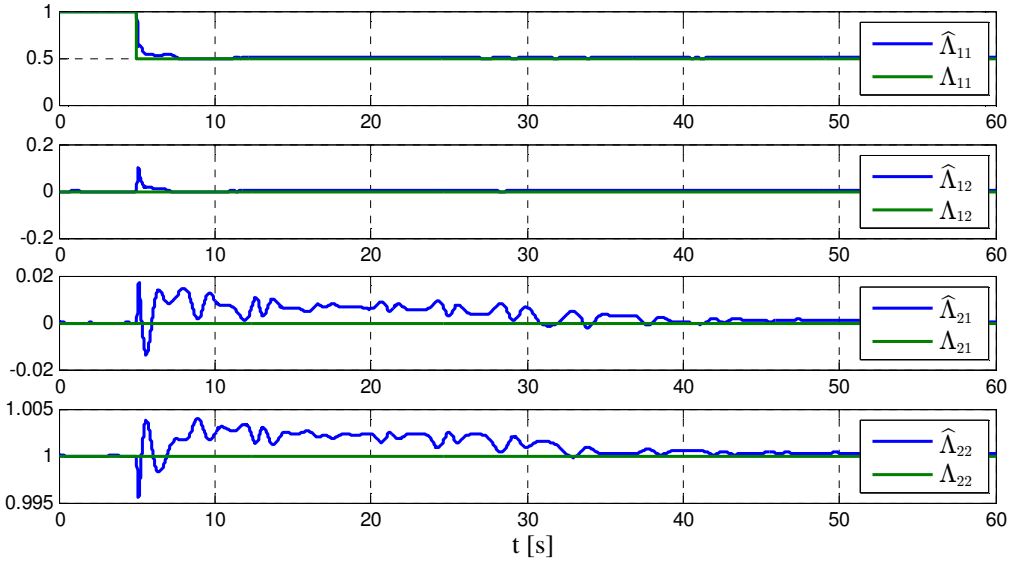


Figure 5.10 MMQ Modification – Control Effectiveness

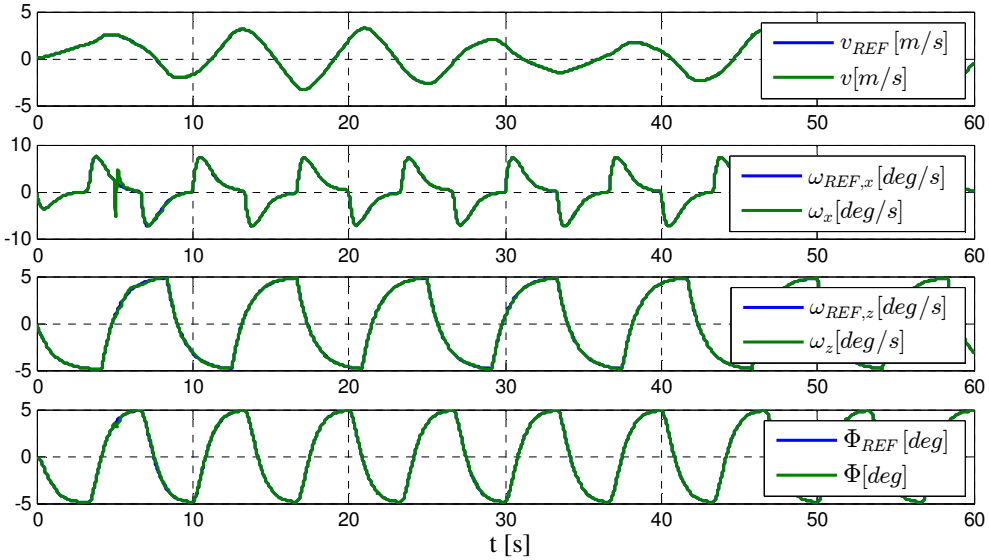


Figure 5.11 MMQ Modification – Tracking Performance

5.2 SVD Update Algorithm

The advantage of updating the control effectiveness matrix in terms of its singular value decomposition (section 4.7) is the fact, that it allows excluding singular matrices without the diagonal property constraint that is commonly used instead.

The control system setup of the following example is the same as in section 5.1 but without MMQ modification. In order to testify advantage of the proposed concept, the control effectiveness degrades to a nondiagonal matrix at $5s$.

$$\mathbf{\Lambda} = \begin{bmatrix} 0.9848 & -0.1736 \\ 0.1736 & 0.9848 \end{bmatrix}$$

Hence, effectiveness of both controls is not merely scaled down, but also changes its effective direction. In the first simulation run, the control effectiveness is tried to be updated by means of classical integration with diagonal constraint. Figure 5.12 reveals that tracking performance significantly deteriorates at $5s$ and Figure 5.13 shows that the reason therefore is that the off-diagonal of the estimated control effectiveness elements cannot converge to the true values due to limitation.

In simulation run 2 the SVD update algorithm is implemented instead. Table 5.2 lists the parameters that are used for the simulation example.

Table 5.2 Parameters of SVD Update Algorithm

numeric correction gains for enforcement of orthogonality (equation (4.396))	$k_V = k_U = 1$
minimum distance of singular values to be considered equal (equation (4.394))	$\bar{\varepsilon} = 0.001$
lower bound on singular values	$\underline{\sigma} = 0.1$

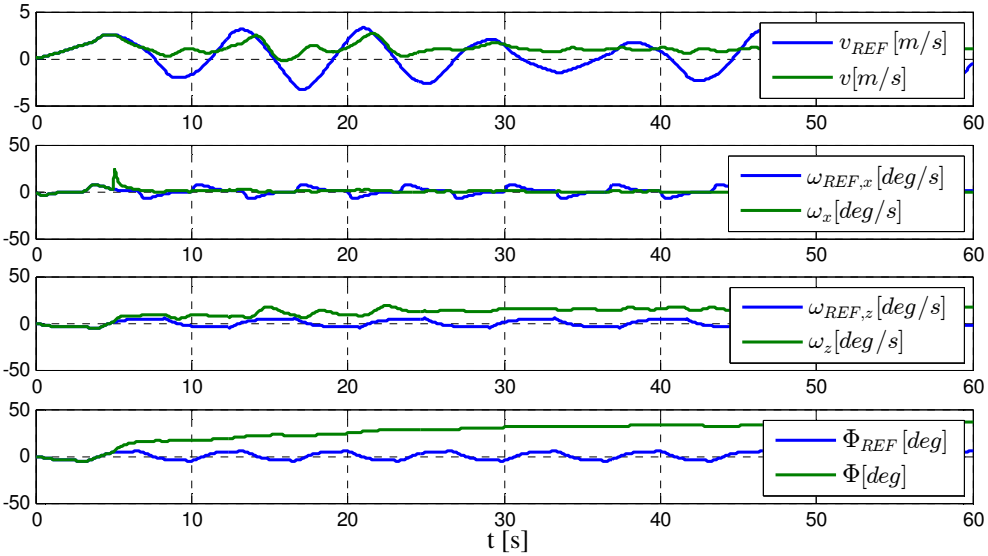


Figure 5.12 Adaptation with Diagonal Constraint – Tracking Performance

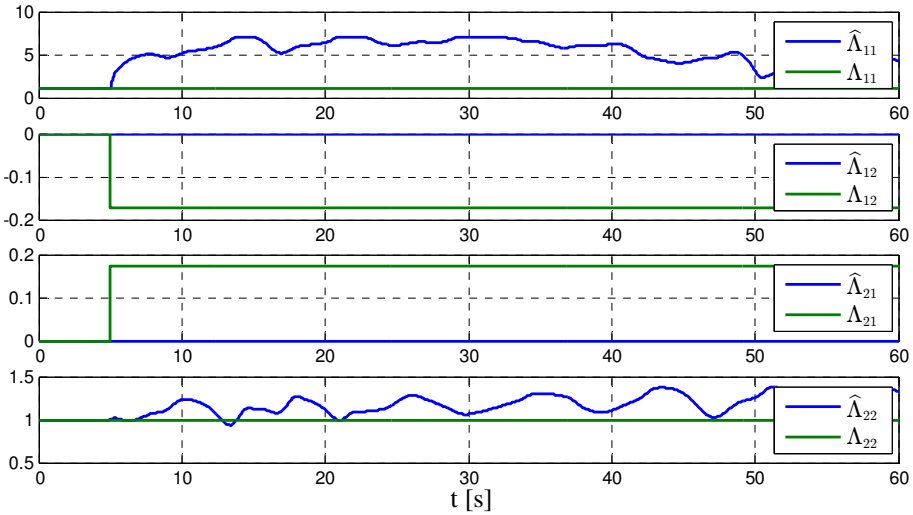


Figure 5.13 Adaptation with Diagonal Constraint – Control Effectiveness

Figure 5.14 reveals that tracking performance is strongly improved compared to the first simulation run. Of course, the reason therefore is that the estimated control effectiveness is allowed to converge to the true value as is depicted in Figure 5.15. Beside the 4 entries of the control effectiveness matrix, the 5th and 6th diagrams of Figure 5.15 show singular values of $\hat{\mathbf{\Lambda}}$ and the reset signal respectively that is triggered by the SVD algorithm if input and output directions have to be rotated due to equal singular values. Obviously, no reset is necessary, thanks to the small value that is chosen for $\bar{\varepsilon}$.

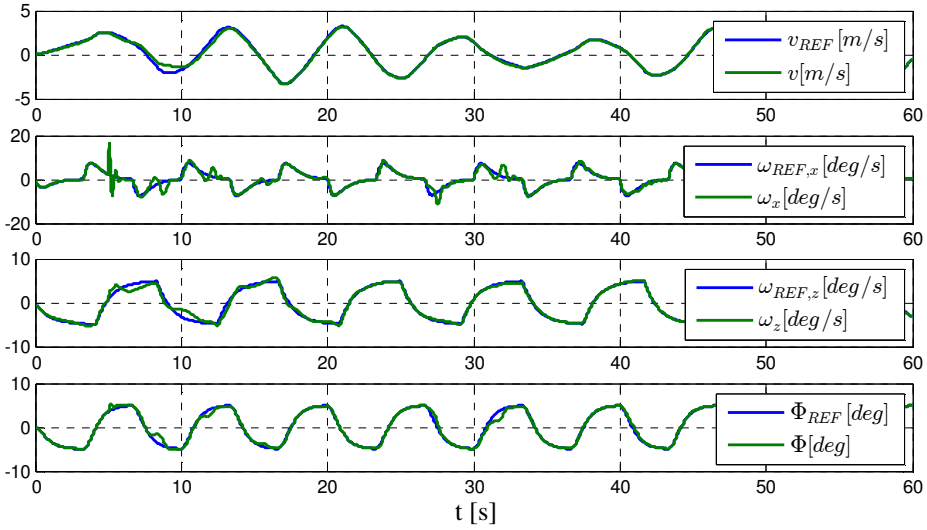


Figure 5.14 Adaptation with SVD Update Algorithm – Tracking Performance

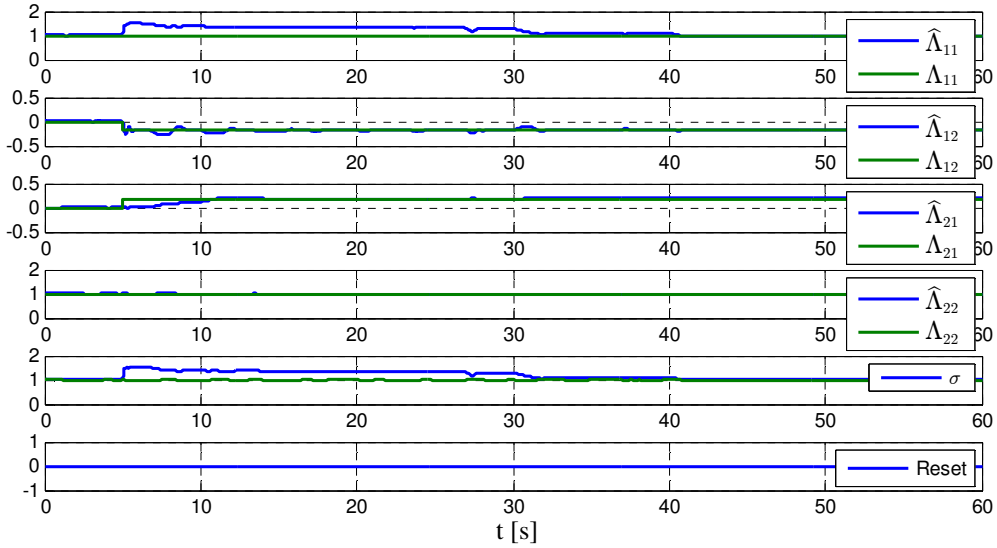


Figure 5.15 Adaptation with SVD Update Algorithm – Control Effectiveness

5.3 Application of Variant 1

5.3.1 Basic Equations

In Chapter 2, differential equations have been derived as they are used for the high fidelity simulation model. It is desirable for simulation to represent the real dynamics as realistic as possible since it is used as testbed for the developed flight control system. However, many of the considered effects have a minor influence and, as the equation have to be inverted for linearizing state feedback, it is desirable to have manageable terms as basis for the control algorithms. It is therefore appropriate to neglect parts of secondary importance.

Simplified Translation and Rotation Dynamics

For controller design, we use simplified dynamic equations, which are subjected to the following assumptions.

- the earth is flat and non-rotating, i.e.
  - no earth rotation: $\left(\vec{\boldsymbol{\omega}}^{IE}\right) \equiv \mathbf{0}$
  - no transport rate: $\left(\vec{\boldsymbol{\omega}}^{EO}\right) = \mathbf{0}$
  - which implies: $\left(\vec{\boldsymbol{\omega}}^{IB}\right) = \left(\vec{\boldsymbol{\omega}}^{OB}\right)$
  - ECEF frame is considered inertial, and also the NED is nonrotating, therefore
    - $\left(\vec{\mathbf{V}}_K^R\right)^I = \left(\vec{\mathbf{V}}_K^R\right)^E = \left(\vec{\mathbf{V}}_K^R\right)^O$
    - $\left(\dot{\vec{\mathbf{V}}}_K^R\right)^{II} = \left(\dot{\vec{\mathbf{V}}}_K^R\right)^{EO} = \left(\dot{\vec{\mathbf{V}}}_K^R\right)^{OO}$
- the dynamics are formulated w.r.t. center of gravity:
  - $\left(\vec{\mathbf{r}}^{RG}\right) = \mathbf{0}$
  - $\mathbf{I}^R = \mathbf{I}^G$
- wind is neglected
  - kinematic velocity equals aerodynamic velocity: $\left(\vec{\mathbf{V}}_K^G\right)^O = \left(\vec{\mathbf{V}}_A^G\right)^O =: \left(\vec{\mathbf{V}}^G\right)^O$
  - kinematic AoA equals aerodynamic AoA: $\alpha_K = \alpha_A =: \alpha$
  - kinematic AoS equals aerodynamic AoS: $\beta_K = \beta_A =: \beta$
  - kinematic bank angle equals aerodynamic bank angle: $\mu_K = \mu_A =: \mu$

A direct consequence of neglect of wind is that the $\overline{K}$-frame (the kinematic frame, rotated about the kinematic bank angle μ_K, refer to Appendix A) equals the aerodynamic frame (Index A). The shift of the reference point to the c.g. is in fact no simplifying assumption, though it essentially reduces the complexity of the translation and rotation equations (2.22), which can be solved decoupled of each other.

$$\left(\dot{\vec{\mathbf{V}}}^G\right)_B^{OB} + \left(\vec{\boldsymbol{\omega}}^{OB}\right)_B \times \left(\vec{\mathbf{V}}_K^G\right)_B^O = \frac{\sum \left(\vec{\mathbf{F}}^G\right)_B}{m} \tag{5.2}$$

$$\left(\dot{\vec{\boldsymbol{\omega}}}^{OB}\right)_B^B = \left(\mathbf{I}^G\right)_B^{-1} \cdot \left\{\sum \left(\vec{\mathbf{M}}^G\right)_B - \left(\vec{\boldsymbol{\omega}}^{OB}\right)_B \times \left[\left(\mathbf{I}^G\right)_B \cdot \left(\vec{\boldsymbol{\omega}}^{OB}\right)_B\right]\right\} \tag{5.3}$$

Choice of Internal states

The control theory, developed in Chapter 4, is applied to a rate command system and hence the angular momentum equations, which are excited by the external moments, form the external dynamics. However, as the aerodynamic moments also depend on aerodynamic flow angles and velocity, these states also have to be considered and are part of the internal dynamics. The dynamics of velocity and flow angles, in turn, depend on the external forces, which contain the gravity force whose direction relative

to the aircraft depends on the attitude for which reason the attitude is part of the internal dynamics, too.

In order to obtain the differential equation for velocity V, angle of attack α and angle of side-slip β we rewrite the time derivative of (5.2) in terms of the $\overline{K}$.

$$\left(\dot{\vec{\mathbf{V}}}^{G}\right)_{\overline{K}}^{OB} = \left(\dot{\vec{\mathbf{V}}}^{G}\right)_{\overline{K}}^{O\overline{K}} + \left(\vec{\boldsymbol{\omega}}^{B\overline{K}}\right)_{\overline{K}} \times \left(\vec{\mathbf{V}}^{G}\right)_{\overline{K}}^{O} \tag{5.4}$$

Since the x-axis of $\overline{K}$ points in the direction of the velocity per definition, we obviously have

$$\left(\vec{\mathbf{V}}^{G}\right)_{\overline{K}}^{O} = \begin{pmatrix} V \\ 0 \\ 0 \end{pmatrix} \text{ and } \left(\dot{\vec{\mathbf{V}}}^{G}\right)_{\overline{K}}^{O\overline{K}} = \begin{pmatrix} \dot{V} \\ 0 \\ 0 \end{pmatrix}.$$

Further, by kinematic considerations and definition of α, β ,we get

$$\left(\vec{\boldsymbol{\omega}}^{B\overline{K}}\right)_{\overline{K}} = \begin{pmatrix} -\dot{\alpha}\sin\beta \\ -\dot{\alpha}\cos\beta \\ \dot{\beta} \end{pmatrix}$$

The cross product term in (5.4) adopts a special form.

$$\begin{aligned} \mathbf{M}_{\overline{K}B}\left(\vec{\boldsymbol{\omega}}^{OB}\right)_{B} \times \left(\vec{\mathbf{V}}^{G}\right)_{\overline{K}}^{O} &= \begin{bmatrix} \cos\alpha\cos\beta & \sin\beta & \sin\alpha\cos\beta \\ -\cos\alpha\sin\beta & \cos\beta & -\sin\alpha\sin\beta \\ -\sin\alpha & 0 & \cos\alpha \end{bmatrix} \begin{pmatrix} p \\ q \\ r \end{pmatrix} \times \begin{pmatrix} V \\ 0 \\ 0 \end{pmatrix} \\ &= V \begin{pmatrix} 0 \\ r\cos\alpha - p\sin\alpha \\ p\sin\beta\cos\alpha - q\cos\beta + r\sin\alpha\sin\beta \end{pmatrix} \end{aligned}$$

Hence (5.2), written in kinematic frame and solved for the state derivatives reads as

$$\begin{aligned} \dot{V} &= \frac{\left(X^{G}\right)_{\overline{K}}}{m} \\ \dot{\alpha} &= \frac{\left(Z^{G}\right)_{\overline{K}}}{mV\cos\beta} + q - \tan\beta \cdot \left[p\cos\alpha + r\sin\alpha\right] \\ \dot{\beta} &= \frac{\left(Y^{G}\right)_{\overline{K}}}{mV} \qquad + \qquad p\sin\alpha - r\cos\alpha. \end{aligned} \tag{5.5}$$

For description of the attitude, we use Euler angles, whose dynamics are given by (2.23). Notice that the azimuth dynamics are decoupled and hence, are not necessarily part of the internal dynamics.

$$\begin{aligned} \dot{\Phi} &= \quad p \quad + \tan\Theta\sin\Phi\, q \quad + \tan\Theta\cos\Phi\, r \\ \dot{\Theta} &= \qquad\qquad \cos\Phi\, q \qquad - \sin\Phi\, r \end{aligned} \tag{5.6}$$

5.3.2 Radial Basis Function Neural Networks

The dependence of propulsive forces and moments on states and controls are inherently nonlinear and thrust vectoring introduces additional nonlinearities due to trigonometric functions of the thrust vectoring angles. A simple parameterization, using linear regressors in states and control is thus not appropriate.

In adaptive control, arbitrary continuous nonlinear functions are commonly modeled by neutral networks ([Joh01], [Joh04], [Joh00c], [Shi05]), which allow an arbitrary close approximation of the considered functions over a compact domain, provided the number of neurons is sufficiently large ([Fun89]).

Two types of neural networks are commonly utilized in MRAC, single hidden layer neural networks (SHL), using sigmoid activation functions and radial basis function neural networks (RBF). Both types have their advantages for adaptive control applications as discussed in [And09]. While the SHL requires quite complex update laws because it is nonlinear in its parameters, the RBF is linearly parameterized such that a standard update of the weights, as developed in Chapter 4, results in a stable closed loop system. Hence, the RBF approach is chosen for adaptive approximation of nonlinear functions. The radial basis functions are of the form

$$\phi(\mathbf{x}) = e^{-(\mathbf{x}-\mathbf{x}_0)^T \mathbf{Q}^{-1}(\mathbf{x}-\mathbf{x}_0)} \tag{5.7}$$

where $\mathbf{x} \in \mathbb{R}^q$ is a vector of independent variables, $\mathbf{x}_0$ is the center and $\mathbf{Q}$ is a symmetric positive definite matrix that scales the width of the function in the principle axes of $\mathbf{Q}$. Figure 5.16 shows a radial basis function with center in the origin and $\mathbf{Q} = \mathrm{diag}(5 \quad 2)$. Note that it achieves a maximum of 1 in the center while it decays to e^{-1} in distance $\sqrt{5}$ from the center in x_1 direction and in distance $\sqrt{2}$ from the center in x_2 direction. For a general positive symmetric $\mathbf{Q}$, the principle axes are the orthogonal eigenvectors and the scaling factors are the associated eigenvalues.

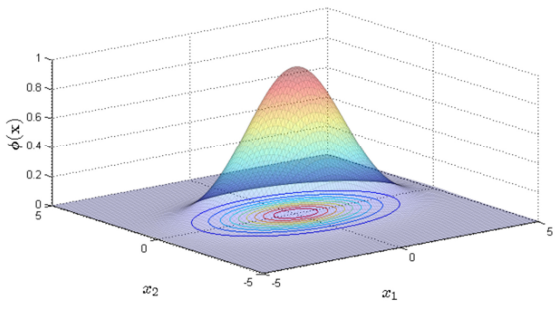


Figure 5.16 Radial Basis Function

Particularly the nonlinear-in-control design requires for the partial derivatives of the radial basis function w.r.t. its independent variables; the partial derivative of (5.7) w.r.t its independent variable is given by

$$\frac{\partial \phi(\mathbf{x})}{\partial \mathbf{x}} = -(\mathbf{x}-\mathbf{x}_0)^T \mathbf{Q}^{-1} e^{-(\mathbf{x}-\mathbf{x}_0)^T \mathbf{Q}^{-1}(\mathbf{x}-\mathbf{x}_0)}. \tag{5.8}$$

For the RBF, the centers of the regressors are equally spaced over the valid domain for $\mathbf{x}$ as depicted in Figure 5.17, where 3 regressors were placed in each dimension.

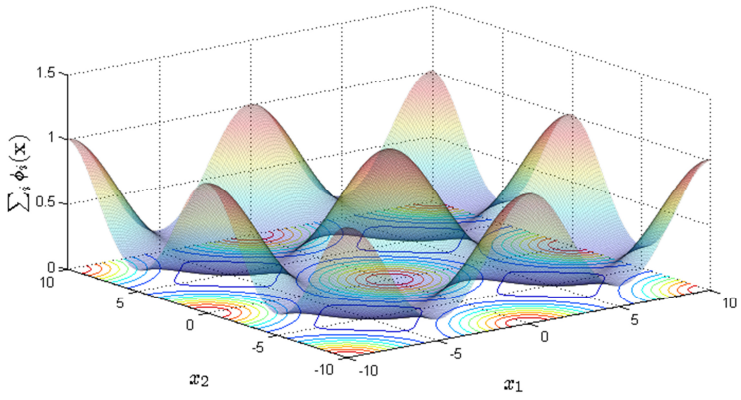


Figure 5.17 Radial Basis Functions with Equally Spaced Centers

The neural network is composed by stacking all regressors into a vector $\boldsymbol{\varphi}(\mathbf{x})$, $i = 1, \cdots, s$ and multiplication with a parameter matrix $\boldsymbol{\Theta} \in \mathbb{R}^{s \times m}$

$$\mathbf{y} = \boldsymbol{\Theta}^T \boldsymbol{\varphi}(\mathbf{x}) \tag{5.9}$$

where m is the dimension of the output space. Figure 5.18 shows a schematic of a RBF. While $\boldsymbol{\Theta}$ adopts the role of an adaptive parameter within the MRAC scheme, the weighing as well as the center locations are fixed a priori, but they leave room for further optimization. If some knowledge about the structure of the nonlinearity is available, one could use narrow functions with dense spacing in regions of high curvature and wide functions, placed with low density in regions of low curvature.

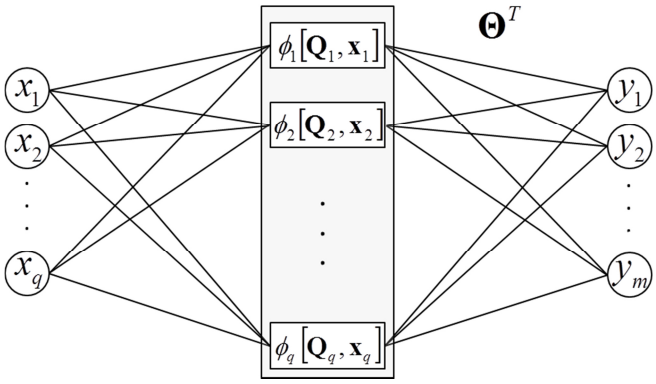


Figure 5.18 Radial Basis Function Neural Network

5.3.3 Parameterizations of Forces and Moments

The external forces and moments within the simulation model are composed of an aerodynamic, propulsive and a gravitational part, according to (2.33), (2.34) and comprise a high level of detail with highly nonlinear characteristic. For the controller, a simplified model will be used as described in the following.

Aerodynamics

The moment coefficients are modeled linearly in states and controls.

$$\begin{aligned}
C_l &= C_{l0} + C_{l\alpha}\alpha + C_{l\beta}\beta + C_{lp}p^* + C_{lp}q^* + C_{lp}r^* \\
&\quad + C_{l\eta_{c,l}}\eta_{c,l} + C_{l\eta_{c,r}}\eta_{c,r} + C_{l\xi_l}\xi_l + C_{l\xi_r}\xi_r + C_{l\delta_l}\delta_l + C_{l\delta_r}\delta_r + C_{l\zeta}\zeta + \delta_{AL}(V,\alpha,\beta,p,q,r,\eta_{c,l},\ldots) \\
C_m &= C_{m0} + C_{m\alpha}\alpha + C_{m\beta}\beta + C_{mp}p^* + C_{mp}q^* + C_{mp}r^* \\
&\quad + C_{m\eta_{c,l}}\eta_{c,l} + C_{m\eta_{c,r}}\eta_{c,r} + C_{m\xi_l}\xi_l + C_{m\xi_r}\xi_r + C_{m\delta_l}\delta_l + C_{m\delta_r}\delta_r + C_{m\zeta}\zeta + \delta_{AM}(V,\alpha,\beta,p,q,r,\eta_{c,l},\ldots) \\
C_n &= C_{n0} + C_{n\alpha}\alpha + C_{n\beta}\beta + C_{np}p^* + C_{np}q^* + C_{np}r^* \\
&\quad + C_{n\eta_{c,l}}\eta_{c,l} + C_{n\eta_{c,r}}\eta_{c,r} + C_{n\xi_l}\xi_l + C_{n\xi_r}\xi_r + C_{n\delta_l}\delta_l + C_{n\delta_r}\delta_r + C_{n\zeta}\zeta + \delta_{AN}(V,\alpha,\beta,p,q,r,\eta_{c,l},\ldots)
\end{aligned} \tag{5.10}$$

where $\delta_{AL}(\cdot)$, $\delta_{AM}(.)$, $\delta_{AN}(\cdot)$ denote the unmatched uncertainties and

$$p^* = \frac{bp}{2V} \quad , \quad q^* = \frac{\bar{c}q}{2V} \quad , \quad r^* = \frac{br}{2V}$$

are the nondimensional angular rates where b and $\bar{c}$ are half wing span and mean aerodynamics chord respectively. Finally, aerodynamic forces and moments in body-fixed frame are

$$\left(\vec{\mathbf{F}}_A^G\right)_B = \mathbf{M}_{B\bar{K}}(\alpha,\beta)\begin{pmatrix}\left(X_A^G\right)_{\bar{K}}\\ \left(Y_A^G\right)_{\bar{K}}\\ \left(Z_A^G\right)_{\bar{K}}\end{pmatrix} = \mathbf{M}_{B\bar{K}}(\alpha,\beta)\begin{pmatrix}-C_D\\ C_Q\\ -C_L\end{pmatrix}\bar{q}S \qquad \left(\vec{\mathbf{M}}_A^A\right)_B = \bar{q}\cdot S\cdot\begin{pmatrix}b/2\cdot C_l\\ \bar{c}\cdot C_m\\ b/2\cdot C_n\end{pmatrix}.$$

Propulsion

For adaptation, it is advantageous to model uncertainties with a preferably small number of parameters since this increases adaptation performance. As forces and moments, produced by propellers, act mainly in x-direction of the propulsion system (Figure 2.3), the number of adaptive parameters is, at first, significantly reduced by setting the other axes to zero.

$$\left(\vec{\mathbf{F}}_{\mathrm{Pr}}^{\mathrm{Pr}}\right)_{\mathrm{Pr}} = \begin{pmatrix}\left(X_{\mathrm{Pr}}^{\mathrm{Pr}}\right)_{\mathrm{Pr}}\\ 0\\ 0\end{pmatrix} \quad , \quad \left(\vec{\mathbf{M}}_{\mathrm{Pr}}^{\mathrm{Pr}}\right)_{\mathrm{Pr}} = \begin{pmatrix}\left(L_{\mathrm{Pr}}^{\mathrm{Pr}}\right)_{\mathrm{Pr}}\\ 0\\ 0\end{pmatrix}$$

As explained in section 2.3.2, the propulsive forces and moments depend on the total AoA and AoS, which describe in fact the direction of the aerodynamic velocity relative to the respective propulsion frame. Since we only consider the x-axis direction, the dependence of forces and moments on the total aerodynamic flow angles is rotationally symmetric. In missile kinematics, the concept of *included angle of attack* is commonly used ([Nie60]). We will take advantage of this concept in a slight modified version in order to reduce the number of independent variables for the propulsive forces and moments by exploiting the rotational symmetry.

AoA and AoS describe the rotation from the body-fixed to the aerodynamic frame, by the rotation sequence y-axis – z-axis and actually determine the direction of the velocity vector relative to the body-fixed frame. The relative direction can also be parameterized by an alternative set of angles. Consider Figure 5.19, at first the body-fixed frame (respectively propulsion frame) is rotated about the x-axis until the velocity vector is aligned within the x-z-plane, where the associated angle ϕ is denoted as azimuth. Then, the resulting frame is rotated about the y-axis until the x-axis is aligned with the velocity. The associated angle α_c is denoted as *included angle of attack*.

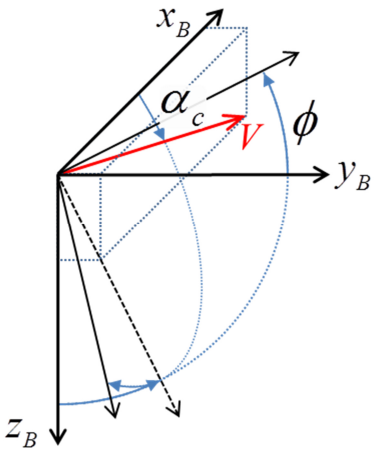


Figure 5.19 Included Angle of Attack

The relationship between total AoA, AoS and the velocity components in propulsion axes is, according to (2.30)

$$u = V\cos\beta_T\cos\alpha_T \;,\; v = V\sin\beta_T \;,\; w = V\cos\beta_T\sin\alpha_T \,.$$

where α_T, β_T denote the total AoA and AoS of the respective propulsion according to (2.46) – (2.49). On the other hand, from Figure 5.19, we derive

$$u = V\cos\alpha_c \quad , \quad v = V\sin\alpha_c\sin\phi \quad , \quad w = V\sin\alpha_c\cos\phi$$

and by comparison, finally

$$\begin{aligned} \alpha_c &= \arccos\left(\cos\beta_T\cos\alpha_T\right) \\ \phi &= \arctan\left(\frac{\tan\beta_T}{\sin\alpha_T}\right). \end{aligned} \tag{5.11}$$

Then, due to rotation symmetry, the propulsive thrust and torque are independent of the azimuth and, according to equations (2.43), (2.44) depend on included AoA α_c, velocity V and propeller rotation speed ω_M.

$$\left(X_{\mathrm{Pr}}^{\mathrm{Pr}}\right)_{\mathrm{Pr}} = \bar{f}_X\left(\alpha_c, V, \omega_M\right) \quad , \quad \left(L_{\mathrm{Pr}}^{\mathrm{Pr}}\right)_{\mathrm{Pr}} = \bar{f}_L\left(\alpha_c, V, \omega_M\right) \tag{5.12}$$

It is important to notice that the equation for included AoA (5.11) is free of singularities, while the azimuth is not. However, this does not cause any problem since the azimuth is not a dependent variable.

The motor rotation speed cannot be measured and therefore we would rather like to have thrust and torque dependent on the input voltage, which could be measured easily or is at least known, as it is an exogenous input. The propeller is powered by a brushless DC motor and, according to equation (2.60), the steady state motor rotation speed is related to the input voltage and the load moment, which equals the propeller moment. Consequently, by equating (2.60) and (5.11), we obtain

$$\frac{K_e}{R_T}U_0 - \frac{K_E^2 + R_T K_L}{R_T}\omega_M = \bar{f}_L(\alpha_c, V, \omega_M)$$

and, for fixed U_0, α_c, V, the equation can be uniquely solved for ω_M if $\bar{f}_L(\cdot)$ is monotonic in ω_M, which is the case in normal operation conditions, since we could expect that the torque is monotonic increasing with rotation speed. Let the solution of the above equation be

$$\omega_M = g(\alpha_c, V, U_0) \tag{5.13}$$

then, by insertion of (5.13) into (5.12), thrust and torque are dependent on U_0.

$$\left(X_{\mathrm{Pr}}^{\mathrm{Pr}}\right)_{\mathrm{Pr}} = \bar{f}_X(\alpha_c, V, g(\alpha_c, V, U_o)) =: f_X(\alpha_c, V, U_o) \tag{5.14}$$

$$\left(L_{\mathrm{Pr}}^{\mathrm{Pr}}\right)_{\mathrm{Pr}} = \bar{f}_L(\alpha_c, V, g(V, \alpha_c, U_0)) =: f_L(\alpha_c, V, U_0) \tag{5.15}$$

Figure 5.20 – Figure 5.27 show results of numerical computations of thrust and torque characteristic for a single main and back engine respectively. These nonlinear functions are approximated by RBF as introduced in section 5.3.2. For each engine we choose

$$\begin{aligned}\left(X_{\mathrm{Pr}_i}^{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} &= \mathbf{p}_{X,i}^T \boldsymbol{\varphi}_{X,i}(V, \alpha_{c,i}, U_{0,i}) + \delta_{PX,i}(V, \alpha_{c,i}, U_{o,i}) \\ \left(L_{\mathrm{Pr}_i}^{\mathrm{Pr}_i}\right)_{\mathrm{Pr}_i} &= \mathbf{p}_{L,i}^T \boldsymbol{\varphi}_{L,i}(V, \alpha_{c,i}, U_{0,i}) + \delta_{PL,i}(V, \alpha_{c,i}, U_{o,i})\end{aligned} \tag{5.16}$$

where i is replaced by l (left engine), r (right engine) or b (back engine), the included AoAs are computed for the respective engines according to (2.46) – (2.49) and (5.11) and $\delta_{PX,i}(\cdot)$ and $\delta_{PL,i}(\cdot)$ are the unmatched uncertainties. Moreover $\mathbf{p}_{X,i}$ and, $\mathbf{p}_{L,i}$ denote the parameter vectors for thrust and torque respectively and $\boldsymbol{\varphi}_{X,i}(V, \alpha_{c,i}, U_{0,i})$, $\boldsymbol{\varphi}_{L,i}(V, \alpha_{c,i}, U_{0,i})$ are regressor vectors, whose elements contain radial basis functions according to (5.7). The scalar expansion of the regressor vectors reads as

$$\boldsymbol{\varphi}_{X,i}^T(\alpha_{c,i}, V, U_{0,i}) = \left(\phi_{X,i,1}(V, \alpha_{c,i}, U_{0,i}) \quad \cdots \quad \phi_{X,i,s_{Xi}}(V, \alpha_{c,i}, U_{0,i})\right) \tag{5.17}$$

$$\boldsymbol{\varphi}_{L,i}^T(\alpha_{c,i}, V, U_{0,i}) = \left(\phi_{L,i,1}(V, \alpha_{c,i}, U_{0,i}) \quad \cdots \quad \phi_{L,i,s_{Li}}(V, \alpha_{c,i}, U_{0,i})\right) \tag{5.18}$$

where s_{Xi}, s_{Li} denote the length of the respective regressors and

$$\phi_{X,i,j} = \exp\left[\begin{pmatrix} V - V^{(X_{i,j})} & \alpha_{c,i} - \alpha_{c,i}^{(X_{i,j})} & U_{0,i} - U_{0,i}^{(X_{i,j})} \end{pmatrix} \begin{bmatrix} q_V^{(X_i)} & 0 & 0 \\ 0 & q_\alpha^{(X_i)} & 0 \\ 0 & 0 & q_U^{(X_i)} \end{bmatrix}^{-1} \begin{pmatrix} V - V^{(X_{i,j})} \\ \alpha_{c,i} - \alpha_{c,i}^{(X_{i,j})} \\ U_{0,i} - U_{0,i}^{(X_{i,j})} \end{pmatrix}\right] \tag{5.19}$$

$$\phi_{L,i,j} = \exp\left[\begin{pmatrix} V - V^{(L_{i,j})} & \alpha_{c,i} - \alpha_{c,i}^{(L_{i,j})} & U_{0,i} - U_{0,i}^{(L_{i,j})} \end{pmatrix} \begin{bmatrix} q_V^{(Li)} & 0 & 0 \\ 0 & q_\alpha^{(Li)} & 0 \\ 0 & 0 & q_U^{(Li)} \end{bmatrix}^{-1} \begin{pmatrix} V - V^{(L_{i,j})} \\ \alpha_{c,i} - \alpha_{c,i}^{(L_{i,j})} \\ U_{0,i} - U_{0,i}^{(L_{i,j})} \end{pmatrix} \right]. \quad (5.20)$$

Thereby $V^{(X_{i,j})}, V^{(L_{i,j})}, \alpha_{c,i}^{(X_{i,j})}, \alpha_{c,i}^{(L_{i,j})}, U_{0,i}^{(X_{i,j})}, U_{0,i}^{(L_{i,j})}$ denote the center location of the respective radial basis function and $q_V^{(X_i)}, q_\alpha^{(X_i)}, q_U^{(X_i)}, q_V^{(L_i)}, q_\alpha^{(L_i)}, q_U^{(L_i)}$ are the scaling factors in the respective direction. Accordingly, the scalar expansion of the parameter vectors read as

$$\mathbf{p}_{X,i}^T = \begin{pmatrix} p_{X,i,1} & \cdots & p_{X,i,s_{Xi}} \end{pmatrix}, \quad \mathbf{p}_{L,i}^T = \begin{pmatrix} p_{L,i,1} & \cdots & p_{L,i,s_{Li}} \end{pmatrix}.$$

Finally, for insertion of the parameterized forces and moments into the EOMs, the reference point has to be shifted to c.g. and the notation frame has to be transformed appropriately.

$$\begin{aligned} \left(\vec{\mathbf{F}}_P^G\right)_{\overline{K}} = \quad & \mathbf{M}_{B\,\mathrm{Pr}_l} \cdot \left(\vec{\mathbf{F}}_{\mathrm{Pr}_l}^{\mathrm{Pr}_l}\right)_{\mathrm{Pr}_l} + \mathbf{M}_{B\,\mathrm{Pr}_l} \cdot \left(\vec{\mathbf{F}}_{\mathrm{Pr}_r}^{\mathrm{Pr}_r}\right)_{\mathrm{Pr}_r} + \mathbf{M}_{B\,\mathrm{Pr}_l} \cdot \left(\vec{\mathbf{F}}_{\mathrm{Pr}_b}^{\mathrm{Pr}_b}\right)_{\mathrm{Pr}_b} \\ \left(\vec{\mathbf{M}}_P^G\right)_B = \quad & \mathbf{M}_{B\mathrm{Pr}_l} \cdot \left(\vec{\mathbf{M}}_{\mathrm{Pr}_l}^{\mathrm{Pr}_l}\right)_{\mathrm{Pr}_l} + \mathbf{M}_{B\mathrm{Pr}_r} \cdot \left(\vec{\mathbf{M}}_{\mathrm{Pr}_r}^{\mathrm{Pr}_r}\right)_{\mathrm{Pr}_r} + \mathbf{M}_{B\mathrm{Pr}_b} \cdot \left(\vec{\mathbf{M}}_{\mathrm{Pr}_b}^{\mathrm{Pr}_b}\right)_{\mathrm{Pr}_b} \\ & + \left(\vec{\mathbf{r}}^{G\mathrm{Pr}_l}\right)_B \times \left(\vec{\mathbf{F}}_{\mathrm{Pr}_l}^{\mathrm{Pr}_l}\right)_B + \left(\vec{\mathbf{r}}^{G\mathrm{Pr}_r}\right)_B \times \left(\vec{\mathbf{F}}_{\mathrm{Pr}_r}^{\mathrm{Pr}_r}\right)_B + \left(\vec{\mathbf{r}}^{G\mathrm{Pr}_b}\right)_B \times \left(\vec{\mathbf{F}}_{\mathrm{Pr}_b}^{\mathrm{Pr}_b}\right)_B \end{aligned} \quad (5.21)$$

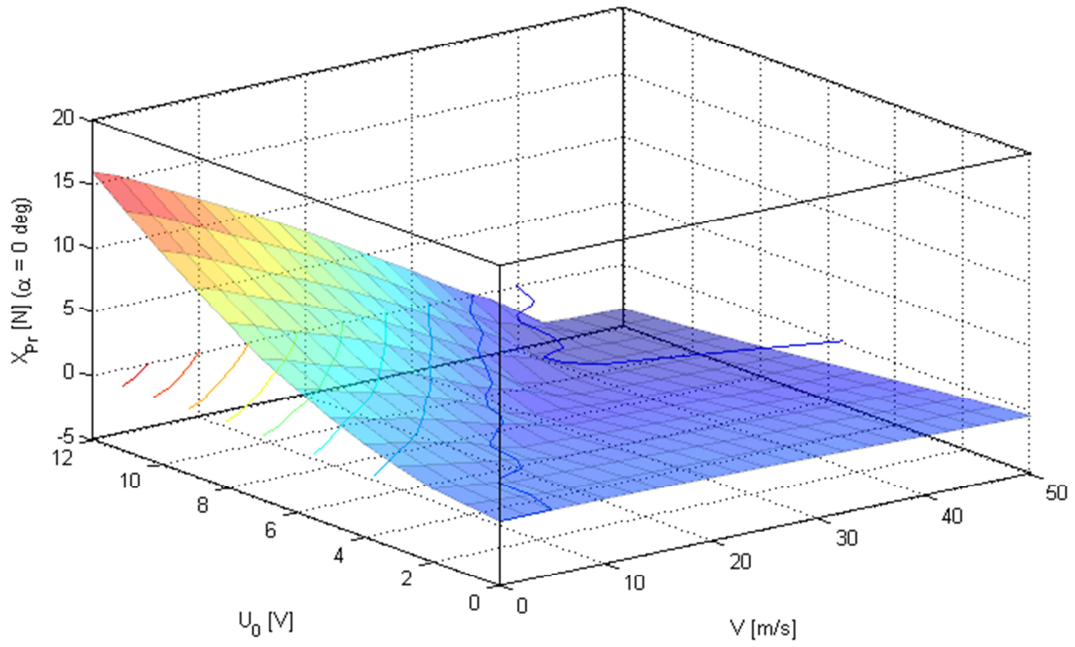


Figure 5.20 Main Engine Thrust– Input Voltage, Aerodynamic Velocity

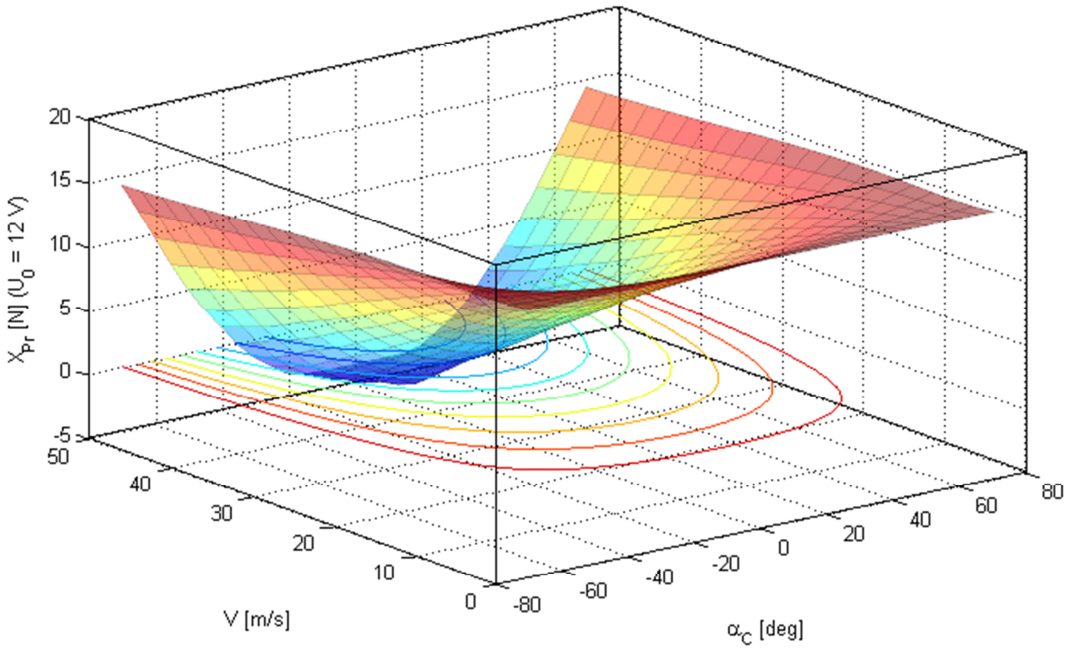


Figure 5.21 Main Engine Thrust– Aerodynamic Velocity, Included AoA

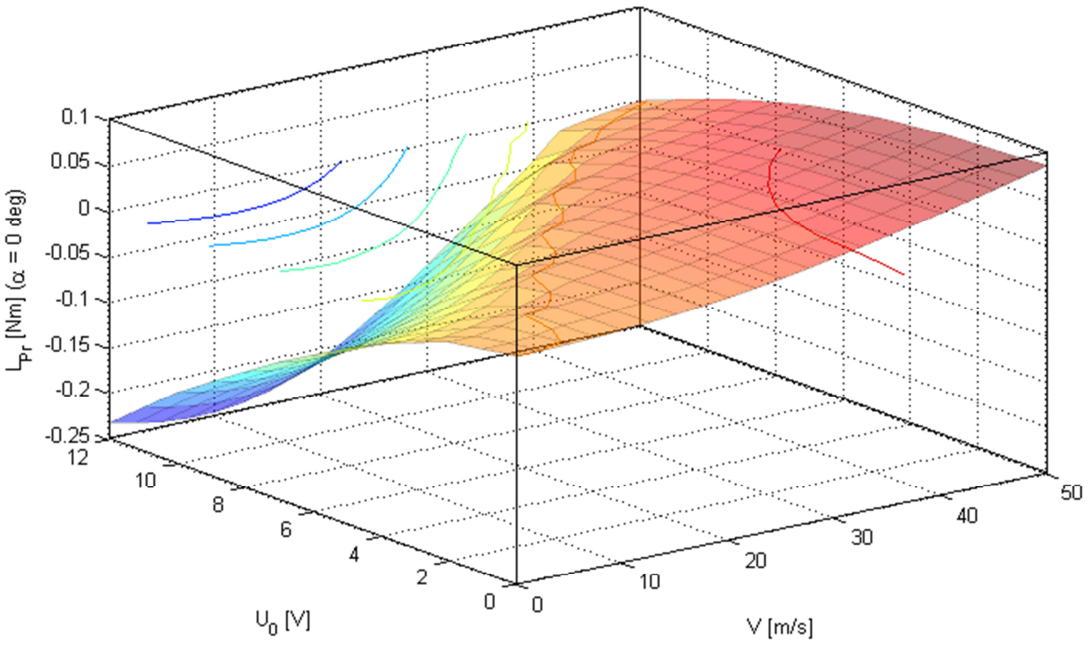


Figure 5.22 Main Engine Torque– Input Voltage, Aerodynamic Velocity

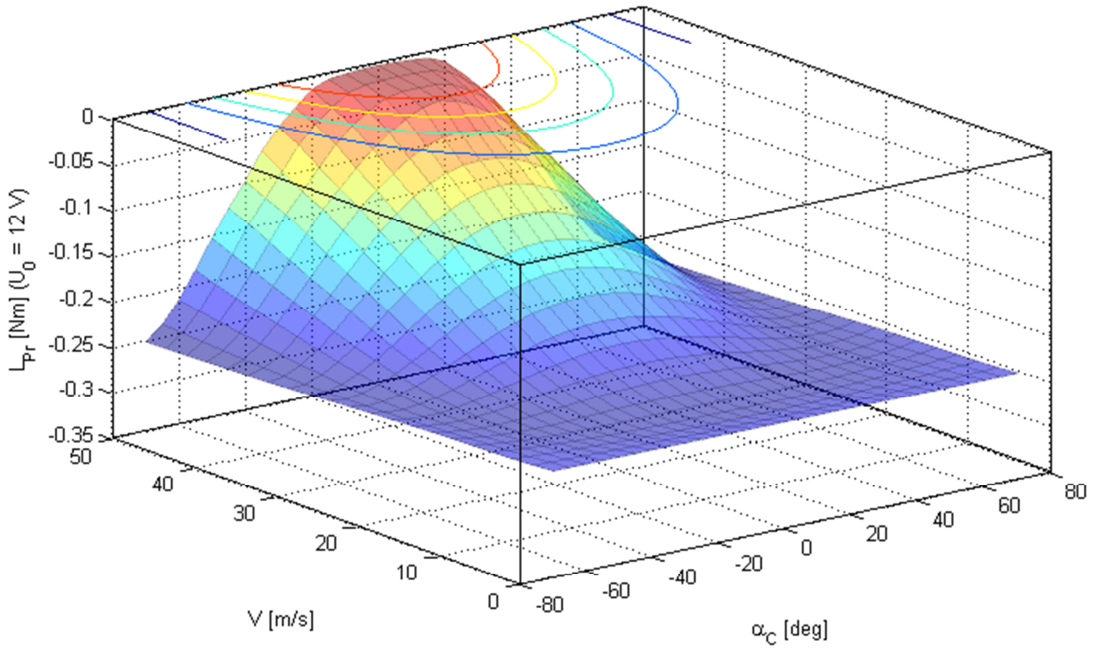


Figure 5.23 Main Engine Torque – Aerodynamic Velocity, Included AoA

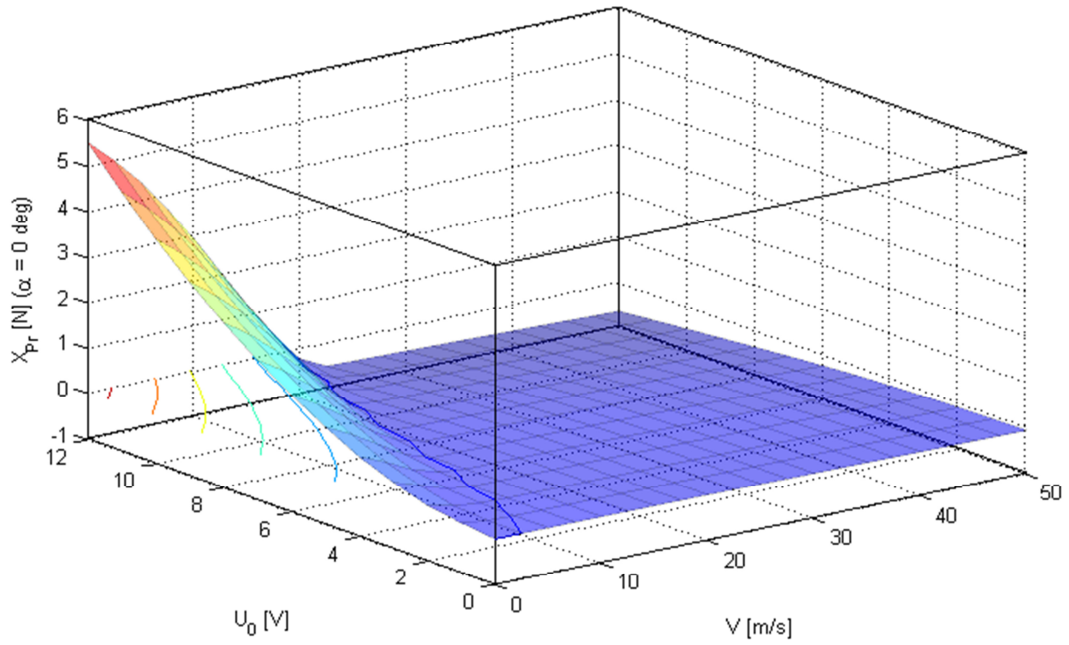


Figure 5.24 Back Engine Thrust – Input Voltage, Aerodynamic Velocity

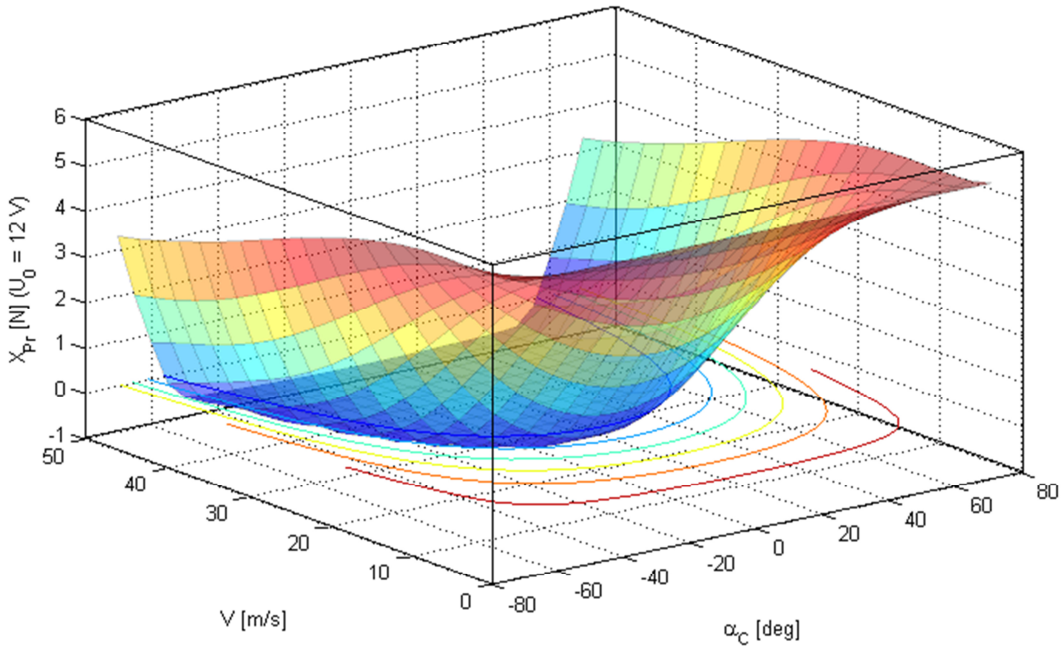


Figure 5.25 Back Engine Thrust – Aerodynamic Velocity, Included AoA

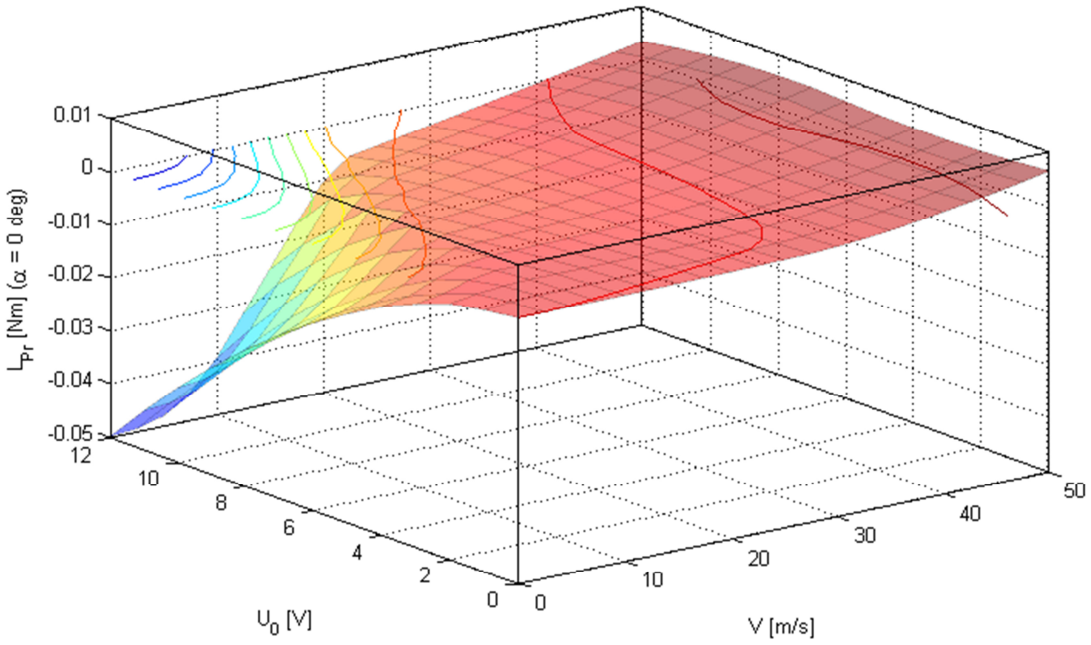


Figure 5.26 Back Engine Torque – Input Voltage, Aerodynamic Velocity

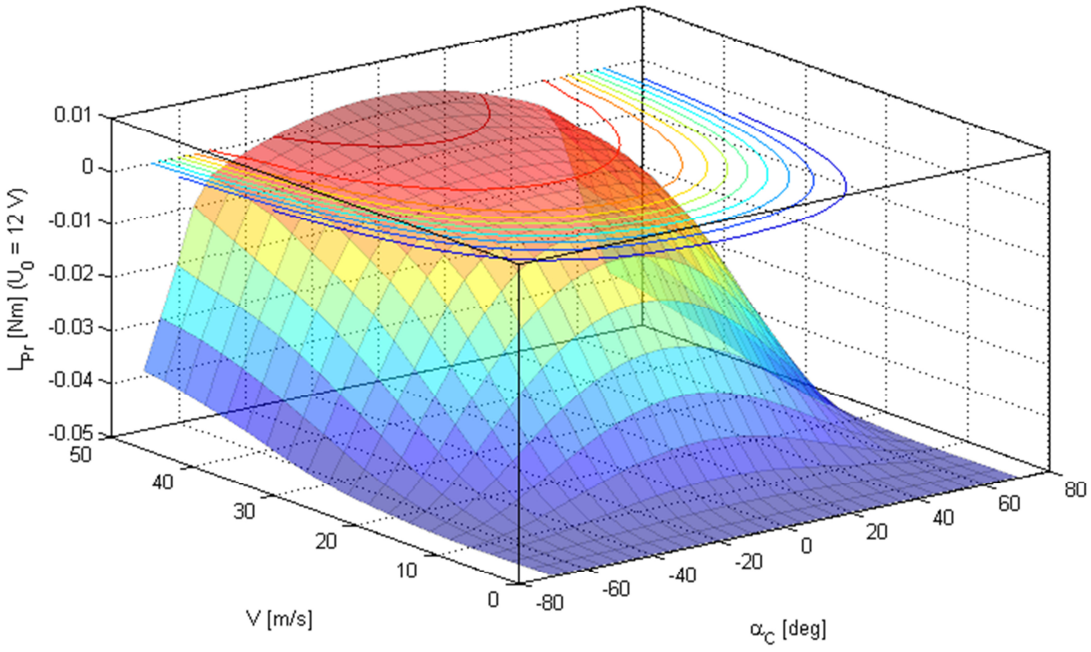


Figure 5.27 Back Engine Torque – Aerodynamic Velocity, Included AoA

5.3.4 Control System Implementation

As a rate command system is implemented, the linearizing state feedback is governed by moment equation (5.3). Using parameterizations (5.10) and (5.16) for aerodynamic and propulsive moments, we obtain differential equations with a structure, compliant with (4.90), comprising a pure state dependent part, a part that depends on the nonaffine thrust vector controls and a control affine part. It revealed that the radial basis function regressors in (5.16) for approximation of thrust and torque can be chosen to be equal without significant loss of accuracy, i.e.

$$\boldsymbol{\varphi}_{X,i}\left(V,\alpha_{c,i},U_{0,i}\right)=\boldsymbol{\varphi}_{L,i}\left(V,\alpha_{c,i},U_{0,i}\right)=:\boldsymbol{\varphi}_{i}\left(V,\alpha_{c,i},U_{0,i}\right).$$

Notice that the controller requires measurements of AoA and AoS. Currently, there are no sensors available at the aircraft that measure aerodynamic angles, however, the hardware platform is prepared for incorporation of an air data system (ADS) in the future. For simulation results, we will assume that an ADS is presently available. Using the simplified notation $\dot{\boldsymbol{\omega}}=\left(\dot{\vec{\boldsymbol{\omega}}}^{OB}\right)_B^B$ and inserting parameterizations into the moment equations (5.3),yields

$$\begin{aligned}\dot{\boldsymbol{\omega}}=&\mathbf{K}_x^T\boldsymbol{\varphi}_x\left(V,\alpha,\beta,p,q,r\right)+\mathbf{K}_N^T\boldsymbol{\varphi}_N\left(V,\alpha_{c,l},\alpha_{c_r},\alpha_{c,b},\sigma_l,\sigma_r,\sigma_b,\kappa_b,U_{0l},U_{0r},U_{0b}\right)+b(V)\mathbf{B}_L\mathbf{u}\\&+\boldsymbol{\delta}\left(V,\alpha,\beta,p,q,r,\sigma_l,\sigma_r,\sigma_b,\kappa_b,U_{0l},U_{0r},U_{0b}\right)\end{aligned}\tag{5.22}$$

where the single terms are defined on the subsequent pages. Thereby, the state dependent regressor $\boldsymbol{\varphi}(\cdot)$ contain terms that describe the aerodynamic moments, as well as the terms that occur due to inertia coupling (cross product term in equation (5.3)), $\boldsymbol{\varphi}_N(\cdot)$ contains all terms due to propulsion moments – basically the radial basis

functions – and $\mathbf{B}_L$ describes the influences due to control surfaces. Notice that $\mathbf{B}_L$ is a constant matrix while the state dependency is collected in $b(V)$ as is required for MMQ modification in (4.184). In the subsequent expression, the following quantities have been used.

- Position vectors from c.g. to the origin of the respective propulsion frame (Figure 2.3)

$$\left(\vec{\mathbf{r}}^{GPr_l}\right)=\begin{pmatrix} x^{GPr_l} \\ y^{GPr_l} \\ z^{GPr_l} \end{pmatrix} \qquad \left(\vec{\mathbf{r}}^{GPr_r}\right)=\begin{pmatrix} x^{GPr_r} \\ y^{GPr_r} \\ z^{GPr_r} \end{pmatrix} \qquad \left(\vec{\mathbf{r}}^{GPr_b}\right)=\begin{pmatrix} x^{GPr_b} \\ y^{GPr_b} \\ z^{GPr_b} \end{pmatrix}$$

- Inertia tensor relative to c.g. in body- fixed components

$$\left(\mathbf{I}^G\right)_B=\begin{bmatrix} I_x & 0 & I_{xz} \\ 0 & I_y & 0 \\ I_{xz} & 0 & I_z \end{bmatrix}$$

$$\boldsymbol{\varphi}_{N,l}\left(V,\alpha_{c,l},\sigma_l,U_{0,l}\right)=\begin{pmatrix}\sin\sigma_l\cdot\phi_{l,1}\left(V,\alpha_{c,l},U_{0,l}\right)\\ \cos\sigma_l\cdot\phi_{l,1}\left(V,\alpha_{c,l},U_{0,l}\right)\\ \vdots\\ \sin\sigma_l\cdot\phi_{l,s_l}\left(V,\alpha_{c,l},U_{0,l}\right)\\ \cos\sigma_l\cdot\phi_{l,s_l}\left(V,\alpha_{c,l},U_{0,l}\right)\end{pmatrix}\quad \boldsymbol{\varphi}_{N,r}\left(V,\alpha_{c,r},\sigma_r,U_{0,r}\right)=\begin{pmatrix}\sin\sigma_r\cdot\phi_{r,1}\left(V,\alpha_{c,r},U_{0,r}\right)\\ \cos\sigma_r\cdot\phi_{r,1}\left(V,\alpha_{c,r},U_{0,r}\right)\\ \vdots\\ \sin\sigma_r\cdot\phi_{r,s_r}\left(V,\alpha_{c,r},U_{0,r}\right)\\ \cos\sigma_r\cdot\phi_{r,s_r}\left(V,\alpha_{c,r},U_{0,r}\right)\end{pmatrix}$$

$$\mathbf{K}_{N,l}=\begin{bmatrix}\frac{I_z y^{G\,\mathrm{Pr}_l} p_{X,l,1}-I_{xz}p_{L,l,1}}{I_{xz}^2-I_xI_z} & \frac{x^{G\,\mathrm{Pr}_l}p_{X,l,1}}{I_y} & \frac{I_x p_{L,l,1}-I_{xz}y^{G\,\mathrm{Pr}_l}p_{X,l,1}}{I_{xz}^2-I_xI_z}\\ -\frac{I_z p_{L,l,1}+I_{xz}y^{G\,\mathrm{Pr}_l}p_{X,l,1}}{I_{xz}^2-I_xI_z} & \frac{z^{G\,\mathrm{Pr}_l}p_{X,l,1}}{I_y} & \frac{I_{xz}p_{L,l,1}+I_x y^{G\,\mathrm{Pr}_l}p_{X,l,1}}{I_{xz}^2-I_xI_z}\\ \vdots & \vdots & \vdots\\ \frac{I_z y^{G\,\mathrm{Pr}_l} p_{X,l,s_l}-I_{xz}p_{L,l,s_l}}{I_{xz}^2-I_xI_z} & \frac{x^{G\,\mathrm{Pr}_l}p_{X,s_l}}{I_y} & \frac{I_x p_{L,l,s_l}-I_{xz}y^{G\,\mathrm{Pr}_l}p_{X,l,s_l}}{I_{xz}^2-I_xI_z}\\ -\frac{I_z p_{L,l,s_l}+I_{xz}y^{G\,\mathrm{Pr}_l}p_{X,l,s_l}}{I_{xz}^2-I_xI_z} & \frac{z^{G\,\mathrm{Pr}_l}p_{X,l,s_l}}{I_y} & \frac{I_{xz}p_{L,l,s_l}+I_x y^{G\,\mathrm{Pr}_l}p_{X,l,s_l}}{I_{xz}^2-I_xI_z}\end{bmatrix}$$

$$\mathbf{K}_{N,r}=\begin{bmatrix}\frac{I_z y^{G\,\mathrm{Pr}_r} p_{X,r,1}-I_{xz}p_{L,r,1}}{I_{xz}^2-I_xI_z} & \frac{x^{G\,\mathrm{Pr}_r}p_{X,r,1}}{I_y} & \frac{I_x p_{L,r,1}-I_{xz}y^{G\,\mathrm{Pr}_r}p_{X,r,1}}{I_{xz}^2-I_xI_z}\\ -\frac{I_z p_{L,r,1}+I_{xz}y^{G\,\mathrm{Pr}_r}p_{X,r,1}}{I_{xz}^2-I_xI_z} & \frac{z^{G\,\mathrm{Pr}_r}p_{X,r,1}}{I_y} & \frac{I_{xz}p_{L,r,1}+I_x y^{G\,\mathrm{Pr}_r}p_{X,r,1}}{I_{xz}^2-I_xI_z}\\ \vdots & \vdots & \vdots\\ \frac{I_z y^{G\,\mathrm{Pr}_r} p_{X,r,s_r}-I_{xz}p_{L,r,s_r}}{I_{xz}^2-I_xI_z} & \frac{x^{G\,\mathrm{Pr}_r}p_{X,s_r}}{I_y} & \frac{I_x p_{L,r,s_r}-I_{xz}y^{G\,\mathrm{Pr}_l}p_{X,r,s_r}}{I_{xz}^2-I_xI_z}\\ -\frac{I_z p_{L,r,s_r}+I_{xz}y^{G\,\mathrm{Pr}_r}p_{X,r,s_r}}{I_{xz}^2-I_xI_z} & \frac{z^{G\,\mathrm{Pr}_l}p_{X,l,s_r}}{I_y} & \frac{I_{xz}p_{L,r,s_r}+I_x y^{G\,\mathrm{Pr}_r}p_{X,r,s_r}}{I_{xz}^2-I_xI_z}\end{bmatrix}$$

$$\boldsymbol{\varphi}_{N,b}\left(V,\alpha_{c,b},\sigma_b,\kappa_b,U_{0,l}\right)=\begin{pmatrix}\cos\kappa_b\sin\sigma_b\cdot\phi_{b,1}\left(V,\alpha_{c,b},U_{0,b}\right)\\ \cos\kappa_b\cos\sigma_b\cdot\phi_{b,1}\left(V,\alpha_{c,b},U_{0,b}\right)\\ \sin\kappa_b\cdot\phi_{b,1}\left(V,\alpha_{c,b},U_{0,b}\right)\\ \vdots\\ \cos\kappa_b\sin\sigma_b\cdot\phi_{b,s_b}\left(V,\alpha_{c,b},U_{0,b}\right)\\ \cos\kappa_b\cos\sigma_b\cdot\phi_{b,s_b}\left(V,\alpha_{c,b},U_{0,b}\right)\\ \sin\kappa_b\cdot\phi_{b,s_b}\left(V,\alpha_{c,b},U_{0,b}\right)\end{pmatrix}$$

$$\mathbf{K}_{N,b}=\begin{bmatrix}\frac{I_z y^{G\mathrm{Pr}_b}p_{X,b,1}-I_{xz}p_{L,b,1}}{I_{xz}^2-I_xI_z} & \frac{x^{\mathrm{Pr}_b}p_{X,b,1}}{I_y} & \frac{I_x p_{L,b,1}-I_{xz}y^{G\mathrm{Pr}_b}p_{X,b,1}}{I_{xz}^2-I_xI_z}\\ -\frac{I_z p_{L,b,1}+I_{xz}y^{G\mathrm{Pr}_b}p_{X,b,1}}{I_{xz}^2-I_xI_z} & \frac{z^{G\mathrm{Pr}_b}p_{X,b,1}}{I_y} & \frac{I_{xz}p_{L,b,1}+I_x y^{G\mathrm{Pr}_b}p_{X,b,1}}{I_{xz}^2-I_xI_z}\\ \frac{I_{xz}x^{G\mathrm{Pr}_b}p_{X,b,1}+I_z z^{G\mathrm{Pr}_b}p_{X,b,1}}{I_{xz}^2-I_xI_z} & \frac{p_{L,b,1}}{I_y} & -\frac{I_x x^{G\mathrm{Pr}_b}p_{X,b,1}+I_{xz}z^{G\mathrm{Pr}_b}p_{X,b,1}}{I_{xz}^2-I_xI_z}\\ \vdots & \vdots & \vdots\\ \frac{I_z y^{G\mathrm{Pr}_b}p_{X,b,s_b}-I_{xz}p_{L,b,s_b}}{I_{xz}^2-I_xI_z} & \frac{x^{\mathrm{Pr}_b}p_{X,b,s_b}}{I_y} & \frac{I_x p_{L,b,s_b}-I_{xz}y^{G\mathrm{Pr}_b}p_{X,b,s_b}}{I_{xz}^2-I_xI_z}\\ -\frac{I_z p_{L,b,s_b}+I_{xz}y^{G\mathrm{Pr}_b}p_{X,b,s_b}}{I_{xz}^2-I_xI_z} & \frac{z^{G\mathrm{Pr}_b}p_{X,b,s_b}}{I_y} & \frac{I_{xz}p_{L,b,s_b}+I_x y^{G\mathrm{Pr}_b}p_{X,b,s_b}}{I_{xz}^2-I_xI_z}\\ \frac{I_{xz}x^{G\mathrm{Pr}_b}p_{X,b,s_b}+I_z z^{G\mathrm{Pr}_b}p_{X,b,s_b}}{I_{xz}^2-I_xI_z} & \frac{p_{L,b,s_b}}{I_y} & -\frac{I_x x^{G\mathrm{Pr}_b}p_{X,b,s_b}+I_{xz}z^{G\mathrm{Pr}_b}p_{X,b,s_b}}{I_{xz}^2-I_xI_z}\end{bmatrix}$$

$$\boldsymbol{\varphi}_N\left(V,\alpha_{c,l},\alpha_{c_r},\alpha_{c,b},\sigma_l,\sigma_r,\sigma_b,\kappa_b,U_{0l},U_{0r},U_{0b}\right)=\begin{pmatrix}\boldsymbol{\varphi}_{N,l}\left(V,\alpha_{c,l},\sigma_l,U_{0,l}\right)\\ \boldsymbol{\varphi}_{N,r}\left(V,\alpha_{c,r},\sigma_r,U_{0,r}\right)\\ \boldsymbol{\varphi}_{N,b}\left(V,\alpha_{c,b},\sigma_b,\kappa_b,U_{0,l}\right)\end{pmatrix}$$

(5.23)

$$\mathbf{K}_N=\begin{bmatrix}\mathbf{K}_{N,l}\\ \mathbf{K}_{N,r}\\ \mathbf{K}_{N,b}\end{bmatrix}$$

$$\boldsymbol{\delta}_A(V,\alpha,\beta,p,q,r)=\frac{\rho}{2}V^2S\begin{pmatrix} b/2\,\delta_{AL} \\ \bar{c}\,\delta_{AM} \\ b/2\,\delta_{AN}\end{pmatrix}$$

$$\boldsymbol{\delta}_P\left(V,\alpha_{c,l},\sigma_{c,r},\sigma_{c,b},\sigma_{l,},\sigma_l,\sigma_b,\kappa_b,U_{0l},U_{0r}U_{0b}\right)=$$

$$\begin{pmatrix} \delta_{PL,l}\cos\sigma_l-\delta_{PX,l}y^{G\mathrm{Pr}_l}\sin\sigma_l+\delta_{PL,r}\cos\sigma_r-\delta_{PX,r}y^{G\mathrm{Pr}_r}\sin\sigma_r \\ \delta_{PX,l}z^{G\mathrm{Pr}_l}\cos\sigma_l+\delta_{PX,l}x^{G\mathrm{Pr}_l}\sin\sigma_l+\delta_{PX,r}z^{G\mathrm{Pr}_r}\cos\sigma_r+\delta_{PX,r}x^{G\mathrm{Pr}_r}\sin\sigma_r \\ -\delta_{PL,l}\sin\sigma_l-\delta_{PX,l}y^{G\mathrm{Pr}_l}\cos\sigma_l-\delta_{PL,r}\sin\sigma_r-\delta_{PX,r}y^{G\mathrm{Pr}_r}\cos\sigma_r \end{pmatrix}$$

$$+\begin{pmatrix} \delta_{PL,b}\cos\kappa_b\cos\sigma_b-\delta_{PX,b}y^{G\mathrm{Pr}_b}\cos\kappa_b\sin\sigma_b \\ \delta_{PL,b}\sin\kappa_b+\delta_{PX,b}z^{G\mathrm{Pr}_b}\cos\kappa_b\cos\sigma_b+\delta_{PX,b}x^{G\mathrm{Pr}_b}\cos\kappa_b\sin\sigma_b \\ \delta_{PX,b}x^{G\mathrm{Pr}_b}\sin\kappa_b-\delta_{PL,b}\cos\kappa_b\sin\kappa_b-\delta_{PX,b}y^{GPR_b}\cos\kappa_b\cos\sigma_b \end{pmatrix}$$

$$\boldsymbol{\delta}(V,\alpha,\beta,p,q,r,\sigma_l,\alpha_{c,l},\alpha_{c,r},\alpha_{c,b},\sigma_r,\sigma_b,\kappa_b,U_{0l},U_{0r},U_{0b})=\boldsymbol{\delta}_A+\boldsymbol{\delta}_P$$

$$\mathbf{B}_L = \rho S \begin{bmatrix} \frac{b}{4}\frac{I_{xz}C_{n\eta_{c,l}} - I_z C_{l,\eta_{c,l}}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m,\eta_{c,l}}}{2I_y} & \frac{b}{4}\frac{-I_x C_{n\eta_{c,l}} + I_{xz}C_{l\eta_{c,l}}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\eta_{c,r}} - I_z C_{l,\eta_{c,r}}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m,\eta_{c,r}}}{2I_y} & \frac{b}{4}\frac{-I_x C_{n\eta_{c,r}} + I_{xz}C_{l\eta_{c,r}}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\xi_l} - I_z C_{l,\xi_l}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m,\xi_l}}{2I_y} & \frac{b}{4}\frac{-I_x C_{n\xi_l} + I_{xz}C_{l\xi_l}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\xi_r} - I_z C_{l,\xi_r}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m,\xi_r}}{2I_y} & \frac{b}{4}\frac{-I_x C_{n\xi_r} + I_{xz}C_{l\xi_r}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\delta_l} - I_z C_{l,\delta_l}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m,\delta_l}}{2I_y} & \frac{b}{4}\frac{-I_x C_{n\delta_l} + I_{xz}C_{l\delta_l}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\delta_r} - I_z C_{l,\delta_r}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m,\delta_r}}{2I_y} & \frac{b}{4}\frac{-I_x C_{n\delta_r} + I_{xz}C_{l\delta_r}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\eta_l} - I_z C_{l,\eta_l}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m,\eta_l}}{2I_y} & \frac{b}{4}\frac{-I_x C_{n\eta_l} + I_{xz}C_{l\eta_l}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\eta_r} - I_z C_{l,\eta_{c,r}}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m,\eta_r}}{2I_y} & \frac{b}{4}\frac{-I_x C_{n\eta_r} + I_{xz}C_{l\eta_r}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\zeta} - I_z C_{l,\zeta}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m,\zeta}}{2I_y} & \frac{b}{4}\frac{-I_x C_{n\zeta} + I_{xz}C_{l\zeta}}{I_{xz}^2 - I_x I_z} \end{bmatrix}$$

$$\mathbf{u} = \begin{pmatrix} \eta_{c,l} \\ \eta_{c,r} \\ \xi_l \\ \xi_r \\ \delta_l \\ \delta_r \\ \eta_l \\ \eta_r \\ \zeta \end{pmatrix}$$

$$b(V) = V^2$$

$$\mathbf{K}_x = \rho S \begin{bmatrix} \frac{b}{4}\frac{I_{xz}C_{n0} - I_z C_{l0}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m0}}{I_y} & \frac{b}{4}\frac{I_{xz}C_{l0} - I_x C_{n0}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\alpha} - I_z C_{l\alpha}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m\alpha}}{I_y} & \frac{b}{4}\frac{I_{xz}C_{l\alpha} - I_x C_{n\alpha}}{I_{xz}^2 - I_x I_z} \\ \frac{b}{4}\frac{I_{xz}C_{n\beta} - I_z C_{l\beta}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}C_{m\beta}}{I_y} & \frac{b}{4}\frac{I_{xz}C_{l\beta} - I_x C_{n\beta}}{I_{xz}^2 - I_x I_z} \\ \frac{I_{xz}\left(I_x - I_z\right)}{I_{xz}^2 - I_x I_z} & 0 & \frac{I_x\left(I_y - I_x\right)}{I_{xz}^2 - I_x I_z} \\ \frac{b^2}{8}\frac{I_{xz}C_{np} - I_z C_{lp}}{I_{xz}^2 - I_x I_z} & \frac{b\bar{c}}{4}\frac{C_{mp}}{I_y} & \frac{b^2}{8}\frac{I_{xz}C_{lp} - I_x C_{np}}{I_{xz}^2 - I_x I_z} \\ \frac{I_{xz}^2}{I_{xz}^2 - I_x I_z} & 0 & -\frac{I_x I_{xz}}{I_{xz}^2 - I_x I_z} \\ \frac{b\bar{c}}{8}\frac{I_{xz}C_{nq} - I_z C_{lq}}{I_{xz}^2 - I_x I_z} & \frac{\bar{c}^2}{4}\frac{C_{mq}}{I_y} & \frac{b\bar{c}}{8}\frac{I_{xz}C_{lq} - I_x C_{nq}}{I_{xz}^2 - I_x I_z} \\ \frac{b^2}{8}\frac{I_{xz}C_{nr} - I_z C_{lr}}{I_{xz}^2 - I_x I_z} & \frac{b\bar{c}}{4}\frac{C_{mr}}{I_y} & \frac{b^2}{8}\frac{I_{xz}C_{lr} - I_x C_{nr}}{I_{xz}^2 - I_x I_z} \\ 0 & \frac{I_{xz}}{I_y} & 0 \\ 0 & \frac{I_z - I_x}{I_y} & 0 \\ 0 & -\frac{I_{xz}}{I_y} & 0 \end{bmatrix}$$

$$\boldsymbol{\varphi}_x(V, \alpha, \beta, p, q, r) = \begin{pmatrix} V^2 \\ V^2\alpha \\ V^2\beta \\ pq \\ Vp \\ qr \\ Vq \\ Vr \\ p^2 \\ pr \\ r^2 \end{pmatrix}$$

The parameters $\mathbf{K}_x$ and $\mathbf{K}_N$ are, of course, subjected to model uncertainties as well as $\mathbf{B}_L$. In section 4.3 the uncertainty in the decoupling matrix has been modeled by means of a control effectiveness matrix $\mathbf{\Lambda}_L$ using two different parameterizations. For this application, we will utilize the relative parameterization according to (4.114). The true decoupling matrix is

$$\mathbf{B}_L\left(\mathbf{I}+\mathbf{\Lambda}_L\right)$$

with an unknown matrix $\mathbf{\Lambda}_L$. Overall, the number of uncertain parameters is:

- $\mathbf{K}_x$: $11\cdot 3=33$
- $\mathbf{K}_N$: $2(s_l+s_r)+3s_b$
- $\mathbf{\Lambda}_L$: $9\cdot 9=81$

According to section 4.3, the linearizing state feedback is

$$\mathbf{u}=\hat{\mathbf{B}}^T(V)\left[\hat{\mathbf{B}}(V)\hat{\mathbf{B}}^T(V)\right]^{-1}\left[\boldsymbol{\nu}-\hat{\mathbf{K}}_x^T\boldsymbol{\varphi}_x-\left(\hat{\mathbf{K}}_N^T+\hat{\mathbf{\Theta}}_N^T\right)\boldsymbol{\varphi}_N\right] \tag{5.23}$$

where

$$\hat{\mathbf{B}}(V)=b(V)\mathbf{B}_L\left(\mathbf{I}+\hat{\mathbf{\Lambda}}_L\right),$$

$\hat{\mathbf{K}}_x$, $\hat{\mathbf{K}}_N$ are nonadaptive values, assumed for $\mathbf{K}_x$, $\mathbf{K}_N$ and $\hat{\mathbf{\Theta}}_N$ is an adaptive estimate for the deviation between $\hat{\mathbf{K}}_N$ and $\mathbf{K}_N$ according to (4.117), (4.119) and $\hat{\mathbf{\Lambda}}_L$ is an adaptive estimate for $\mathbf{\Lambda}_L$. Note that the independent variables of the regressors have been dropped for readability. Hence, the feedback linearized moment dynamics are

$$\dot{\boldsymbol{\omega}}=\boldsymbol{\nu}+\mathbf{\Theta}_x^T\boldsymbol{\varphi}_x-\tilde{\mathbf{\Theta}}_N^T\boldsymbol{\varphi}_N-b\mathbf{B}_L\tilde{\mathbf{\Lambda}}_L\mathbf{u}+\boldsymbol{\delta}$$

with parameter-estimation-errors $\tilde{\mathbf{\Lambda}}_L$, $\tilde{\mathbf{\Theta}}_N$, according to (4.112), (4.119) and $\mathbf{\Theta}_x$ according to (4.109). Further, the reference dynamics are designed as a first order lag.

$$\dot{\boldsymbol{\omega}}_R=\mathbf{K}^T\left(-\boldsymbol{\omega}_R+\boldsymbol{\omega}_C\right)$$

Thereby

$$\mathbf{K}=\begin{bmatrix}T_p^{-1} & 0 & 0\\ 0 & T_q^{-1} & 0\\ 0 & & T_r^{-1}\end{bmatrix}$$

with time constants T_p, T_q, $T_r>0$, reference angular rates $\boldsymbol{\omega}_R^T=\begin{pmatrix}p_R & q_R & r_R\end{pmatrix}$ and angular rates command $\boldsymbol{\omega}_C^T=\begin{pmatrix}p_C & q_C & r_C\end{pmatrix}$. The pseudo control contains feed forward reference command, error feedback and an adaptive part

$$\boldsymbol{\nu}=\boldsymbol{\nu}_R+\boldsymbol{\nu}_E+\boldsymbol{\nu}_A$$
$$\boldsymbol{\nu}_R=\mathbf{K}^T\left(-\boldsymbol{\omega}_R+\boldsymbol{\omega}_C\right)\quad,\quad\boldsymbol{\nu}_E=\mathbf{C}^T\mathbf{e}\quad,\quad\boldsymbol{\nu}_A=-\hat{\mathbf{\Theta}}_x^T\boldsymbol{\varphi}_x$$

where $\hat{\boldsymbol{\Theta}}_x$ is an adaptive estimate for $\boldsymbol{\Theta}_x$,

$$\mathbf{e}_\omega = \boldsymbol{\omega}_R - \boldsymbol{\omega}$$

is the tracking error and

$$\mathbf{C} = \begin{bmatrix} k_p & 0 & 0 \\ 0 & k_q & 0 \\ 0 & 0 & k_r \end{bmatrix}$$

contains the error feedback gains k_p, k_q, k_r>0. According to (4.127) tracking error dynamics are

$$\dot{\mathbf{e}}_\omega = -\mathbf{C}^T\mathbf{e}_\omega + \tilde{\boldsymbol{\Theta}}_x^T\boldsymbol{\varphi}_x + \tilde{\boldsymbol{\Theta}}_N^T\boldsymbol{\varphi}_N + b\mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L - \boldsymbol{\delta}$$

with $\tilde{\boldsymbol{\Theta}}_x$ according to (4.124) and, according to Corollary 4.2, the updates of the parameter estimates are

$$\dot{\hat{\boldsymbol{\Theta}}}_x = \mathrm{Proj}\left[\hat{\boldsymbol{\Theta}}_x, -\boldsymbol{\Gamma}_x\boldsymbol{\varphi}_x\mathbf{e}_\omega^T\mathbf{P}_E - \sigma_x\bar{f}_{\varepsilon_x,\theta_x,\boldsymbol{\Gamma}_x}\left(\hat{\boldsymbol{\Theta}}_x\right)\hat{\boldsymbol{\Theta}}_x - \kappa\boldsymbol{\Gamma}_x\mathbf{Q}_x\tilde{\mathbf{C}}^T\right]_{\varepsilon_x,\theta_{x,\max},\boldsymbol{\Gamma}_x}$$

$$\dot{\hat{\boldsymbol{\Theta}}}_N = \mathrm{Proj}\left[\hat{\boldsymbol{\Theta}}_N, -\boldsymbol{\Gamma}_N\boldsymbol{\varphi}_N\mathbf{e}_\omega^T\mathbf{P}_E - \sigma_N\bar{f}_{\varepsilon_N,\theta_N,\boldsymbol{\Gamma}_N}\left(\hat{\boldsymbol{\Theta}}_N\right)\hat{\boldsymbol{\Theta}}_N - \kappa\boldsymbol{\Gamma}_N\mathbf{Q}_N\tilde{\mathbf{C}}^T\right]_{\varepsilon_N,\theta_{N,\max},\boldsymbol{\Gamma}_N}$$

$$\dot{\hat{\boldsymbol{\Lambda}}}_L^T = \mathrm{Proj}\left[\hat{\boldsymbol{\Lambda}}_L^T, -\boldsymbol{\Gamma}_L\mathbf{u}\mathbf{e}_\omega^T\mathbf{P}_E\mathbf{B}_L b - \sigma_L\bar{f}_{\varepsilon_L,\lambda_L,\boldsymbol{\Gamma}_L}\left(\hat{\boldsymbol{\Lambda}}_L\right)\hat{\boldsymbol{\Lambda}}_L^T - \kappa\boldsymbol{\Gamma}_L\mathbf{Q}_L\tilde{\mathbf{C}}^T\mathbf{B}_L\right]_{\varepsilon_L,\lambda_{L,\max},\boldsymbol{\Gamma}_L}$$

where $\mathbf{P}_E$ is the symmetric positive definite solution of the Lyapunov equation

$$-\mathbf{C}^T\mathbf{P}_E - \mathbf{P}_E\mathbf{C} = -\mathbf{Q}_E$$

for some symmetric positive definite $\mathbf{Q}_E$. With stable filters $G_i(s)$, $i=1,\ldots,N$ we compute for MMQ modification

- the filtered uncertainty according to (4.181)
  $\mathbf{c}_i(s) = G_i(s)s\boldsymbol{\omega}(s) + G_i(s)\left[-\mathbf{v}(s) + \left[\hat{\boldsymbol{\Theta}}_N^T\boldsymbol{\varphi}_N\right](s) + \left[b\mathbf{B}_L\hat{\boldsymbol{\Lambda}}_L\mathbf{u}\right](s)\right]$
- the filtered regressor according to (4.186)
  $\mathbf{q}_{x,i}(s) = G_i(s)\boldsymbol{\varphi}_x(s)$, $\mathbf{q}_{N,i}(s) = G_i(s)\boldsymbol{\varphi}_N(s)$, $\mathbf{q}_{L,i}(s) = G_i(s)[b\mathbf{u}](s)$
- filtered estimated uncertainty according to (4.189)
  $\hat{\mathbf{c}}_i(s) = \hat{\boldsymbol{\Theta}}_x^T\mathbf{q}_{x,i} + \hat{\boldsymbol{\Theta}}_N^T\mathbf{q}_{N,i} + \mathbf{B}_L\hat{\boldsymbol{\Lambda}}_L\mathbf{q}_{L,i}$
- filtered uncertainty estimation error according to (4.190) $\tilde{\mathbf{c}}_i(\mathbf{s}) = \hat{\mathbf{c}}_i(s) - \mathbf{c}_i(s)$

and stack the filtered quantities into matrices.

$$\mathbf{Q}_x(s) = [\mathbf{q}_{x,1}(s) \quad \cdots \quad \mathbf{q}_{x,N}(s)] \ , \ \mathbf{Q}_N(s) = [\mathbf{q}_{N,1}(s) \quad \cdots \quad \mathbf{q}_{N,N}(s)] \ , \ \mathbf{Q}_L(s) = [\mathbf{q}_{L,1}(s) \quad \cdots \quad \mathbf{q}_{L,N}(s)]$$
$$\mathbf{C}(s) = [\tilde{\mathbf{c}}_1(s) \quad \cdots \quad \tilde{\mathbf{c}}_N(s)]$$

Still an open topic is the design of the filters and a possible design will be presented in the following.

Design of MMQ Filters

It is desirable that the single MMQ filters provide different information, since this increases the chance that columns of the matrix of filtered regressors contain a maximum number of linearly independent columns. This in turn leads to increased adaptation performance as has been explained in section 4.4. Therefore, we will use frequency separating filters, with a low-pass that covers the low frequencies and $N\text{-}1$ band-pass filters, which separate the relevant frequency band up to a certain maximum frequency. The low-pass filter is of the form

$$G_L = \frac{\omega_0}{s+\omega_0}$$

with natural frequency ω_0 and the band-passes are of the form

$$G_B(s) = \frac{2\zeta\omega_0 s}{s^2+2\zeta\omega_0 s+\omega_0^2}$$

with a damping ratio ζ. Before designing the frequency separation, at first, recall some basic facts on low-pass and band-pass. The amplitude response of low-pass and band-pass is depicted in Figure 5.28. The low-pass has a constant amplification gain up to the natural frequency and then it decays with $20dB$ per decade. The band-pass only transmits frequencies between some low frequency ω_L and a high frequency ω_H, which evaluate to

$$\omega_L = \omega_0\left(-\zeta+\sqrt{\zeta^2+1}\right) \quad , \quad \omega_H = \omega_0\left(\zeta+\sqrt{\zeta^2+1}\right)$$

and outside the transmission band the amplitude response decays with $20dB$ per decade. The maximum amplitude is achieved at ω_0, but it does not lie in the middle of the transmission band whose bandwidth evaluates to

$$B = \omega_H - \omega_L = 2\zeta\omega_0 . \tag{5.24}$$

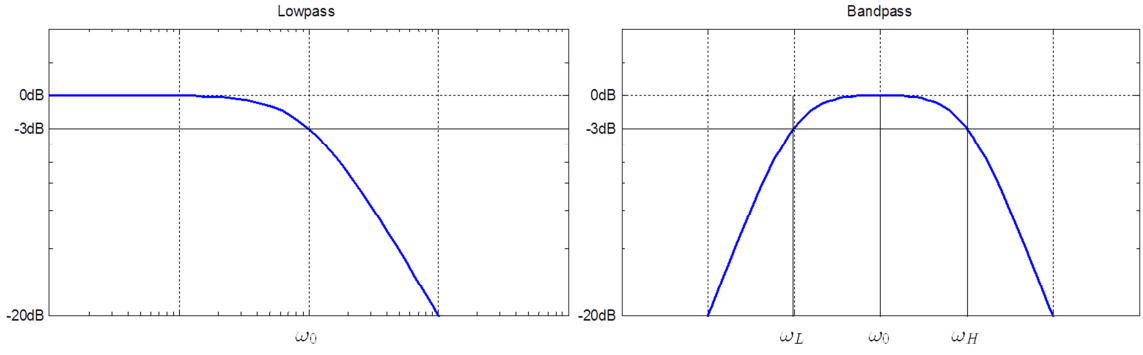


Figure 5.28 Amplitude Response Low-pass and Band-pass

Now, we want to design as set of N frequency separating filters that transmit frequencies up to some maximum $\overline{\Omega}$. The low pass filter

$$G_L(s) = \frac{\Omega_0}{s + \Omega_0}$$

thereby covers the low frequencies up to some $\Omega_0 < \overline{\Omega}$, and the remaining $N\text{-}1$ filters are designed as band-passes

$$G_{B,i}(s) = \frac{2\zeta\omega_{0,i}s}{s^2 + 2\zeta\omega_{0,i}s + \omega_{0,i}^2}$$

such that each filter covers a separate frequency band without overlap, as depicted in Figure 5.29. The low frequency of the first band-pass equals the bandwidth of the low-pass $\omega_{L1}=\Omega_0$ and its high frequency equals the low frequency of the second band pass $\omega_{L2}=\omega_{H1}$. This procedure is repeated until the high frequency of the N^{th} filter equals the maximum frequency $\overline{\Omega} = \omega_{H,N}$. As physical systems, such as aircraft typically have a limited bandwidth, it is expectable that the information content of the time response reduces with increased frequency. Therefore, it is nearby to widen the bandwidth of the filters with increasing frequency.

It follows from the definition of the bandwidth (5.24) that this could be achieved by setting ζ constant for all filters, while the position of the transmission band is shifted by the natural frequency ω_0.

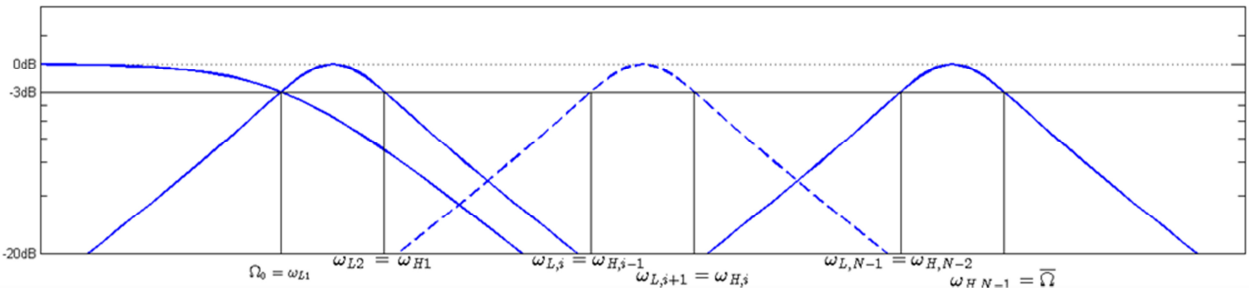


Figure 5.29 Frequency Separation

Se we need to find the constant damping ratio ζ and natural frequencies $\omega_{0,i}$ such that the frequency band from Ω_0 to $\overline{\Omega}$ is divided into $N\text{-}1$ separate parts. To this end, consider the relationship between natural and low frequency of the i^{th} filter.

$$\omega_{0,i} = \frac{\omega_{L,i}}{-\zeta + \sqrt{\zeta^2+1}} = \frac{\omega_{L,i}\left(\zeta + \sqrt{\zeta^2+1}\right)}{\left(-\zeta + \sqrt{\zeta^2+1}\right)\left(\zeta + \sqrt{\zeta^2+1}\right)} = \frac{\omega_{L,i}\left(\zeta + \sqrt{\zeta^2+1}\right)}{-\zeta^2 + \zeta^2 + 1}$$

$$\omega_{0,i} = \omega_{L,i}\left(\zeta + \sqrt{\zeta^2+1}\right) \tag{5.25}$$

Further, using the bandwidth, the high frequency is

$$\omega_{H,i} = \omega_{L,i} + B = \omega_{L,i} + 2\zeta\omega_{0,i} = \omega_{L,i} + 2\zeta\omega_{L,i}\left(\zeta + \sqrt{\zeta^2+1}\right)$$

$$\omega_{H,i} = \omega_{L,i}\left(1 + 2\zeta^2 + 2\zeta\sqrt{\zeta^2+1}\right). \tag{5.26}$$

We set $\omega_{H,i}=\omega_{L,i+1}$ and obtain a recursive definition for the respective low frequencies.

$$\omega_{L,i+1} = \omega_{L,i}\left(1 + 2\zeta^2 + 2\zeta\sqrt{\zeta^2+1}\right). \tag{5.27}$$

Further, since $\omega_{L,1}=\Omega_0$ we get. by successive application of (5.27)

$$\omega_{L,i} = \Omega_0\left(1 + 2\zeta^2 + 2\zeta\sqrt{\zeta^2+1}\right)^{i-1} \tag{5.28}$$

and by (5.26)

$$\omega_{H,i} = \Omega_0\left(1 + 2\zeta^2 + 2\zeta\sqrt{\zeta^2+1}\right)^{i}.$$

For the $(N\text{-}1)^{th}$ band-pass, we have

$$\omega_{H,N-1} = \Omega_0\left(1 + 2\zeta^2 + 2\zeta\sqrt{\zeta^2+1}\right)^{N-1} =: \overline{\Omega} \tag{5.29}$$

which is a condition for ζ. In order to solve the equation, as an interim step, we define

$$C = 1 + 2\zeta^2 + 2\zeta\sqrt{\zeta^2+1}\,. \tag{5.30}$$

By (5.29) we obtain

$$C = \left(\frac{\overline{\Omega}}{\Omega_0}\right)^{\frac{1}{N-1}} \tag{5.31}$$

and finally, (5.30) can be solved for C.

$$\begin{aligned}\left(2\zeta\sqrt{\zeta^2+1}\right)^2 &= \left(C-1-2\zeta^2\right)^2\\ 4\zeta^4 + 4\zeta^2 &= (C-1)^2 - 4(C-1)\zeta^2 + 4\zeta^4\\ \zeta &= \frac{(C-1)}{2\sqrt{C}}\end{aligned} \tag{5.32}$$

Using this result, the natural frequencies are obtained by insertion of (5.28) into (5.25).

$$\omega_{0,i} = \Omega_0\left(1 + 2\zeta^2 + 2\zeta\sqrt{\zeta^2+1}\right)^{i-1}\left(\zeta + \sqrt{\zeta^2+1}\right) \tag{5.33}$$

$\mathcal{L}_1$-norms of MMQ filters

The impulse response of a low-pass is

$$g_L(t) = e^{-\omega_0 t}$$

and the $\mathcal{L}_1$-norm is the integral over the absolute value of impulse response

$$\begin{aligned}\left\|G_L(s)\right\|_{\mathcal{L}_1} &= \int_0^\infty |g_L(t)|dt\\ \left\|G_L(s)\right\|_{\mathcal{L}_1} &= \frac{1}{\omega_0}\end{aligned} \tag{5.34}$$

For weakly damped band-passes ($\zeta<1$), the impulse response is obtained by completing squares of the denominator of the transfer function

$$G_B(s) = \frac{2\zeta\omega_0 s}{(s+\zeta\omega_0)^2 + (1-\zeta^2)\omega_0^2}\,.$$

and the associated time domain signal is obtained from tables available in the literature.

$$g_B(t) = 2\zeta\omega_0 e^{-\zeta\omega_0 t}\left[\cos\left(\omega_0\sqrt{1-\zeta^2}\,t\right) - \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\left(\omega_0\sqrt{1-\zeta^2}\,t\right)\right]$$

Direct integration of the magnitude of this signal is not even solved by symbolic mathematics software tools such as Wolfram MATHEMATICA and Mathwork's MUPAD but, using some smaller reformulations, the $\mathcal{L}_1$-norm can be computed analytically. At first, using addition theorems for trigonometric functions, the impulse response also reads as

$$g_B(t) = 2\frac{\zeta}{\sqrt{1-\zeta^2}}\,\omega_0 e^{-\zeta\omega_0 t}\sin\left[-\omega_0\sqrt{1-\zeta^2}\,t + \arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)\right].$$

Moreover, the signal has a constant sign within the intervals $[T_{k-1}, T_k]$ where

$$T_k = \frac{\arctan\left(\dfrac{\sqrt{1-\zeta^2}}{\zeta}\right) + k\pi}{\omega_0\sqrt{1-\zeta^2}}$$

for $k \in \mathbb{Z}$ and, using symbolic mathematics software, the integral of $g_B(t)$ over the k^{th} interval evaluates to

$$I_k = \int_{T_{k-1}}^{T_k} |g_B(t)|\,dt = 2(-1)^k\,\zeta\, e^{-\frac{k\pi\zeta}{\sqrt{1-\zeta^2}}}\, e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)}\left(1 + e^{\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}\right).$$

The integral of $|g_B(t)|$ equals the infinite sum of $|I_k|$ which results in a geometric series.

$$\int_0^\infty |g_B(t)|\,dt = \int_0^{T_0} |g_B(t)|\,dt + \sum_{k=1}^{\infty}\int_{T_{k-1}}^{T_k} |g_B(t)|\,dt = \int_0^{T_0} |g_B(t)|\,dt + \sum_{k=0}^{\infty}|I_k| - I_0$$

Thereby, the first integral evaluates to

$$\int_0^{T_0} |g_B(t)|\,dt = 2\zeta e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)},$$

the geometric sum is

$$\sum_{k=0}^{\infty}|I_k| = 2\zeta\, e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)}\frac{1 + e^{\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}}{1 - e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}}$$

and hence

$$\|G_B(s)\|_{\mathcal{L}_1} = \int_0^\infty |g_B(t)|dt = 2\zeta e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)}\left(1+\frac{1+e^{\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}}{1-e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}}-\left(1+e^{\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}\right)\right)$$

$$\|G_B(s)\|_{\mathcal{L}_1} = 4\zeta\left(1-e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}\right)^{-1} e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)}. \tag{5.35}$$

For the strongly damped case ($\zeta > 1$), the transfer function is reformulated by means of partial fraction decomposition, which yields

$$G_B(s) = \frac{\zeta\omega_0}{\sqrt{\zeta^2-1}}\frac{\zeta+\sqrt{\zeta^2-1}}{s-\omega_0\left(\zeta+\sqrt{\zeta^2-1}\right)} - \frac{\zeta\omega_0}{\sqrt{\zeta^2-1}}\frac{\zeta-\sqrt{\zeta^2-1}}{s-\omega_0\left(\zeta-\sqrt{\zeta^2-1}\right)}.$$

Hence the impulse response is

$$g_B(t) = \frac{\zeta\omega_0\left(\zeta+\sqrt{\zeta^2-1}\right)}{\sqrt{\zeta^2-1}}e^{-\omega_0\left(\zeta+\sqrt{\zeta^2-1}\right)t} - \frac{\zeta\omega_0\left(\zeta-\sqrt{\zeta^2-1}\right)}{\sqrt{\zeta^2-1}}e^{-\omega_0\left(\zeta-\sqrt{\zeta^2-1}\right)t}$$

which can also be written in terms of hyperbolic functions

$$g_B(t) = \frac{2\zeta\omega_0}{\sqrt{\zeta^2-1}}e^{-\zeta\omega_0 t}\left[\sqrt{\zeta^2-1}\cosh\left(\sqrt{\zeta^2-1}\omega_0 t\right) - \zeta\sinh\left(\sqrt{\zeta^2-1}\omega_0 t\right)\right].$$

By use of addition theorems, we obtain

$$g_B(t) = \frac{2\zeta\omega_0}{\sqrt{\zeta^2-1}}e^{-\zeta\omega_0 t}\sinh\left[-\sqrt{\zeta^2-1}\omega_0 t + \operatorname{arctanh}\frac{\sqrt{\zeta^2-1}}{\zeta}\right].$$

Moreover, the impulse response function is positive in the interval $[0, T_0]$ and negative in the interval $[T_0, \infty]$ where

$$T_0 = \frac{\operatorname{arctanh}\frac{\sqrt{\zeta^2-1}}{\zeta}}{\sqrt{\zeta^2-1}\omega_0}.$$

Hence, we have

$$\int_0^\infty |g_B(t)|dt = \int_0^{T_0} g_B(t)dt - \int_{T_0}^\infty g_B(t)dt$$

and the $\mathcal{L}_1$-norm is

$$\|G_B(s)\|_{\mathcal{L}_1} = \int_0^\infty |g_B(t)|dt = 2\zeta e^{-\frac{\zeta}{\sqrt{\zeta^2-1}}\arctan\left(\frac{\sqrt{\zeta^2-1}}{\zeta}\right)} - \left(-2\zeta e^{-\frac{\zeta}{\sqrt{\zeta^2-1}}\arctan\left(\frac{\sqrt{\zeta^2-1}}{\zeta}\right)}\right)$$

$$\|G_B(s)\|_{\mathcal{L}_1} = 4\zeta e^{-\frac{\zeta}{\sqrt{\zeta^2-1}}\arctan\left(\frac{\sqrt{\zeta^2-1}}{\zeta}\right)} \tag{5.36}$$

For the critically damped case ($\zeta = 1$), the partial fraction decomposition of the transfer function is

$$G_B(s) = \frac{\omega_0}{s-\omega_0} - \frac{\omega_0^2}{(s-\omega_0)^2}$$

which yields

$$g_B(t) = e^{-\omega_0 t}(1-\omega_0 t)$$

in the time domain. Clearly the impulse response is positive for $t < \omega_0^{-1}$ and negative for $t > \omega_0^{-1}$ and hence

$$\int_0^\infty |g_B(t)|dt = \int_0^{\omega_0^{-1}} g_B(t)dt - \int_{\omega_0^{-1}}^\infty g_B(t)dt \ .$$

The $\mathcal{L}_1$-norm, finally is

$$\|G_B(s)\|_{\mathcal{L}_1} = 4e^{-1}.$$

Summing up,

$$\|G_B(s)\|_{\mathcal{L}_1} = \begin{cases} 4\zeta\left(1 - e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}}\right)^{-1} e^{-\frac{\zeta}{\sqrt{1-\zeta^2}}\arctan\left(\frac{\sqrt{1-\zeta^2}}{\zeta}\right)} & \text{for } 0 < \zeta < 1 \\ 4e^{-1} & \text{for } \zeta = 1 \\ 4\zeta e^{-\frac{\zeta}{\sqrt{\zeta^2-1}}\arctan\left(\frac{\sqrt{\zeta^2-1}}{\zeta}\right)} & \text{for } \zeta > 1. \end{cases} \quad (5.37)$$

Thrust Vector Control

The nonaffine controls (the thrust vector angles, Table 2.6) are not used within the linearizing state feedback but their effect is compensated for, using the affine controls, according to equation (5.23). As explained in remark 2 of section 4.3.3, the external controls are allowed to adopt any value within the set $\mathcal{U}$ without destroying the proof of ultimate boundedness of tracking- and parameter-estimation-error according to Theorem 4.4. It is hence reasonable to tune the nonaffine controls within a nonaffine-in-control framework, as introduced in section 4.5, while it is assured that the nonaffine controls do not leave $\mathcal{U}$.

The estimated effect of the thrust vectoring angles is described by a nonlinear map

$$\mathbf{w} = \hat{\mathbf{g}}(V, \alpha_{c,l}, \alpha_{c_r}, \alpha_{c,b}, \sigma_l, \sigma_r, \sigma_b, \kappa_b, U_{0l}, U_{0r}, U_{0b}) \quad (5.38)$$

where the map is linearly parameterized according to equation (5.23)

$$\hat{\mathbf{g}}(\cdot)=\left(\hat{\mathbf{K}}_N^T+\hat{\boldsymbol{\Theta}}_N^T\right)\boldsymbol{\varphi}_N\left(V,\alpha_{c,l},\alpha_{c_r},\alpha_{c,b},\sigma_l,\sigma_r,\sigma_b,\kappa_b,U_{0l},U_{0r},U_{0b}\right),$$

with assumed parameters $\hat{\mathbf{K}}_N$ and adaptive estimate $\hat{\boldsymbol{\Theta}}_N$. The nonaffine controls are, of course, the 4 thrust vectoring angles

$$\mathbf{u}_N^T=\begin{pmatrix}\sigma_l & \sigma_r & \sigma_b & \kappa_b\end{pmatrix}. \tag{5.39}$$

The motor input voltages are free to be prescribed from extern and could be used for speed control in the next outer loop. Hence, from the considered inner loop, they are considered as external and known disturbance. Additionally, the input voltages to all 3 engines will be equal, such that $U_{0l}=U_{0r}=U_{0b}=:U_0$ and hence, in compliance with notation of section 4.1.8, the external disturbance is $d=U_0$.

The remaining quantities, namely the aerodynamic velocity, AoA and AoS form the state dependent part of the regressors.

$$\mathbf{x}^T=\begin{pmatrix}V & \alpha & \beta\end{pmatrix}$$

However, the regressors have been designed, using the concept of included AoA. But, they depend on aerodynamic as well as thrust vectoring angles according to equations (2.47) – (2.50), (5.11).

$$\alpha_{c,l}=\alpha_{c,l}(\alpha,\beta,\sigma_l)\quad,\quad\alpha_{c,r}=\alpha_{c,r}(\alpha,\beta,\sigma_r)\quad,\quad\alpha_{c,b}=\alpha_{c,b}(\alpha,\beta,\sigma_b,\kappa_b)$$

Hence, the effect of thrust vector angle onto the dynamics is compactly written as

$$\mathbf{w}=\hat{\mathbf{g}}(\mathbf{x},\mathbf{u}_N,d)=\left(\hat{\mathbf{K}}_N^T+\hat{\boldsymbol{\Theta}}_N^T\right)\boldsymbol{\varphi}_N(\mathbf{x},\mathbf{u}_N,d). \tag{5.40}$$

Partitioning of Nonaffine Controls

Obviously, the output dimension of the nonaffine control map (5.40) is 3, while there are 4 thrust vectoring angles available for control. However (5.40) needs to be diffeomorphism w.r.t. the controls in order for the nonlinear-in-control algorithm to be exponentially stable, according to Theorem 4.5. This in turn requires that input and output dimensions are equal. In order to comply with this requirement, the nonaffine controls are restructured, based on their primary physical effects.

At first, consider the main engines. As illustrated in Figure 5.30 the main engine thrust only produces negative pitch moments due to their location relative to the aircraft c.g.. Also, yaw moments cannot be produced effectively, since the input voltages to both engines are chosen to be equal. Hence, the roll axis is the only one that can effectively be controlled by the main engines, using an antimetric deflection of left and right thrust vector inclination angle. We will therefore introduced a virtual thrust vector angle for

the main engines σ_m that effects a positive angle of the left engine and a negative angle of the right engine.

$$\sigma_l = \sigma_m \quad , \quad \sigma_r = -\sigma_m \tag{5.41}$$

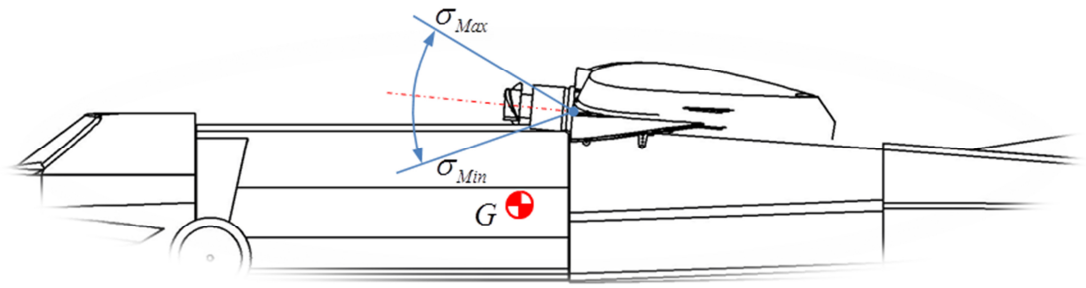


Figure 5.30 Main Propulsion Geometry

The back engine, however, effectively produces moments about pitch and yaw axes, not yet about the roll axis. Moreover, since the nondiagonal entry of the inertia tensor I_{xz} is small compared to I_x, I_y, I_z, the moments along the body-fixed axes approximately produce a time derivative of the respective body-fixed angular rate. We will hence use σ_m for production of roll moments and the backward thrust vectoring angles σ_b, κ_b to produce pitch and yaw moments. To this end, the nonaffine control map (5.23) is split up into a main engine part and a back engine part

$$\hat{\mathbf{g}}_m = \left(\hat{\mathbf{K}}_{N,m}^T + \hat{\mathbf{\Theta}}_{N,m}^T\right)\boldsymbol{\varphi}_{N,m}\left(V, \alpha_{c,l}, \alpha_{c,r}, \sigma_l, \sigma_r, U_0\right) \tag{5.42}$$

$$\hat{\mathbf{g}}_b = \left(\hat{\mathbf{K}}_{N,b}^T + \hat{\mathbf{\Theta}}_{N,b}^T\right)\boldsymbol{\varphi}_{N,b}\left(V, \alpha_{c,b}, \sigma_b, \kappa_b, U_0\right) \tag{5.43}$$

with the main engine regressor

$$\boldsymbol{\varphi}_{N,m}^T(\cdot) = \left(\boldsymbol{\varphi}_{N,l}^T\left(V, \alpha_{c,l}, \sigma_l, U_0\right) \quad \boldsymbol{\varphi}_{N,r}^T\left(V, \alpha_{c,r}, \sigma_r, U_0\right)\right),$$

their associated parameters

$$\hat{\mathbf{K}}_m^T = \left[\hat{\mathbf{K}}_l^T \quad \hat{\mathbf{K}}_R^T\right] \quad , \quad \hat{\mathbf{\Theta}}_{N,m}^T = \left[\hat{\mathbf{\Theta}}_l^T \quad \hat{\mathbf{\Theta}}_r^T\right].$$

and their analogs for the back engine $\boldsymbol{\varphi}_{N,b}$, $\hat{\mathbf{K}}_{N,b}$, $\hat{\mathbf{\Theta}}_{N,b}$. Hence, the first component of the virtual control $\mathbf{w}$, according to (5.38), is controlled by the main engines, while the remaining components are controlled by the back engine.

Now the thrust vector controls are partitioned as depicted in Figure 5.31. The desired virtual control $\mathbf{w}^*$ is separated into a roll axis part w_1^* and into a pitch-yaw axis part $\mathbf{w}_{(2,3)}^*$. The virtual thrust vectoring angle σ_m of the main engine is used to produce the desired roll axis control, while the back engine thrust vectoring angles σ_b, κ_b are used to produce the desired pitch and yaw axis control. Nevertheless, both, main and back engines produce moments in all 3 axes and hence the part that has been produced undesirably by one part is compensated for by the respective other part (dashed lines).

The deviation in the respective control effectiveness is finally tuned by the gradient based algorithm as introduced in section 4.5.

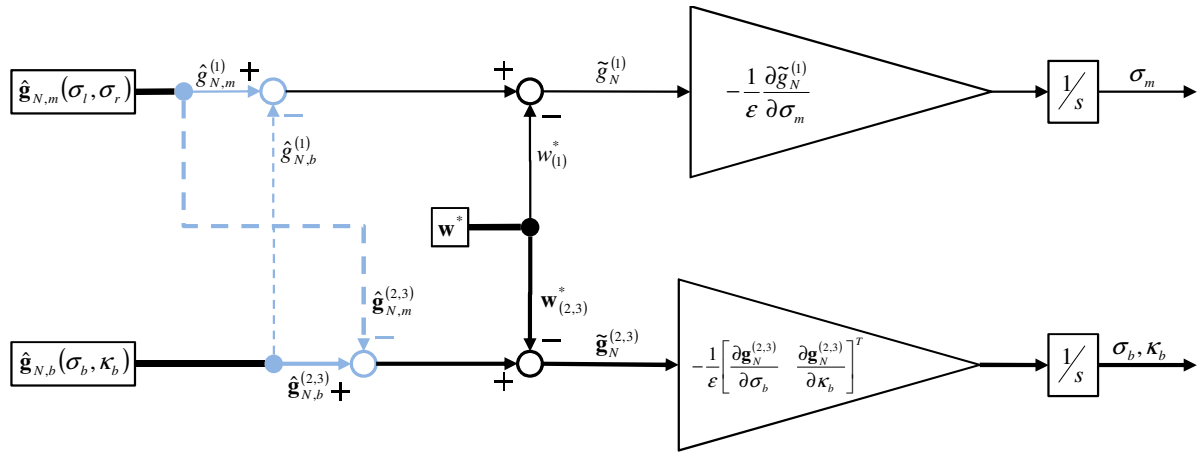


Figure 5.31 Partitioning of Thrust Vector Controls

Trim Value Assignment

The main reason, why variant 2 (section 4.6) is not suitable for the considered aircraft is the fact that the effectiveness of thrust vectoring is quite nonuniform across the flight envelope, particularly since thrust is decaying with increased aerodynamic velocity. Moreover, due to the disadvantageous positioning of the main engines relative to the aircraft c.g., the main engine cannot produce a pitch moment effectively. Furthermore, the thrust, produced by the back engine, is considerably smaller than the main engine thrust. Simulations discovered that the back engine thrust vanishes or even becomes negative in normal cruise flight conditions due to its insufficient power.

It is hence inappropriate to demand the aircraft to be trimmed by the thrust vector controls as proposed in section 4.1.8. The best, one can expect is that the propellers (at least partially) compensate for the moments that are produced by them, by setting $\mathbf{w}^*=\mathbf{0}$.

Computation of Control Map Jacobian

The computation of the Jacobians of (5.42) (5.43) need application of the chain rule, since the propulsion regressors are designed by use of the included AoA which implicitly depends on the thrust vector controls. According to (5.42), we have

$$\left[\frac{d\hat{\mathbf{g}}_m}{d\sigma_l}\quad\frac{d\hat{\mathbf{g}}_m}{d\sigma_r}\right]=\left(\hat{\mathbf{K}}_N^T+\hat{\mathbf{\Theta}}_N^T\right)\left[\frac{d\boldsymbol{\varphi}_{N,m}\left(V,\alpha_{c,l},\alpha_{c,r},\sigma_l,\sigma_r,U_0\right)}{d\sigma_l}\quad\frac{d\boldsymbol{\varphi}_{N,m}\left(V,\alpha_{c,l},\alpha_{c,r},\sigma_l,\sigma_r,U_0\right)}{d\sigma_r}\right]. \tag{5.44}$$

However, for the algorithm, illustrated in Figure 5.31, we need the derivative w.r.t. the virtual thrust vector inclination angle σ_m.

$$\frac{d\hat{\mathbf{g}}_m}{d\sigma_m}=\frac{d\hat{\mathbf{g}}_m}{d\sigma_l}\frac{d\sigma_l}{d\sigma_m}+\frac{d\hat{\mathbf{g}}_m}{d\sigma_r}\frac{d\sigma_r}{d\sigma_m}=\frac{d\hat{\mathbf{g}}_m}{d\sigma_l}-\frac{d\hat{\mathbf{g}}_m}{d\sigma_r}$$

$$\frac{d\hat{\mathbf{g}}_m}{d\sigma_m}=\left(\hat{\mathbf{K}}_N^T+\hat{\mathbf{\Theta}}_N^T\right)\left(\frac{d\boldsymbol{\varphi}_{N,m}\left(V,\alpha_{c,l},\alpha_{c,r},\sigma_l,\sigma_r,U_0\right)}{d\sigma_l}-\frac{d\boldsymbol{\varphi}_{N,m}\left(V,\alpha_{c,l},\alpha_{c,r},\sigma_l,\sigma_r,U_0\right)}{d\sigma_r}\right) \tag{5.45}$$

According to (5.43), the Jacobian for the back engine is

$$\left[\frac{d\hat{\mathbf{g}}_{N,b}}{d\sigma_b}\quad\frac{d\hat{\mathbf{g}}_{N,b}}{d\kappa_b}\right]=\left(\hat{\mathbf{K}}_N^T+\hat{\mathbf{\Theta}}_N^T\right)\left[\frac{d\boldsymbol{\varphi}_{N,b}\left(V,\alpha_{c,b},\sigma_b,\kappa_b,U_0\right)}{d\sigma_b}\quad\frac{d\boldsymbol{\varphi}_{N,b}\left(V,\alpha_{c,b},\sigma_b,\kappa_b,U_0\right)}{d\kappa_b}\right] \tag{5.46}$$

The included AoA $\alpha_{c,l}$, $\alpha_{c,r}$, $\alpha_{c,b}$ directly depend on the respective total AoA $\alpha_{T,l}$, $\alpha_{T,r}$, $\alpha_{T,b}$ and AoS $\beta_{T,l}$, $\beta_{T,r}$, $\beta_{T,b}$ according to (5.11) which in turn depend on the thrust vectoring angles by (2.47) – (2.50). Hence, the derivatives of the regressors w.r.t. thrust vector angles evaluate to

$$\frac{d\boldsymbol{\varphi}_{N,m}\left(V,\alpha_{c,l},\alpha_{c,r},\sigma_l,\sigma_r,U_0\right)}{d\sigma_l}=\frac{\partial\boldsymbol{\varphi}_{N,m}}{\partial\alpha_{c,l}}\left[\frac{\partial\alpha_{c,l}}{\partial\alpha_{T,l}}\frac{\partial\alpha_{T,l}}{\partial\sigma_l}+\frac{\partial\alpha_{c,l}}{\partial\beta_{T,l}}\frac{\partial\beta_{T,l}}{\partial\sigma_l}\right]+\frac{\partial\boldsymbol{\varphi}_{N,m}}{\partial\sigma_l} \tag{5.47}$$

$$\frac{d\boldsymbol{\varphi}_{N,m}\left(V,\alpha_{c,l},\alpha_{c,r},\sigma_l,\sigma_r,U_0\right)}{d\sigma_r}=\frac{\partial\boldsymbol{\varphi}_{N,m}}{\partial\alpha_{c,r}}\left[\frac{\partial\alpha_{c,r}}{\partial\alpha_{T,r}}\frac{\partial\alpha_{T,r}}{\partial\sigma_r}+\frac{\partial\alpha_{c,r}}{\partial\beta_{T,r}}\frac{\partial\beta_{T,r}}{\partial\sigma_r}\right]+\frac{\partial\boldsymbol{\varphi}_{N,m}}{\partial\sigma_r} \tag{5.48}$$

$$\frac{d\boldsymbol{\varphi}_{N,b}\left(V,\alpha_{c,b},\sigma_b,\kappa_b,U_0\right)}{d\sigma_b}=\frac{\partial\boldsymbol{\varphi}_{N,m}}{\partial\alpha_{c,b}}\left[\frac{\partial\alpha_{c,b}}{\partial\alpha_{T,b}}\frac{\partial\alpha_{T,b}}{\partial\sigma_b}+\frac{\partial\alpha_{c,b}}{\partial\beta_{T,b}}\frac{\partial\beta_{T,b}}{\partial\sigma_b}\right]+\frac{\partial\boldsymbol{\varphi}_{N,b}}{\partial\sigma_b} \tag{5.49}$$

$$\frac{d\boldsymbol{\varphi}_{N,b}\left(V,\alpha_{c,b},\sigma_b,\kappa_b,U_0\right)}{d\kappa_b}=\frac{\partial\boldsymbol{\varphi}_{N,m}}{\partial\alpha_{c,b}}\left[\frac{\partial\alpha_{c,b}}{\partial\alpha_{T,b}}\frac{\partial\alpha_{T,b}}{\partial\kappa_b}+\frac{\partial\alpha_{c,b}}{\partial\beta_{T,b}}\frac{\partial\beta_{T,b}}{\partial\kappa_b}\right]+\frac{\partial\boldsymbol{\varphi}_{N,b}}{\partial\kappa_b} \tag{5.50}$$

Thereby, the expressions, involved in the main propulsion regressors evaluate to

$$\frac{\partial\alpha_{c,i}}{\partial\alpha_{T,i}}=\frac{\sin\alpha_{T,i}\cos\beta_{T,i}}{\sqrt{1-\cos^2\alpha_{T,i}\cos^2\beta_{T,i}}}\quad,\quad\frac{\partial\alpha_{c,i}}{\partial\beta_{T,i}}=\frac{\cos\alpha_{T,i}\sin\beta_{T,i}}{\sqrt{1-\cos^2\alpha_{T,i}\cos^2\beta_{T,i}}} \tag{5.51}$$

$$\frac{\partial\alpha_{T,i}}{\partial\sigma_i}=1\qquad\frac{\partial\beta_{T,i}}{\partial\sigma_i}=0$$

$$\frac{\partial\boldsymbol{\varphi}_{N,m}}{\partial\alpha_{c,l}}=\begin{pmatrix}\frac{\partial\boldsymbol{\varphi}_{N,l}}{\partial\alpha_{c,l}}\\ \frac{\partial\boldsymbol{\varphi}_{N,r}}{\partial\alpha_{c,l}}\end{pmatrix}\qquad\frac{\partial\boldsymbol{\varphi}_{N,m}}{\partial\alpha_{c,r}}=\begin{pmatrix}\frac{\partial\boldsymbol{\varphi}_{N,l}}{\partial\alpha_{c,r}}\\ \frac{\partial\boldsymbol{\varphi}_{N,r}}{\partial\alpha_{c,r}}\end{pmatrix}$$

$$\frac{\partial\boldsymbol{\varphi}_{N,l}}{\partial\alpha_{c,l}}=\begin{pmatrix}\sin\sigma_l\frac{\partial\phi_{l,1}}{\partial\alpha_{c,l}}\\ \cos\sigma_l\frac{\partial\phi_{l,1}}{\partial\alpha_{c,l}}\\ \vdots\\ \sin\sigma_l\frac{\partial\phi_{l,s_l}}{\partial\alpha_{c,l}}\\ \cos\sigma_l\frac{\partial\phi_{l,s_l}}{\partial\alpha_{c,l}}\end{pmatrix}\qquad\frac{\partial\boldsymbol{\varphi}_{N,l}}{\partial\alpha_{c,r}}=\mathbf{0}\qquad\frac{\partial\boldsymbol{\varphi}_{N,r}}{\partial\alpha_{c,r}}=\begin{pmatrix}\sin\sigma_r\frac{\partial\phi_{r,1}}{\partial\alpha_{c,r}}\\ \cos\sigma_r\frac{\partial\phi_{r,1}}{\partial\alpha_{c,r}}\\ \vdots\\ \sin\sigma_r\frac{\partial\phi_{r,s_r}}{\partial\alpha_{c,r}}\\ \cos\sigma_r\frac{\partial\phi_{r,s_r}}{\partial\alpha_{c,r}}\end{pmatrix}\qquad\frac{\partial\boldsymbol{\varphi}_{N,r}}{\partial\alpha_{c,l}}=\mathbf{0}$$

$$\frac{\partial \boldsymbol{\varphi}_m}{\partial \sigma_l} = \begin{pmatrix} \frac{\partial \boldsymbol{\varphi}_{N,l}}{\partial \sigma_l} \\ \frac{\partial \boldsymbol{\varphi}_{N,r}}{\partial \sigma_l} \end{pmatrix} \qquad \frac{\partial \boldsymbol{\varphi}_m}{\partial \sigma_r} = \begin{pmatrix} \frac{\partial \boldsymbol{\varphi}_{N,l}}{\partial \sigma_r} \\ \frac{\partial \boldsymbol{\varphi}_{N,r}}{\partial \sigma_r} \end{pmatrix}$$

$$\frac{\partial \boldsymbol{\varphi}_{N,l}}{\partial \sigma_l} = \begin{pmatrix} \cos\sigma_l \cdot \phi_{l,1} \\ -\sin\sigma_l \cdot \phi_{l,1} \\ \vdots \\ \cos\sigma_l \cdot \phi_{l,s_l} \\ -\sin\sigma_l \cdot \phi_{l,s_l} \end{pmatrix} \qquad \frac{\partial \boldsymbol{\varphi}_{N,l}}{\partial \sigma_r} = \mathbf{0} \qquad \frac{\partial \boldsymbol{\varphi}_{N,r}}{\partial \sigma_r} = \begin{pmatrix} \cos\sigma_r \cdot \phi_{r,1} \\ -\sin\sigma_r \cdot \phi_{r,1} \\ \vdots \\ \cos\sigma_r \cdot \phi_{r,s_r} \\ -\sin\sigma_r \cdot \phi_{r,s_r} \end{pmatrix} \qquad \frac{\partial \boldsymbol{\varphi}_{N,r}}{\partial \sigma_l} = \mathbf{0}$$

Notice that the index i is replaced by l (left engine) or r (right engine). For the back engine regressor, we have

$$\frac{\partial \alpha_{c,b}}{\partial \alpha_{T,b}} = \frac{\sin\alpha_{T,b}\cos\beta_{T,b}}{\sqrt{1-\cos^2\alpha_{T,b}\cos^2\beta_{T,b}}} \quad , \quad \frac{\partial \alpha_{c,b}}{\partial \beta_{T,b}} = \frac{\cos\alpha_{T,b}\sin\beta_{T,b}}{\sqrt{1-\cos^2\alpha_{T,b}\cos^2\beta_{T,b}}} \tag{5.52}$$

$$\frac{\partial \alpha_{T,b}}{\partial \sigma_b} = \frac{\cos\beta\left[\cos(\alpha+\sigma_b)\sin\beta\sin\kappa_b + \cos\beta\cos\kappa_b\right]}{\left[\sin(\alpha+\sigma_b)\cos\beta\right]^2 + \left[\sin\beta\sin\kappa_b + \cos\kappa_b\cos\beta\cos(\alpha+\sigma_b)\right]^2} \tag{5.53}$$

$$\frac{\partial \beta_{T,b}}{\partial \sigma_b} = \frac{\cos\beta\sin\kappa_b\sin(\alpha+\sigma_b)}{\sqrt{1-\left[\cos\kappa_b\sin\beta - \cos\beta\sin\kappa_b\cos(\alpha+\sigma_b)\right]^2}} \tag{5.54}$$

$$\frac{\partial \alpha_{T,b}}{\partial \kappa_b} = -\frac{\sin(\alpha+\sigma_b)\cos\beta\left[\cos\kappa_b\sin\beta - \cos\beta\sin\kappa_b\cos(\alpha+\sigma_b)\right]}{\left[\sin(\alpha+\sigma_b)\cos\beta\right]^2 + \left[\sin\beta\sin\kappa_b + \cos\beta\cos\kappa_b\cos(\alpha+\sigma_b)\right]^2} \tag{5.55}$$

$$\frac{\partial \beta_{T,b}}{\partial \kappa_b} = -\frac{\sin\beta\sin\kappa_b + \cos\beta\cos\kappa_b\cos(\alpha+\sigma_b)}{\sqrt{1-\left[\cos\kappa_b\sin\beta - \cos\beta\sin\kappa_b\cos(\alpha+\sigma_b)\right]^2}} \tag{5.56}$$

$$\frac{\partial \boldsymbol{\varphi}_{N,b}}{\partial \alpha_{c,b}} = \begin{pmatrix} \cos\kappa_b\sin\sigma_b\frac{\partial \phi_{b,1}}{\partial \alpha_{c,b}} \\ \cos\kappa_b\cos\sigma_b\frac{\partial \phi_{b,1}}{\partial \alpha_{c,b}} \\ \sin\kappa_b\frac{\partial \phi_{b,1}}{\partial \alpha_{c,b}} \\ \vdots \\ \cos\kappa_b\sin\sigma_b\frac{\partial \phi_{b,s_b}}{\partial \alpha_{c,b}} \\ \cos\kappa_b\cos\sigma_b\frac{\partial \phi_{b,s_b}}{\partial \alpha_{c,b}} \\ \sin\kappa_b\frac{\partial \phi_{b,s_b}}{\partial \alpha_{c,b}} \end{pmatrix} \qquad \frac{\partial \boldsymbol{\varphi}_{N,b}}{\partial \sigma_b} = \begin{pmatrix} \cos\kappa_b\cos\sigma_b\phi_{b,1} \\ -\cos\kappa_b\sin\sigma_b\phi_{b,1} \\ 0 \\ \vdots \\ \cos\kappa_b\cos\sigma_b\phi_{b,s_b} \\ -\cos\kappa_b\sin\sigma_b\phi_{b,s_b} \\ 0 \end{pmatrix} \qquad \frac{\partial \boldsymbol{\varphi}_{N,b}}{\partial \kappa_b} = \begin{pmatrix} -\sin\kappa_b\sin\sigma_b\phi_{b,1} \\ -\sin\kappa_b\cos\sigma_b\phi_{b,1} \\ \cos\kappa_b\phi_{b,1} \\ \vdots \\ -\sin\kappa_b\sin\sigma_b\phi_{b,s_b} \\ -\sin\kappa_b\cos\sigma_b\phi_{b,s_b} \\ \cos\kappa_b\phi_{b,s_b} \end{pmatrix}$$

The partial derivatives of radial basis functions $\phi_{i,j}$ are computed according to (5.8). In normal flight conditions, AoA and AoS do not adopt values beyond 60° magnitude and, with regard to the maximum thrust vector angle deflections in Table 2.6, the following conditions are obviously fulfilled.

$$|\alpha+\sigma_l|<\frac{\pi}{2}\ ,\ |\alpha+\sigma_r|<\frac{\pi}{2}\ ,\ |\alpha+\sigma_b|<\frac{\pi}{2}\ ,\ |\beta-\kappa_b|<\frac{\pi}{2}\ ,\ |\alpha|<\frac{\pi}{2}\ ,\ |\beta|<\frac{\pi}{2}. \qquad (5.57)$$

It can be shown that the expressions, derived recently, are free of singularities or, at least, can be extended continuously. To this end, take a closer look to expressions (5.53) – (5.56) which have singularities if the denominator has a zero. The common denominator of equations (5.53) and (5.55) contains two squared summands. The only zero of the first term is $\alpha+\sigma_b=0$, however, the second summand is greater than zero under this condition, since by application of addition theorems, the expression equals

$$\sin(\beta)\sin\kappa_b+\cos(\beta)\cos\kappa_b=\cos(\beta-\kappa_b)>0.$$

For an analogous consideration of the denominators of equations (5.54) and (5.56), we apply addition theorems to the trigonometric functions and take the absolute value.

$$\begin{aligned}&|\cos\kappa_b\sin\beta-\cos\beta\sin\kappa_b\cos(\alpha+\sigma_b)|\\&\quad=\sqrt{\cos^2\kappa_b+\cos^2(\alpha+\sigma_b)\sin^2\kappa_b}\,|\sin[\beta-\arctan(\cos(\alpha+\sigma_b)\tan\kappa_b)]|.\end{aligned}$$

Then, since $0\le\cos(\alpha+\sigma_b)\le1$ we have

$$\sqrt{\cos^2\kappa_b+\cos^2(\alpha+\sigma_b)\sin^2\kappa_b}\le\sqrt{\cos^2\kappa_b+\sin^2\kappa_b}=1.$$

which implies

$$|\cos\kappa_b\sin\beta-\cos\beta\sin\kappa_b\cos(\alpha+\sigma_b)|\le|\sin[\beta-\arctan(\cos(\alpha+\sigma_b)\tan\kappa_b)]|$$

Moreover, since $0\le\cos(\alpha+\sigma_b)\le1$, we have

$$\begin{aligned}0\le\arctan(\cos(\alpha+\sigma_b)\tan\kappa_b)\le\kappa_b\ \text{ for }\ \kappa_b\ge0\\0\ge\arctan(\cos(\alpha+\sigma_b)\tan\kappa_b)\ge\kappa_b\ \text{ for }\ \kappa_b\le0.\end{aligned}$$

This result, together with

$$-\frac{\pi}{2}<\beta-\kappa_b\le\sin[\beta-\arctan(\cos(\alpha+\sigma_b)\tan\kappa_b)]\le\beta<\frac{\pi}{2}\ \text{ for }\ \kappa_b\ge0$$

$$-\frac{\pi}{2}<\beta\le\sin[\beta-\arctan(\cos(\alpha+\sigma_b)\tan\kappa_b)]\le\beta-\kappa_b<\frac{\pi}{2}\ \text{ for }\ \kappa_b\le0$$

we have

$$|\beta-\arctan(\cos(\alpha+\sigma_b)\tan\kappa_b)|<\frac{\pi}{2}.$$

and consequently

$$|\cos\kappa_b\sin\beta-\cos\beta\sin\kappa_b\cos(\alpha+\sigma_b)|\le\sin|[\beta-\arctan(\cos(\alpha+\sigma_b)\tan\kappa_b)]|<1$$

and hence, expressions (5.54) and (5.56) are free of singularities.

Expressions (5.51), (5.52) actually have a singularity for $\alpha_{T,i}=\beta_{T,i}=0$, but it can be shown, by L'Hospital's rule that

$$\lim_{\substack{\beta_{T,i}\to 0\\ \alpha_{T,i}\to 0}}\left(\frac{\sin\alpha_{T,i}\cos\beta_{T,i}}{\sqrt{1-\cos^2\alpha_{T,i}\cos^2\beta_{T,i}}}\right)=1 \quad , \quad \lim_{\substack{\beta_{T,i}\to 0\\ \alpha_{T,i}\to 0}}\left(\frac{\cos\alpha_{T,i}\sin\beta_{T,i}}{\sqrt{1-\cos^2\alpha_{T,i}\cos^2\beta_{T,i}}}\right)=1$$

and hence, these expressions can be continuously extended by

$$\alpha_{c,i}=\begin{cases}\dfrac{\sin\alpha_{T,i}\cos\beta_{T,i}}{\sqrt{1-\cos^2\alpha_{T,i}\cos^2\beta_{T,i}}} & \text{if } \alpha_{T,i}\neq 1 \text{ and } \beta_{T,i}\neq 1\\ 1 & \text{if } \alpha_{T,i}=1 \text{ or } \beta_{T,i}=1.\end{cases} \tag{5.58}$$

5.3.5 Adaptive Gain Design

For design of the baseline controller, namely reference- and error dynamics, any classical design method could be utilized and is considered state of the technology. Therefore, this section will focus on the design of the adaptive controller parameters. These are learning rates γ_x, γ_N, γ_L, gains of switching σ-modification σ_x, σ_N, σ_L and MMQ gain κ:

Gain Design Procedure

As derived in section 4.4.3, inequalities (4.243) have to be satisfied for ultimate boundedness. Assuming that the bounds on true parameters $\theta_{x,max}$, $\theta_{N,max}$ and $\lambda_{L,max}$ as well as the bound on the unmatched uncertainty D are known, there are some minimal values $k_e^*, k_x^*, k_N^*, k_L^* > 0$ (subsequently denoted as k_i^*, $i=e,x,N,L$) such that (4.243) has a solution for all $k_i \geq k_i^*$. In order to find the minimum values, we divide the problem into two parts. For the first part, we assume that $k_i = k_{i,k}$ are fixed in the k^{th} iteration step, while the independent variables of (4.243) , σ_x, σ_N, σ_L, κ, K_0, are considered as optimization variables.

Optimization 1

minimize: $f_1(\mathbf{x})=\mathbf{x}^T\mathbf{x}$

$$\mathbf{x}^T=(\sigma_x \quad \sigma_N \quad \sigma_L \quad K_0 \quad \kappa)$$

subjected to the nonlinear constraint

$$\mathbf{A}(\mathbf{p}_k)\mathbf{x}-\mathbf{b}_k(\mathbf{p}_k) < 0$$
$$-\mathbf{x} < 0$$

$$\mathbf{A}(\mathbf{p}_k)=\begin{bmatrix}\theta_{x,\max}^2 & \theta_{N,\max}^2 & \lambda_{L,\max}^2 & 2 & k_Q\\ -2k_{x,k} & \theta_{N,\max}^2 & \lambda_{L,\max}^2 & 2 & k_Q\\ \theta_{x,\max}^2 & -2k_{N,k} & \lambda_{L,\max}^2 & 2 & k_Q\\ \theta_{x,\max}^2 & \theta_{N,\max}^2 & -2k_{L,k} & 2 & k_Q\end{bmatrix} \quad \mathbf{b}(\mathbf{p}_k)=\begin{pmatrix}2k_{e,k}\\ -2V_D\\ -2V_D\\ -2V_D\end{pmatrix} \quad k_Q=(mLD)^2N \quad V_D=\frac{\left(\bar{\lambda}_{P_E}D\right)^2}{\underline{\lambda}_{Q_E}}$$

$$\mathbf{p}_k^T=(k_{e,k} \quad k_{x,k} \quad k_{N,k} \quad k_{L,k})$$

This is clearly a convex optimization problem, since both, objective function and inequality constraint, are convex functions and hence, if the problem is feasible, it has a unique local minimum, which is also the global minimum. Such convex optimization problems can efficiently be solved numerically, using various algorithms available in literature ([Boy09], [Smi10], [Rao09]).

So far, by convex optimization, we find whether inequalities (4.243) have a solution for some fixed $\mathbf{p}_k$. In a second step, we apply the bisection method ([Pre07]) to find the minimum values $\mathbf{p}^*$ that render the optimization problem feasible. Bisection is actually a method to find zeros of a function within some interval. By a slight abuse, we will use bisection to find – instead of a zero – the limit value $\mathbf{p}^*$ between feasibility and non-feasibility of optimization 1. Therefore, we start with some $k_{e,0}$, $k_{x,0}$, $k_{N,0}$, $k_{L,0}$,>0, sufficiently large such that optimization 1 is feasible. Moreover, since the minima k_i^* are independent of each other, the minimization of a single entry in $\mathbf{p}$ can be done independently of the others. Therefore, the procedure is explained exemplarily for k_e^*. In the initial iteration, it is known that k_e^* lies in the interval

$$I_0 = [k_{e,L,0} \quad k_{e,H,0}] \text{ where } k_{e,L,0} = 0 \text{ , } k_{e,H,0} = k_{e,0} .$$

In order to confine the interval, optimization 1 is tested in the middle of the interval.

$$k_{e,M,0} = \frac{k_{eH,0} + k_{e,L,0}}{2}$$

If it is feasible at this point, k_e^* lies in the left part of the interval, and we choose for the next iteration step $I_1 = [k_{e,L,0} \quad k_{e,M,0}]$, otherwise $I_1 = [k_{e,M,0} \quad k_{e,H,0}]$. This procedure is repeated until the interval is sufficiently small. Hence, for the k^{th} iteration step, we choose

- $I_{k+1} = [k_{e,L,k} \quad k_{e,M,k}]$ if optimization 1 is feasible at $k_{e,M,k}$ and
- $I_{k+1} = [k_{e,M,k} \quad k_{e,H,k}]$ otherwise.

Hence, in each iteration step, the interval that brackets k_e^* is bisected as depicted in Figure 5.32. Finally, in the last – let us say the N^{th} – iteration, we choose $k_e^* = k_{e,H,N}$. Now, if we have found $\mathbf{p}^*$, the minimum tracking error $\underline{e}_e$ is obtained by equation (4.174). Now, if possible, we choose some r_e that simultaneously fulfills (4.151) and (4.178), i.e.

$$\underline{e}_e \le r_e \le \overline{\zeta} - \zeta_R \qquad (5.59)$$

which allows computation of the minimum learning rates $\underline{\gamma}_x$, $\underline{\gamma}_N$, $\underline{\gamma}_L$ according to equations (4.175) – (4.177).

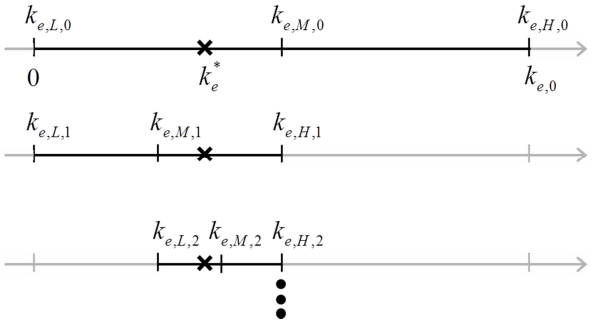


Figure 5.32 Bisection Method

With the lower bound on the maximum tracking error, we can now compute the learning rates as well as modification gains that optimize a suitable objective. From a control system point of view, it is desirable that the transient phase is as short as possible and the ultimate bound is as small as possible. Thereby the behavior of the parameter-estimation-errors is of secondary interest since control performance is tied to the tracking error. Referring to Corollary 4.2, we want to minimize time and ultimate bound of the tracking error T_e and b_e. Notice that T_e depends on the initial condition bound δ_e but, in order to get a worst case bound, we insert the maximum allowable bound ρ_e (equation (4.155)) instead. The optimization problem is summarized on the subsequent page. The objective function is chosen to be a linear combination of transient time T_e and ultimate bound b_e, weighted with a_1, $a_2>0$. The choice of the weights, of course, allows to emphasis either of the objectives. Notice that γ_x, γ_N, γ_L are additionally bounded from above by some $\gamma_{x,max}$, $\gamma_{N,max}$, $\gamma_{L,max}$. This is necessary since high learning rates are suitable to reduce the objective function, however, it is well-known that high gain adaptation results in high frequency oscillation and also could cause problems due to the discretization error that always occur in a digital control systems. In practice, the upper bounds are either set to some fixed values or to a multiple of the minimum learning rates, i.e.

$$\gamma_{x,\max} = c_x \underline{\gamma}_x \quad , \quad \gamma_{N,\max} = c_N \underline{\gamma}_N \quad , \quad \gamma_{L,\max} = c_L \underline{\gamma}_L \text{ with constants } b_x, b_N, b_L \geq 1 .$$

Contrary to optimization 1, this is a general nonlinear optimization problem with potentially several local minima. Since we do not have an initial guess for the parameters, a classical gradient based optimization algorithm potentially only finds a local minimum but not the global. Therefore, it is recommendable to use rather modern optimization algorithms that are able to find the global minimum disregarded the initial guess. For the gain design a genetic algorithm (e.g. [Rao09]) has been chosen. Despite the statistical character of the genetic algorithm, the computations produced quite repeatable results for the adaptive gains. Figure 5.33 illustrates the whole design procedure for the adaptive gains.

Optimization 2

minimize: $f_2(\mathbf{x},\mathbf{y}) = a_1 T_e(\mathbf{x},\mathbf{y}) + a_2 b_e(\mathbf{x},\mathbf{y})$

$$\mathbf{x}^T = (\sigma_x \quad \sigma_N \quad \sigma_L \quad K_0 \quad \kappa) \qquad \mathbf{y}^T = (\gamma_x \quad \gamma_N \quad \gamma_L)$$

subjected to the nonlinear constraint

$$\begin{aligned} \mathbf{A}\mathbf{x} - \mathbf{b} &< 0 \\ -\mathbf{x} &< 0 \\ -\mathbf{y} &< \underline{\mathbf{y}} \\ \mathbf{y} &< \overline{\mathbf{y}} \end{aligned}$$

$$\underline{\mathbf{y}}^T = (\underline{\gamma}_x \quad \underline{\gamma}_N \quad \underline{\gamma}_L) \qquad \overline{\mathbf{y}}^T = (\gamma_{x,\max} \quad \gamma_{N,\max} \quad \gamma_{L,\max})$$

$$\mathbf{A} = \begin{bmatrix} \theta_{x,\max}^2 & \theta_{N,\max}^2 & \lambda_{L,\max}^2 & 2 & k_Q \\ -2k_x & \theta_{N,\max}^2 & \lambda_{L,\max}^2 & 2 & k_Q \\ \theta_{x,\max}^2 & -2k_N & \lambda_{L,\max}^2 & 2 & k_Q \\ \theta_{x,\max}^2 & \theta_{N,\max}^2 & -2k_L & 2 & k_Q \end{bmatrix} \qquad \mathbf{b} = \begin{pmatrix} 2k_e \\ -2V_D \\ -2V_D \\ -2V_D \end{pmatrix} \qquad k_Q = (mLD)^2 N \quad V_D = \frac{\left(\overline{\lambda}_{P_E} D\right)^2}{\underline{\lambda}_{Q_E}}$$

$$k_e = \frac{\underline{\lambda}_{Q_E}\underline{\lambda}_{P_E}}{4\overline{\lambda}_{P_E}} r_e^2 - \sqrt{\underline{\lambda}_{P_E}\overline{\lambda}_{P_E}}\, D r_e \qquad k_x = \frac{\gamma_x \underline{\lambda}_{P_E}}{2} r_e^2 - \sqrt{\gamma_x \underline{\lambda}_{P_E}}\, \theta_{x,\max} r_e$$

$$k_N = \frac{\gamma_N \underline{\lambda}_{P_E}}{2} r_e^2 - \sqrt{\gamma_N \underline{\lambda}_{P_E}}\, \theta_{N,\max} r_e \qquad k_L = \frac{\gamma_L \underline{\lambda}_{P_E}}{2} r_e^2 - \sqrt{\gamma_L \underline{\lambda}_{P_E}}\, \lambda_{L,\max} r_e$$

$$T_e = \max\left[0, K_0^{-1}\left(\rho_e^2 + 3/4 \cdot \overline{\lambda}_{P_E} r_e^2 - \nu\right)\right] \qquad b_e = \sqrt{\underline{\lambda}_{P_E}^{-1}\nu}$$

$$\rho_e = \sqrt{\frac{\underline{\lambda}_{P_E}}{\overline{\lambda}_{P_E}}}\frac{r_e}{2} \qquad \nu = \overline{\lambda}_{P_E}\mu_e^2 + \gamma_x^{-1}\mu_x^2 + \gamma_N^{-1}\mu_N^2 + \gamma_L^{-1}\mu_L^2$$

$$\mu_e = \frac{\overline{\lambda}_{P_E} D}{\underline{\lambda}_{Q_E}} + \sqrt{\left(\frac{\overline{\lambda}_{P_E} D}{\underline{\lambda}_{Q_E}}\right)^2 + \frac{V_e + V_Q + K_0}{\underline{\lambda}_{Q_E}}} \qquad \mu_x = \frac{\theta_{x,\max}}{2} + \sqrt{\left(\frac{\theta_{x,\max}}{2}\right)^2 + \frac{V_D + V_x + V_Q + K_0}{2\sigma_x}}$$

$$\mu_N = \frac{\theta_{N,\max}}{2} + \sqrt{\left(\frac{\theta_{N,\max}}{2}\right)^2 + \frac{V_D + V_N + V_Q + K_0}{2\sigma_N}} \qquad \mu_L = \frac{\lambda_{L,\max}}{2} + \sqrt{\left(\frac{\lambda_{L,\max}}{2}\right)^2 + \frac{V_D + V_L + V_Q + K_0}{2\sigma_L}}$$

$$V_e = \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_N\theta_{N,\max}^2 + \sigma_L\lambda_{L,\max}^2\right) \qquad V_x = \frac{1}{2}\left(\sigma_N\theta_{N,\max}^2 + \sigma_L\lambda_{L,\max}^2\right)$$

$$V_N = \frac{1}{2}\left(\sigma_x\theta_{x,\max}^2 + \sigma_L\lambda_{L,\max}^2\right) \qquad V_L = \frac{1}{2}\left(\sigma_x\theta_x^2 + \sigma_N\theta_{N,\max}^2\right) \qquad V_Q = \frac{1}{2}\kappa(mLD)^2 N$$

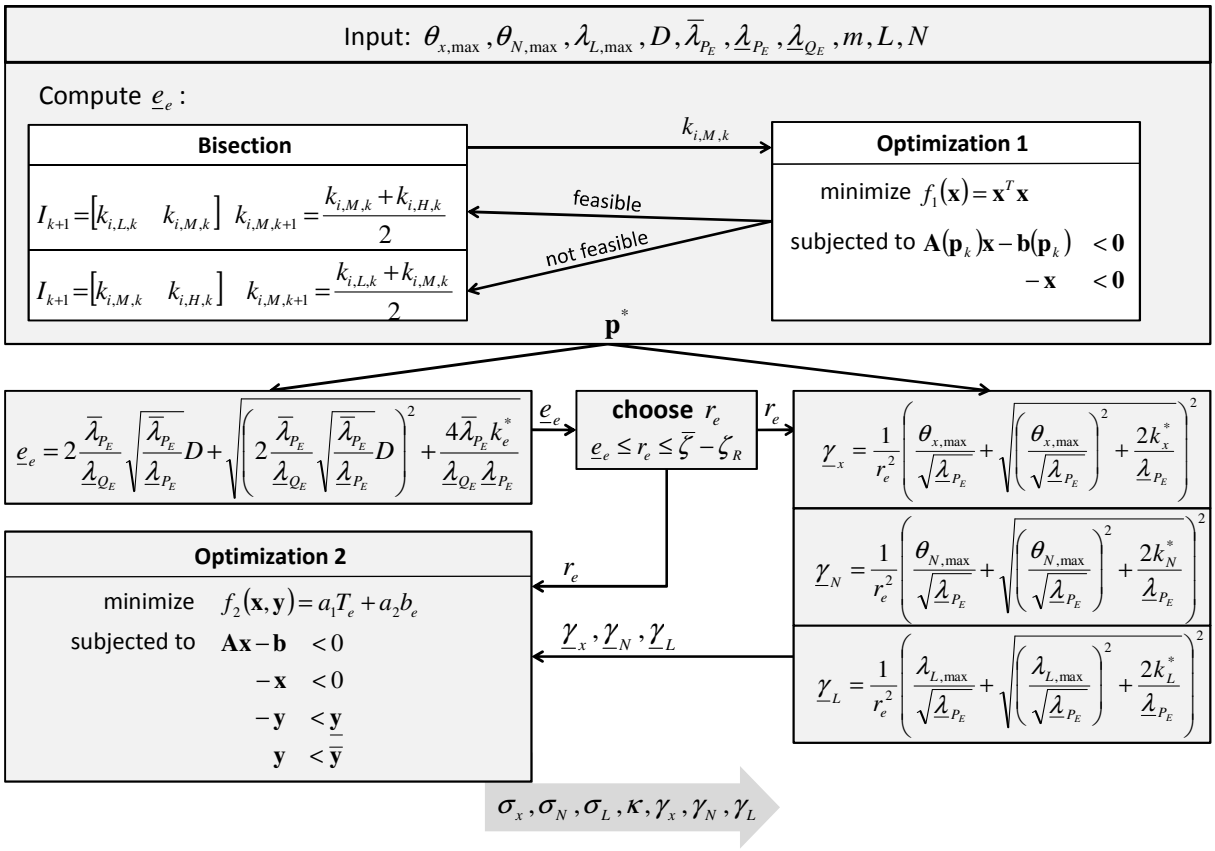


Figure 5.33 Design Procedure for Adaptive Gains

Discussion on Practical Issues

Table 5.3 shows some numeric examples of the gain design procedure that highlight its general behavior as discussed in the following.

It reveals that the lower bound on the tracking error $\underline{e}_e$ is mainly dependent on unmatched uncertainty, rather not on the parameter bounds. This is not surprising since the unmatched uncertainty cannot be cancelled by the adaptation and hence the unmatched uncertainty still remains when adaptation already has taken place and excites the error dynamics (refer to equation (4.127)). Furthermore, large parameter bounds imply a large parameter space, the adaptation should be able to compensate for. Initially, one could think that larger parameter bounds should increase $\underline{e}_e$ since it takes more effort to cross a larger parameter space. However, this effect can be compensated by larger learning rates. This also explains why the minimum learning rates are increased with the parameter bounds.

Table 5.3 Gain Design Numeric Examples

fixed parameters:					
$\overline{\lambda}_{P_E} = 1.25 \cdot 10^{-2}$ $\underline{\lambda}_{P_E} = 6.25 \cdot 10^{-3}$ $\underline{\lambda}_{Q_E} = 1$ $r_e = 40°/s$ $m = 3$ $N = 12$ $L = 1.2743$ $\gamma_{x,\max} = 5$ $\gamma_{N,\max} = 5$ $\gamma_{L,\max} = 5$					
example1:	$\theta_{x,\max} = 0.01$, $\theta_{N,\max} = 0.01$, $\lambda_{L,\max} = 0.01$, $D = 1\,{}^{rad}/_{s^2}$				
$\underline{e}_e$	*5.04°/s*			κ	*2.2204*10-11*
$\underline{\gamma}_x$	*0.33281*	γ_x	*5*	σ_x	*32.839*
$\underline{\gamma}_N$	*0.42373*	γ_N	*5*	σ_N	*32.840*
$\underline{\gamma}_L$	*0.44442*	γ_L	*5*	σ_L	*32.836*
T_e	*47.5s*	b_e	*13.51°/s*		
example 2:	$\theta_{x,\max} = 0.02$, $\theta_{N,\max} = 0.02$, $\lambda_{L,\max} = 0.02$, $D = 1\,{}^{rad}/_{s^2}$				
$\underline{e}_e$	*5.78°/s*			κ	*2.2204*10-11*
$\underline{\gamma}_x$	*1.1512*	γ_x	*5*	σ_x	*17.214*
$\underline{\gamma}_N$	*1.3380*	γ_N	*5*	σ_N	*17.214*
$\underline{\gamma}_L$	*1.3398*	γ_L	*5*	σ_L	*17.214*
T_e	*46.40s*	b_e	*20.94°/s*		
example 3:	$\theta_{x,\max} = 0.01$, $\theta_{N,\max} = 0.01$, $\lambda_{L,\max} = 0.01$, $D = 2\,{}^{rad}/_{s^2}$				
$\underline{e}_e$	*9.19°/s*			κ	*2.2204*10-11*
$\underline{\gamma}_x$	*0.50308*	γ_x	*5*	σ_x	*32.554*
$\underline{\gamma}_N$	*0.50882*	γ_N	*5*	σ_N	*32.564*
$\underline{\gamma}_L$	*0.71295*	γ_L	*5*	σ_L	*32.557*
T_e	*47.4s*	b_e	*14.49°/s*		
example 4:	$\theta_{x,\max} = 0.02$, $\theta_{N,\max} = 0.02$, $\lambda_{L,\max} = 0.02$, $D = 2\,{}^{rad}/_{s^2}$				
$\underline{e}_e$	*10.09°/s*			κ	*2.2204*10-11*
$\underline{\gamma}_x$	*1.3339*	γ_x	*5*	σ_x	*17.181*
$\underline{\gamma}_N$	*1.5195*	γ_N	*5*	σ_N	*17.181*
$\underline{\gamma}_L$	*2.0343*	γ_L	*5*	σ_L	*17.181*
T_e	*46.1s*	b_e	*21.74°/s*		
example 5:	$\theta_{x,\max} = 0.01$, $\theta_{N,\max} = 0.01$, $\lambda_{L,\max} = 0.01$, $D = 0\,{}^{rad}/_{s^2}$				
$\underline{e}_e$	*0.70°/s*			κ	*16.933*
$\underline{\gamma}_x$	*0.24022*	γ_x	*5*	σ_x	*34.626*
$\underline{\gamma}_N$	*0.24022*	γ_N	*5*	σ_N	*34.623*
$\underline{\gamma}_L$	*0.26471*	γ_L	*5*	σ_L	*34.623*
T_e	*47.1s*	b_e	*12.80°/s*		

Since learning rates should not become too large for the reasons, stated recently, large parameter bounds are somehow critical in the design procedure. Moreover, the ultimate bound b_e is increasing with the parameter bounds, while the σ-modification gains are decreasing. In order to interpret this behavior, recall that σ-modification deteriorates performance since it pulls the parameter estimates back to zero (refer to Appendix D.2 and equation (D.9)) and this effect grows with increasing distance from the origin. Now, larger parameter bounds potentially imply a larger distance of the true parameter from zero and the parameter estimate has to "battle" against a stronger σ-modification effect in order to converge to the true parameter. This deteriorating effect explains, why the σ-modification gains are reduced by the design procedure if the parameter bounds are increased. It is important to notice that we use switching σ-modification, which does not have this negative effect, since it is disabled if the parameter estimates reside within some predefined set, which contains the true parameters. However, in stability analysis, this improvement is not considered, since the upper bounds on the Lyapunov function derivatives (in fact the γ-functions, according to Corollary 3.2) are obtained under the assumption that σ-modification is active (refer to equation (4.156)). However, despite reduction of σ-modification gains, the ultimate bound is increased. Notice that learning rates in optimization 2 are bounded from above by a fixed value for all examples. Further computations showed that the ultimate bound is not considerably increased, if the upper bounds are defined as a multiple of the minimum learning rates.

Finally it is observed that MMQ modification is counterproductive in presence of unmatched uncertainty which results in $\kappa=0$ in optimization 2. This also complies with remark 4 of section 4.4.3. In absence of unmatched uncertainties, however, the gain design procedure results in a finite $\kappa>0$ (example 5).

Gain Design Philosophy – Accounting for Certification Aspects

In classical linear flight control systems, robustness of the closed loop system is analyzed by means of gain and phase margin while the specific values have evolved historical on an empirical basis, but not on resilient analytical considerations, however, these are accepted as robustness evidence. It is a today's challenge to provide the certification authorities with robustness metrics for nonlinear adaptive control systems that provide the same level of confidence as gain and phase margin do for linear systems. Although it is in no way claimed to be the ultimate solution, the following proposition might be useful towards that end.

It is a known drawback, that Lyapunov based stability analysis for MRAC systems result in rather conservative stability bounds and also the proposed gain design procedure results in unacceptably high lower bounds on the tracking error $\underline{e}_e$ (several hundred °/s) and unacceptably high minimum learning rates if the bounds on the true

parameters incorporate failure cases. Classical linear flight control systems are commonly designed for the nominal aircraft configuration and are disabled as soon as failure cases occur, nevertheless it is often still possible to operate and land the aircraft safely.

This consideration is incorporated in the adaptive control design by reducing the parameter bounds such that they can cancel model uncertainties in the nominal case but cannot recover performance in case of failures. This restriction results in acceptable values for $\underline{e}_e$ and $\underline{\gamma}_x$, $\underline{\gamma}_N$, $\underline{\gamma}_L$. This gain design procedure guarantees – by Lyapunov's theorems – that the aircraft is operated stably in nominal configuration and in presence of a certain amount of unmatched uncertainties. In any emergency case – when linear control systems are disengaged anyway – the parameter bounds are enlarged by the operator or pilot such that the adaptation is able to recover a safe aircraft operation. I.e. in the emergency case, the parameter bounds are set to

$$\theta_{x,\max,\,E} = d_x \cdot \theta_{x,\max} \qquad \theta_{N,\max,\,E} = d_N \cdot \theta_{N,\max} \qquad \lambda_{L,\max,\,E} = d_L \cdot \lambda_{L,\max} \tag{5.60}$$

for some d_x, d_x, d_x>1.

For argumentation towards certification authorities it can be stated that the flight control system is guaranteed to operate the aircraft safely in the nominal case but, additionally, provides the chance to recover performance in emergency cases, which is not provided by classical control systems. Thus, beside equivalent reliability in the nominal case, adaptive systems provide additional safety in failure cases.

The author is aware of the fact that consideration of unmatched uncertainty does not incorporate unmodeled dynamics. In recent years, research has intensively focused on robustness considerations of adaptive systems against unmodeled dynamics; a promising approach so far is the consideration of time delay margin. This, however, requires an in-depth discussion on the theoretical fundamentals and would go beyond the scope of the thesis. At least the proposed gain design procedure provides an explicit and resilient robustness margins against unmatched uncertainties.

Estimates for Bounds on Parameters and Unmatched Uncertainty

As explained recently, in order get acceptable $\underline{e}_e$ and $\underline{\gamma}_x, \underline{\gamma}_N, \underline{\gamma}_L$, the parameter bounds have to be reduced such that adaptation capabilities are sufficient for compensation of model uncertainties. If this still does not produce satisfactory results, the parameter bounds can even be reduced further such that model uncertainties can only partially be compensated, of course, for the price of an increased unmatched part.

A remaining challenge is the determination of an appropriate value for the unmatched uncertainty bound. This could be accomplished numerically by variation of uncertain parameters and computation of the maximum unmatched uncertainty that occurs

within the flight envelope, but for demonstration purposes, we will utilize a physical approach in order to get imagination for the magnitude of the uncertainty bound. By equation (5.22), the unmatched uncertainty is an angular acceleration caused by unmatched moments.

Referring to Figure 5.20, Figure 5.21 the main engine maximum thrust is about $X_{Pr,max}=10N$ while the torque is about $L_{Pr,max}=0.2Nm$. The nominal position vectors from aircraft c.g. to the left and right main propulsion are

$$\left(\vec{\mathbf{r}}^{G\,\mathrm{Pr}_l}\right)_B = \begin{pmatrix} -42.4 \\ -197.7 \\ -41.3 \end{pmatrix} mm \quad , \quad \left(\vec{\mathbf{r}}^{G\,\mathrm{Pr}_r}\right)_B = \begin{pmatrix} -42.4 \\ 200.3 \\ -41.3 \end{pmatrix} mm \,,$$

and a worst case estimate for the total propulsive moments M_{tot} (i.e. the Euclidean norm of the moment vector) relative to the c.g. (if the back engine is assumed not to contribute to the moments considerably) is given by the distance of the respective engine to the c.g., multiplied with the maximum thrust.

$$M_{tot} = L_{\mathrm{Pr,max}} + \left(\left\| \left(\vec{\mathbf{r}}^{G\,\mathrm{Pr}_l}\right)_B \right\|_2 + \left\| \left(\vec{\mathbf{r}}^{G\,\mathrm{Pr}_r}\right)_B \right\|_2 \right) X_{\mathrm{Pr,max}} = 5.35Nm \,.$$

Now, 10% of the maximal moment should be a very conservative approximation for the unmatched moment. Moreover the inertia tensor relative to c.g. is

$$\left(\mathbf{I}^G\right)_B = \begin{bmatrix} 129 & 0 & 17 \\ 0 & 276 & 0 \\ 17 & 0 & 392 \end{bmatrix} \cdot 10^{-3} kgm^2$$

and, since inertia about the roll axis is the smallest one, the worst case is a unmatched moment about the roll axis and hence

$$D = \frac{0.1 \cdot M_{tot}}{I_x} = \frac{0.1 \cdot 5.35Nm}{129 \cdot 10^{-3} kgm^2} = 237.622 \frac{\mathrm{deg}}{s^2} \,. \qquad (5.61)$$

At this point, it has to be noticed that this procedure does not provide less confidence than gain and phase margins provide for linear system, since values for these robustness margins are based on experience but not on hard mathematical facts. The main difference is the lack of empirical values for D, but they could be gained by numerical variation of parameters in the mathematical model such as inertia sensor, mass or c.g., however those investigations are left for the future and we will rely on the conservative bound in equation (5.61).

Valid Domain for External States

The bound on the external states $\bar{\zeta}$ is chosen, based on the gyroscope sensors, which have a range from $-200^{\mathrm{deg}}/_{s}$ to $200^{\mathrm{deg}}/_{s}$ in each axis. Hence, the admissible set of the

external states is a cube of edge length $400^{\deg}/_{s}$ around the origin. The set of external states that is defined by equation (4.137), is a ball of radius $\bar{\zeta}$. Hence, if we choose

$$\bar{\zeta} = 200\ ^{\deg}/_{s} \tag{5.62}$$

the ball is completely covered by the cube.

5.3.6 Controller Parameter Assignment

Baseline Controller

The following parameters have been chosen for the reference model time constants and the error feedback gains.

Table 5.4 Parameters Baseline Controller

axis	reference model time constant		error controller gain	
roll axis	T_p	*0.15s*	k_p	*80/s*
pitch axis	T_q	*0.12s*	k_q	*40/s*
yaw axis	T_r	*0.2s*	k_r	*40/s*

In order to find r_e that satisfies the right hand side of (5.59) the reference model commands and states have to be limited.

Table 5.5 Reference Model Limitation

axis	reference model limitation	
roll axis	p_{max}	*130°/s*
pitch axis	q_{max}	*80°/s*
yaw axis	r_{max}	*20°/s*

Hence, we obtain for the reference model bound

$$\bar{\zeta}_R = \sqrt{130^2 + 80^2 + 20^2}\ ^{\deg}/_{s} = 153.95^{\deg}/_{s}\ . \tag{5.63}$$

Adaptive Element

In view of equations (5.62) and (5.63), the tracking error shall be not be bigger than

$$r_e = 45^{\deg}/_{s}\ . \tag{5.64}$$

This choice assures that the plant state will not exceed the limit of $200^{\deg}/_{s}$ given by the measurement range of the sensors. Table 5.6 lists the parameters that have been chosen for the adaptation algorithm. The first part contains values that are defined a priori while the second part contains results of the gain design procedure, described in section 5.3.5.

In order to achieve the bound on the tracking error of equation (5.59), the parameter bounds have been reduced to an extent such that they only partially cancel the model uncertainties for the undamaged nominal aircraft. As will be shown in simulation results, the parameter estimates operate at their respective boundaries even in the nominal case, nevertheless tracking performance is acceptable.

Table 5.6 Parameters of Adaptive Element

fixed parameters:					
Lyapunov equation	$\mathbf{P}_E = \begin{bmatrix} 6.25 & 0 & 0 \\ 0 & 12.5 & 0 \\ 0 & 0 & 12.5 \end{bmatrix} \cdot 10^{-3}$ $\mathbf{Q}_E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$				
weighting matrices	$\mathbf{\Gamma}_x = \mathrm{diag}[0.7 \quad 0.7 \quad 0.7 \quad 1 \quad 0.7 \quad 1 \quad 0.2 \quad 0.7 \quad 1 \quad 1 \quad 1]$ $\mathbf{\Gamma}_N = 15 \cdot \mathbf{I}^{(126\times126)}$ $\mathbf{\Gamma}_L = \mathrm{diag}[0.5 \quad 0.5 \quad 0.5 \quad 0.5 \quad 0.5 \quad 0.5 \quad 0.5 \quad 0.5 \quad 1]$				
fade-in factors (projection and switching σ-mod.)	$\varepsilon_x = \varepsilon_N = \varepsilon_L = 0.1$				
parameter bounds	$\theta_{x,\max} = 0.04$, $\theta_{N,\max} = 0.12$, $\lambda_{L,\max} = 0.03$				
emergency parameter bound enlargement	$d_x = 20$, $d_N = 60$, $d_L = 30$				
unmatched uncert. bound	$D = 237.622 \ {}^{\mathrm{deg}}/_{s^2}$				
MMQ modification	low-pass bandwidth:	$\Omega_0 = 2\pi \ {}^{rad}/_{s}$			
	overall bandwidth:	$\overline{\Omega} = 8 \cdot (2\pi) \ {}^{rad}/_{s}$			
	Number of filters	$N = 12$			
	damping ratio:	$\zeta = 9.4661 \cdot 10^{-2}$			
	$\mathcal{L}_1$-norm bound:	1.2743			
gain design					
$\overline{\lambda}_{P_E} = 1.25 \cdot 10^{-2}$ $\underline{\lambda}_{P_E} = 6.25 \cdot 10^{-3}$ $\underline{\lambda}_{Q_E} = 1$ $m = 3$ $r_e = 45°/s$					
$\underline{e}_e$	*23.60°/s*			κ	*2.2204*10^{-11}*
$\underline{\gamma}_x$	*4.2344*	γ_x	*4.6579*	σ_x	*10.616*
$\underline{\gamma}_N$	*26.066*	γ_N	*28.673*	σ_N	*1.5154*
$\underline{\gamma}_L$	*4.7615*	γ_L	*5.2376*	σ_L	*12.963*
T_e	*44.63s*	b_e	*39.32°/s*		

Nonlinear-in-Control Design

Since a minimum control authority of the nonaffine controls cannot be guaranteed, as stated in the introductory remark at the beginning of the chapter, convergence of the NIC algorithm to the desired value is not assured. However, since thrust vectoring is not used for control within the linearizing state feedback, but only for trimming, while the remaining effect is cancelled by the linear controls, it is not necessary for the thrust vector controls to converge. Nevertheless, it has to be assured that the thrust vector angles do not exceed their maximum deflection, however, this can be easily assured by limitation of the integrators of the gradient minimization algorithm.

The nonlinear-in-control algorithm is implemented as normalized gradient system according to equation (4.261). Thereby singular values of the control map Jacobian

could get zero due to insufficient control authority, which requires for a limitation of the singular values from below by some small positive value, since otherwise differential equation (4.261) becomes singular. Notice that the time scale separating parameters have been chosen quite large, since they are only used for trimming, which does not require for high dynamics. Moreover, slower commands preserve the thrust vector actuators.

Table 5.7 Nonlinear-in-Control Parameters

main engine	
time scale separation	$\varepsilon = 2.5$
minimum SV of control map Jacobian	$\underline{\sigma} = 0.01$
back engine	
time scale separation	$\varepsilon = 2.5$
minimum SV of control map Jacobian	$\underline{\sigma} = 0.01$

5.3.7 Simulation Results

The section presents simulation results that show the performance of the control system, including the dynamic inversion baseline controller, the adaptive part as well as the nonlinear-in-control design for the thrust vector angles.

The simulation will show the nominal performance of the baseline controller and improved performance, if adaptation is activated. Additionally, capabilities of the adaptive system to recover the aircraft in case of actuator failures will be presented. Compliant with the gain design philosophy, as introduced in section 5.3.4, bounds on the parameter estimates are restricted such that the adaptation can only compensate for model uncertainties of the undamaged aircraft. In emergency cases, the parameter bounds are enlarged according to equation (5.60), which is denoted as “emergency button” subsequently.

Damage Scenarios

Besides, for nominal performance, the flight control system will also be tested for its capabilities to handle severe aircraft damages. Therefore, three damage scenarios are defined. In all scenarios, the simulation begins with an undamaged aircraft and, in each case, after $5s$, different severe failures occur, depending on the scenario according to Table 5.8.

Table 5.8 Damage Scenarios

scenario 1	scenario 2	scenario 3
left canard missing	left canard missing	50% of left wing missing
left aileron missing	right canard missing	30% of right wing missing
right flap missing	left aileron missing	
right elevator missing	right aileron missing	

Simulation Setup

The simulation is set up with the high fidelity simulation model, comprising rigid body and actuator dynamics as described in Chapter 2. Additionally, sensor models are implemented which comprise measurement insufficiencies such as inaccurate mounting, bias and noise. The simulation is run at a fixed sample rate of $500Hz$ and an additional time delay of one sample is introduced to the sensor channel to account for digital delays. In order to compare the capabilities of baseline controller and adaptive system for the undamaged and damaged case, the following simulation scenarios are defined.

Table 5.9 Simulation Scenarios

no.	description	damage scenario	adaptation	emergency
1	nominal performance, baseline controller	no	no	no
2	nominal performance, with adaptation	no	yes	no
3	nominal performance, emergency button	no	yes	yes
4	damage scenario 1, with adaptation	1	yes	no
5	damage scenario 1, emergency button	1	yes	yes
6	damage scenario 2, emergency button	2	yes	yes
7	damage scenario 3, emergency button	3	yes	yes

Nominal Performance

Simulation scenario 1 shows performance of the baseline controller (without adaptation) for the nominal aircraft configuration. Figure 5.34 presents the tracking error of roll, pitch and yaw axis as well as Euclidean norm of the tracking error. Obviously, the model, assumed for linearizing state feedback does not match reality perfectly, which results in a finite deviation between reference model and plant of about $10deg/s$ mean Euclidean norm.

Figure 5.35 shows the internal states, namely velocity, angle of attack (AoA), angle of side-slip (AoS), bank angle and pitch angle. It reveals that, particularly AoA and AoS remain within an acceptable range. (notice that boundedness of the internal states has

not been shown but was assumed a priori). Also, the aircraft attitude shows an acceptable behavior, although it is not actively controlled. However, it has to be admitted that the attitude behavior has been achieved by an offset to the external command hat compensates for the steady state tacking error.

Figure 5.36 shows the control surface commands, where the dashed lines indicate the maximum deflection of the respective controls. It reveals that there is still sufficient control authority left. Figure 5.37 presents the thrust vector angle commands. The main engine thrust vector angles are deflected antisymmetrically and compensate for the roll moment, produced by the propeller torque. The range of the back engine thrust vector angles is obviously not sufficient to compensate for pitch and yaw moment, which results in saturated commands. On the one hand, the back engine has negligible power compared to the main engines, which effects that the back engine thrust is very low at the current velocity or even changes sign. The sign changes also explain why the back engine thrust vector angle commands run from one saturation to the other. On the other hand, the pitch moment that is produced by the main engines has to be compensated by the back engine inclination angle as explained section 5.3.4. However, due to lack of power the back engine compensation is insufficient. The behavior is certainly not satisfying, however, this is not a problem of control but of the aircraft design.

Figure 5.38 and Figure 5.39 show the nonlinear-in-control error as well as the singular values of the nonlinear control map Jacobian (refer to Figure 5.31). Thereby the main engine has one channel since it only compensates for the roll axis using the virtual thrust vector angle σ_m according to equation (5.41), while the back engine has two channels, compensating for pitch and yaw axes, using back engine thrust vector inclination and azimuth angle. In Figure 5.38 it is observed that the main engine nonlinear-in-control error (about the roll axis) is compensated while Figure 5.39 discovers that pitch and yaw axes cannot be compensated by the back engine. This complies with saturated thrust vector angles in Figure 5.37.

Figure 5.42 shows the tracking error of simulation scenario 2. It reveals that the tracking error is decreased compared to scenario 1. But there is still a smaller deviation left which occurs from the parameter bounds that are restricted to values such that the adaptive part can only partially compensate for model uncertainties, which is observed in Figure 5.41. It displays the weighted Frobenius norms of the adaptive parameters and, obviously, the parameters quickly run into saturation, which is defined by the projection operator. Accordingly Figure 5.43 – Figure 5.45 show the scalar entries of the adaptive parameter matrices. Particularly in Figure 5.44 and Figure 5.45, it is well observed that $\mathbf{\Lambda}_L$ and $\mathbf{Q}_N$ are saturated.

The parameter bounds are enlarged in simulation scenario 3 (“emergency button”) which effects that the tracking error is further reduced as can be seen in Figure 5.46.

Figure 5.47 and Figure 5.48 – Figure 5.50 show the weighted Frobenius norms and the adaptive parameters itself which are, now, not restricted by limitation and evolve freely.

Simulation Scenario1: Nominal Performance, Baseline Controller

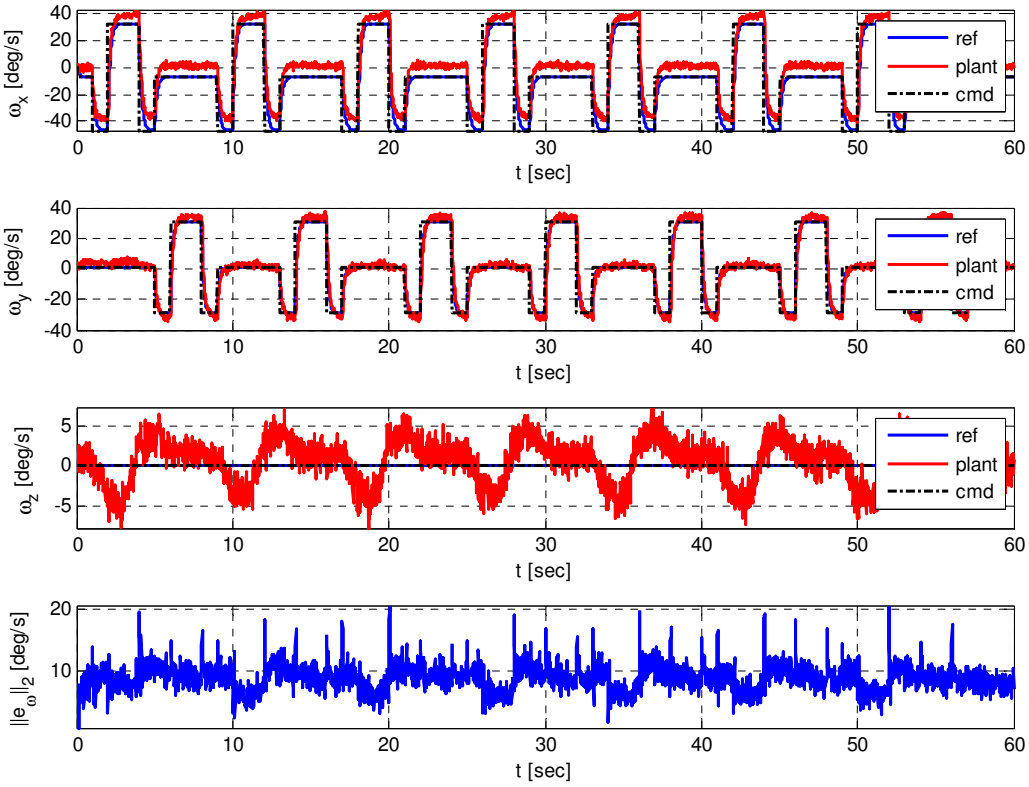


Figure 5.34 Simulation 1 – Tracking Error

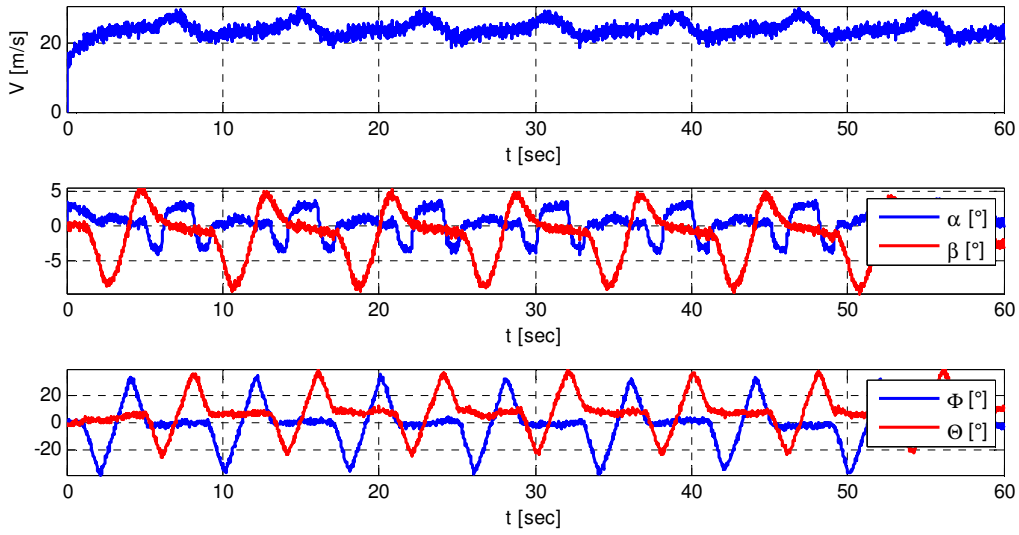


Figure 5.35 Simulation 1 – Internal States

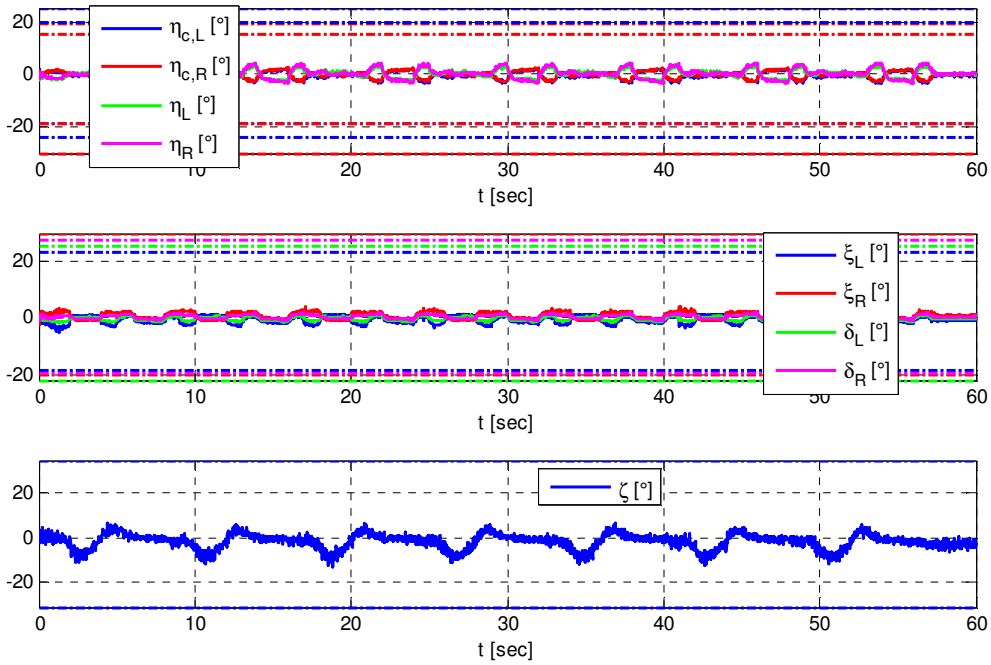


Figure 5.36 Simulation 1 – Control Surface Commands

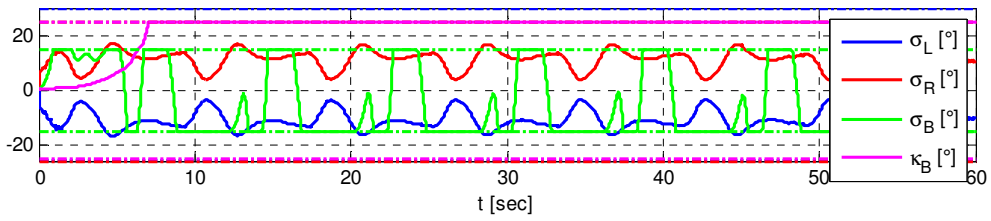


Figure 5.37 Simulation 1 – Thrust Vector Commands

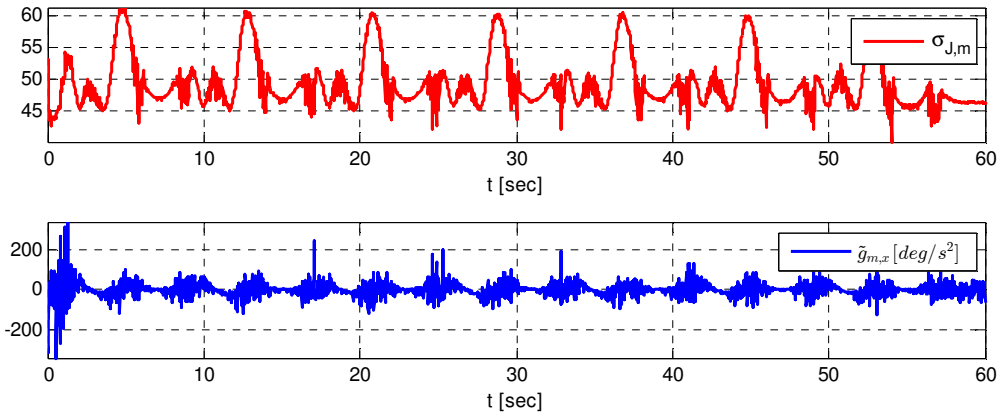


Figure 5.38 Simulation 1 – Nonlinear-in-Control Error, Main Engine

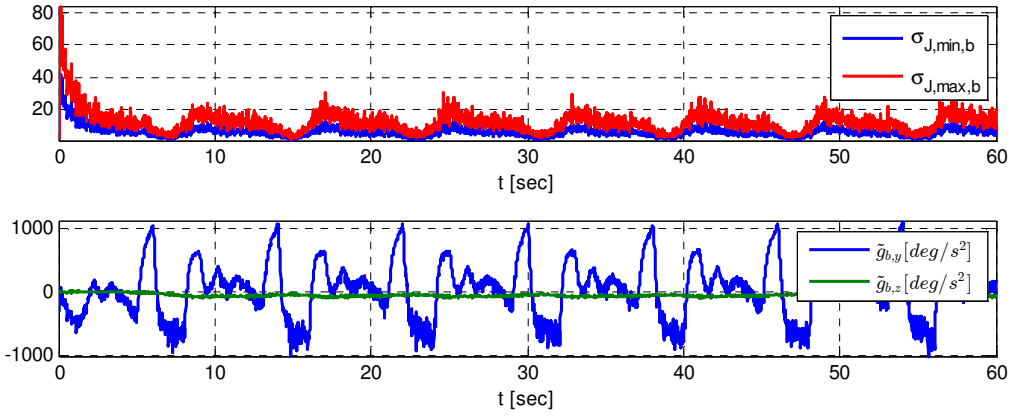


Figure 5.39 Simulation 1 – Nonlinear-in-Control Error, Back Engine

Simulation Scenario 2: Nominal Performance, with Adaptation

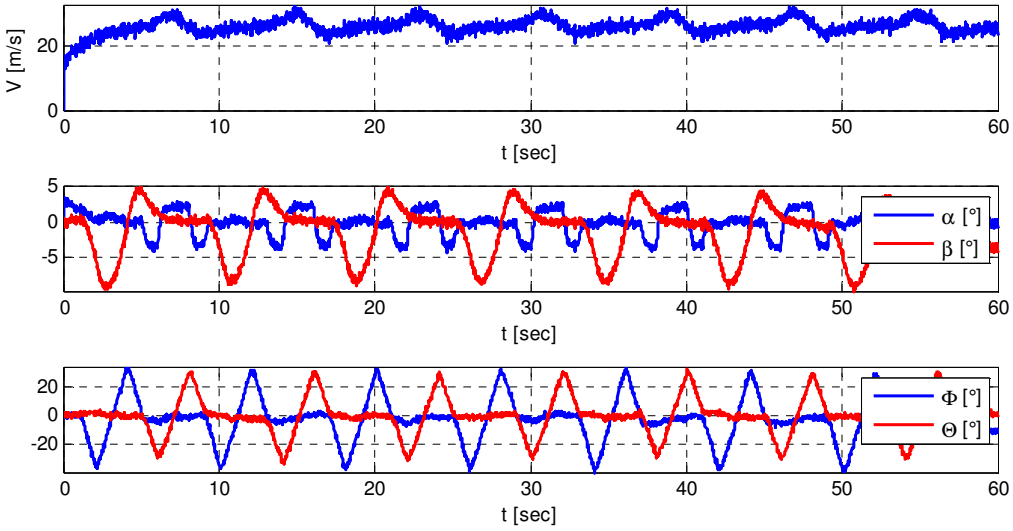


Figure 5.40 Simulation 2 – Internal States

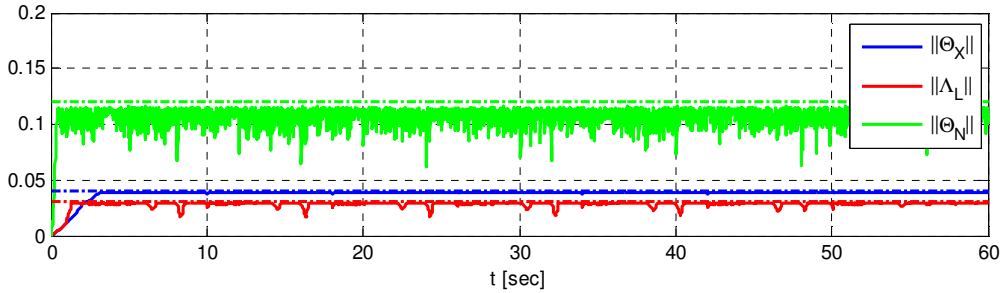


Figure 5.41 Simulation 2 – Weighted Frobenius Norms of Adaptive Parameters

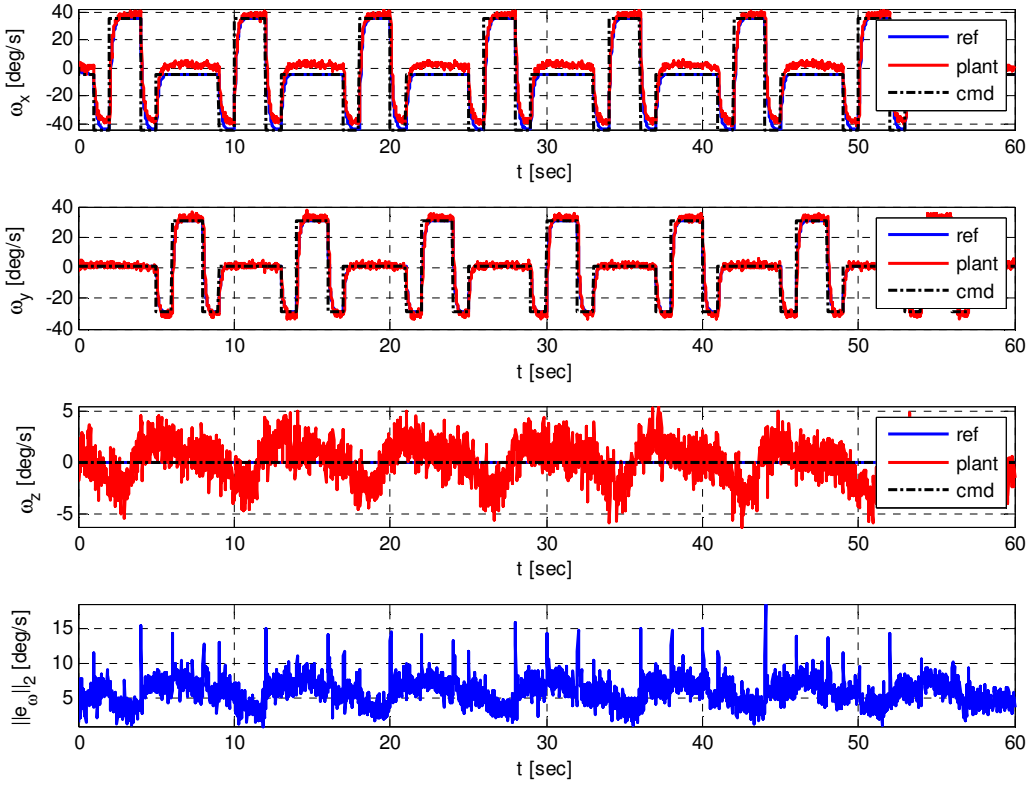


Figure 5.42 Simulation 2 – Tacking Error

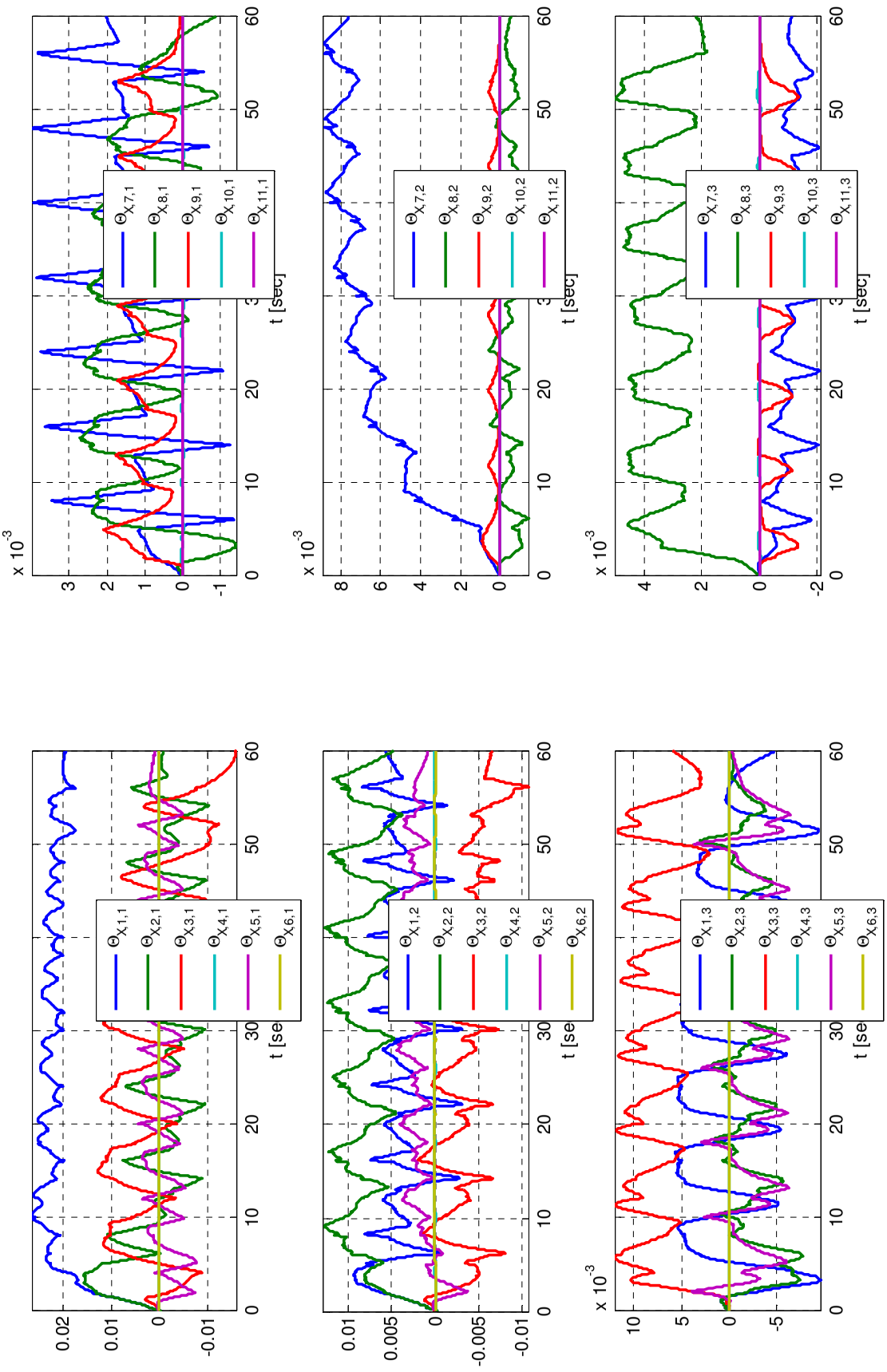


Figure 5.43 Simulation 2 – Theta_X

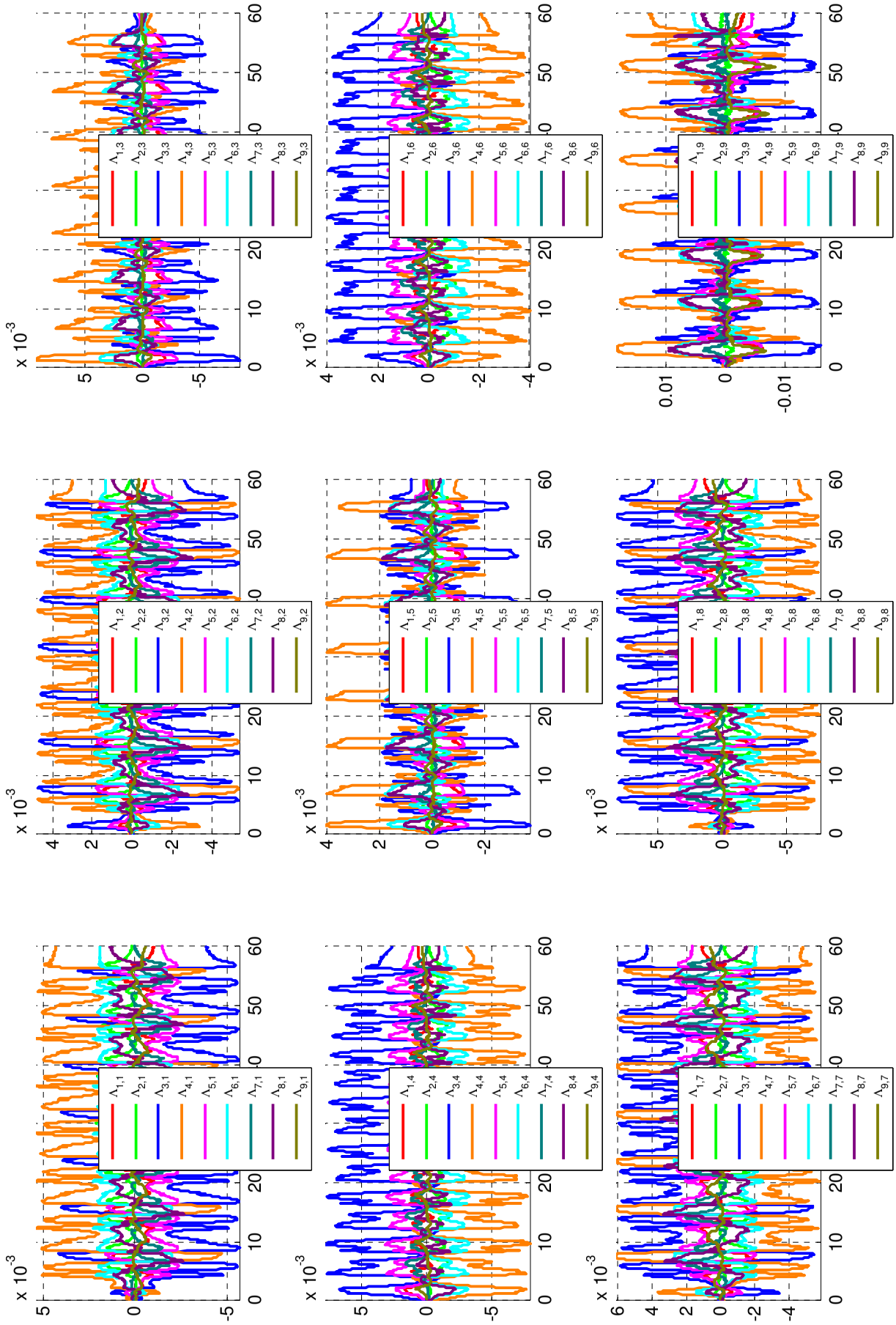


Figure 5.44 Simulation 2 – Lambda_L

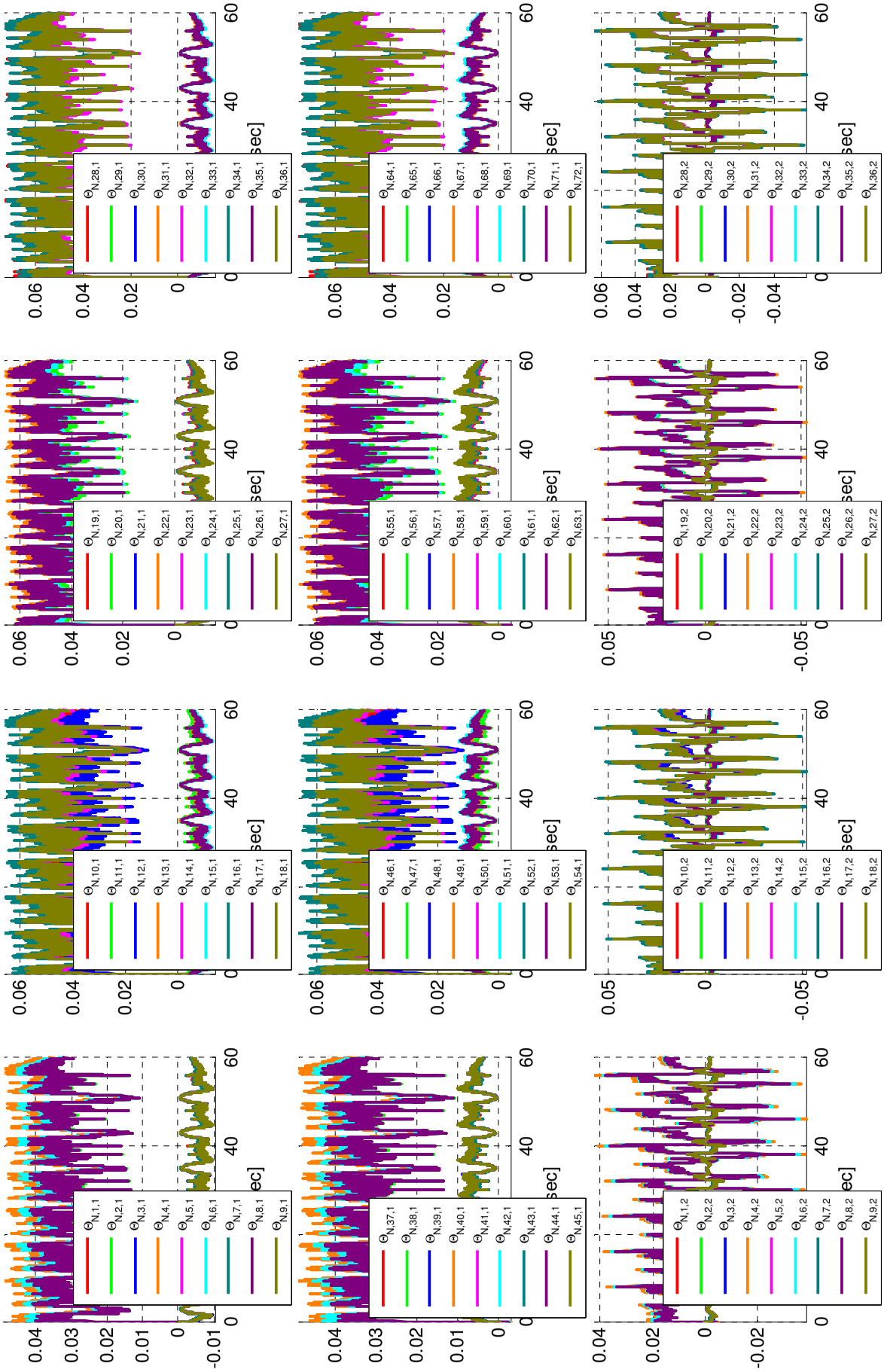


Figure 5.45 Simulation 2 – Theta_N (Main Engine, Partially)

Simulation Scenario 3: Nominal Performance, Emergency Button

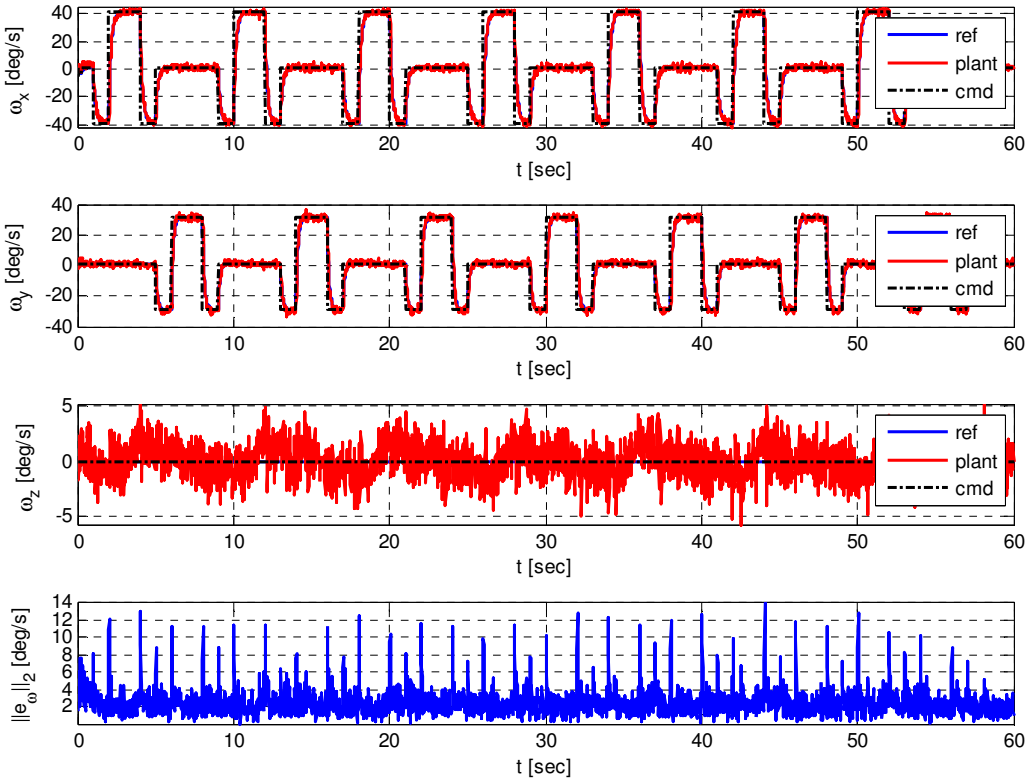


Figure 5.46 Simulation 3 – Tracking Error

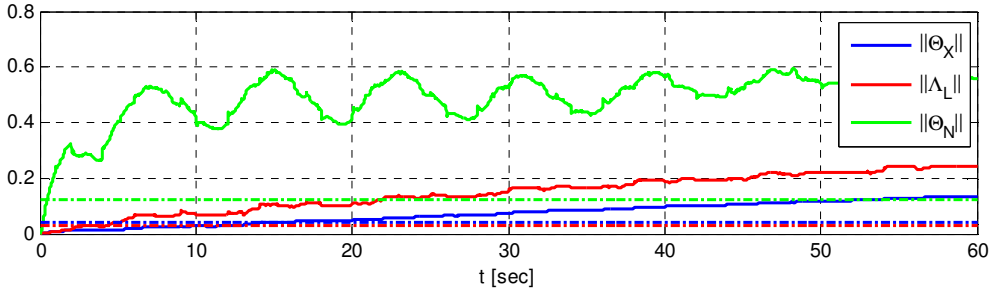


Figure 5.47 Simulation 3 – Weighted Frobenius Norm of Adaptive Parameter

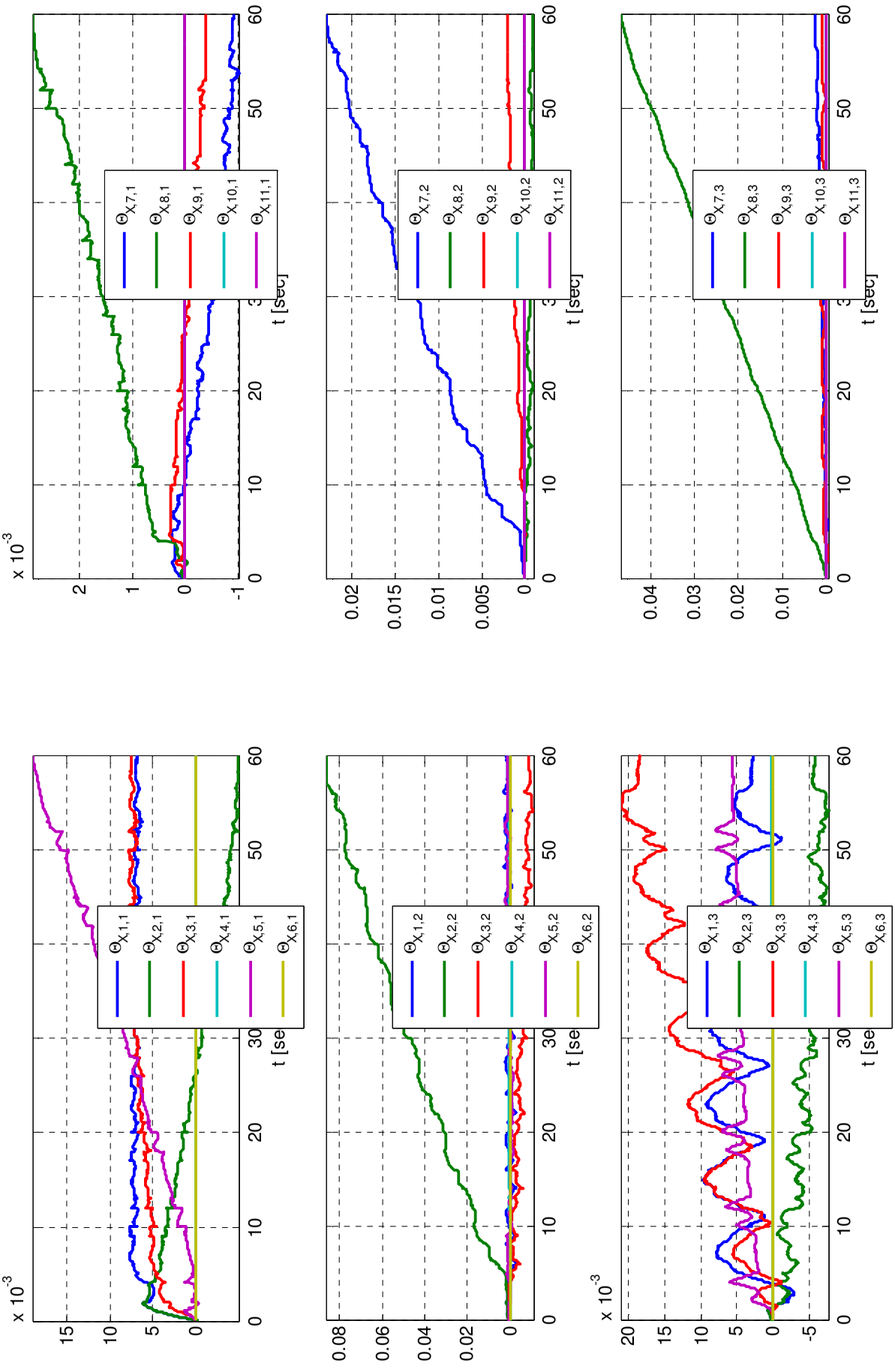


Figure 5.48 Simulation 3 – Theta_X

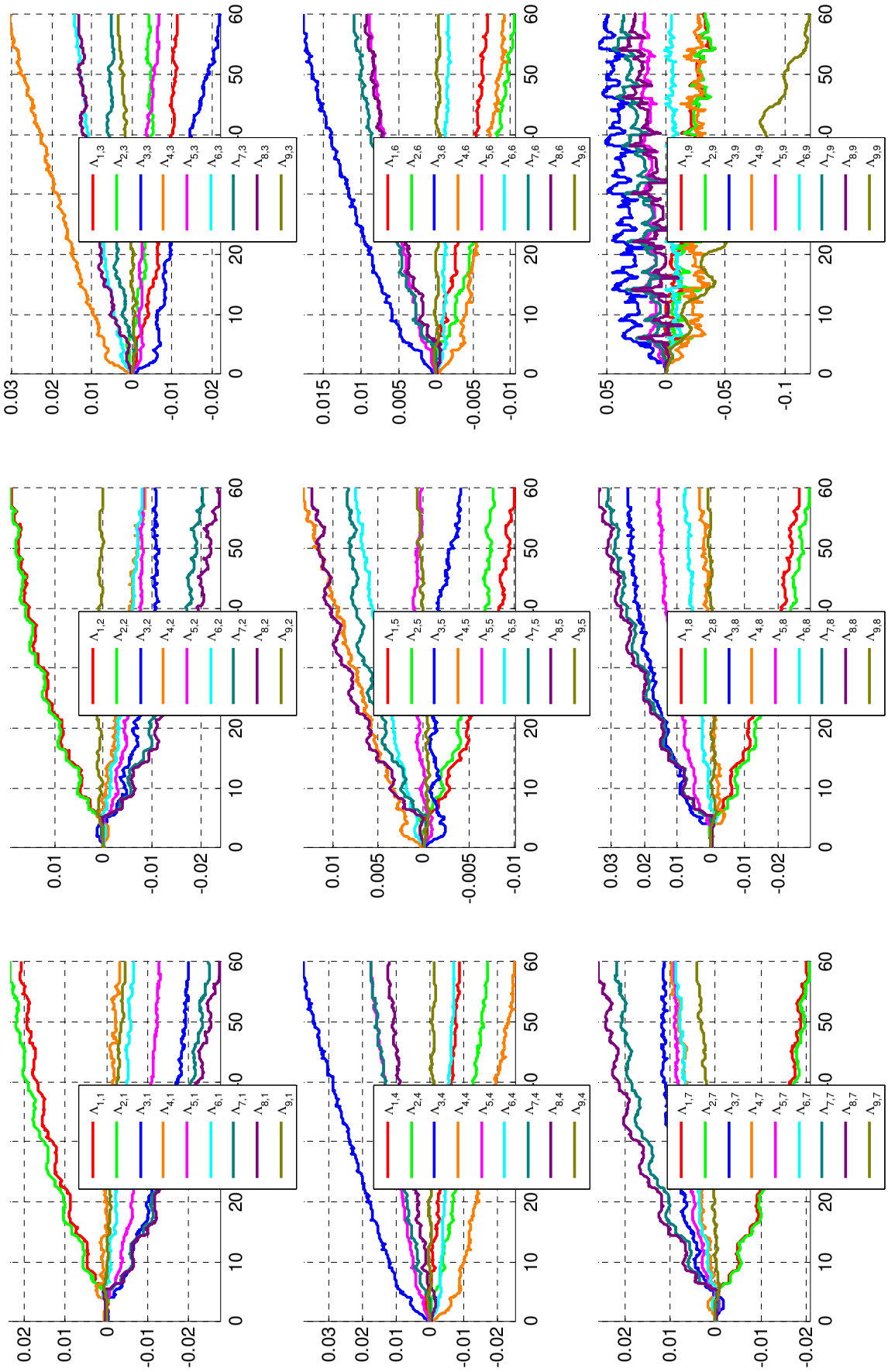


Figure 5.49 Simulation 3 – Lambda_L

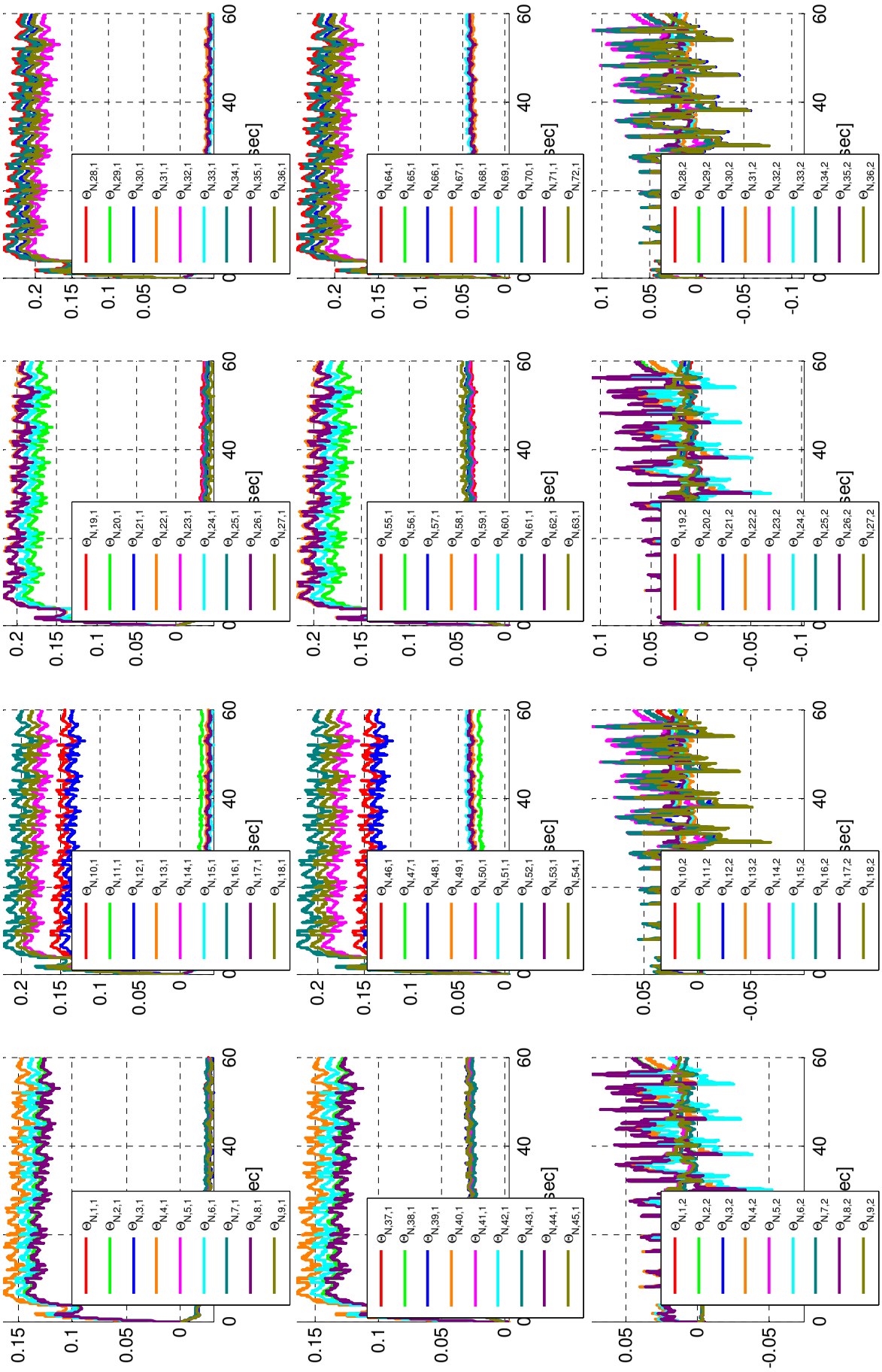


Figure 5.50 Simulation 3 – Theta_N (Main Engine, Partially)

Performance in Case of Severe Damage

Simulation scenario 4 considers damage scenario 1 with adaptation activated, but with disabled emergency button. Figure 5.54, shows that the adaptive parameters immediately run into saturation and cannot compensate for further changes in the aircraft configuration. As seen in Figure 5.51, tracking performance degrades immediately when the damage occurs at $5s$. This is particularly visible in Figure 5.52 where the aircraft attitude is running away very fast, leading to inverted flight. Figure 5.53 shows that, despite 4 missing control surfaces, the control surfaces still have substantial distance to actuator saturation.

If the emergency button is activated, the adaptation recovers nominal tracking performance as can be observed in Figure 5.55, Figure 5.60 and Figure 5.63. Likewise the aircraft attitude does not diverge, although not controlled, according to Figure 5.56, Figure 5.59, Figure 5.64, which highlights the excellent tracking performance even in case of severe damage. In all three damage scenarios, the control surface commands do not run into saturation as depicted in Figure 5.57, Figure 5.61, and Figure 5.65.

Simulation Scenario 4: Damage Scenario 1, with Adaptation

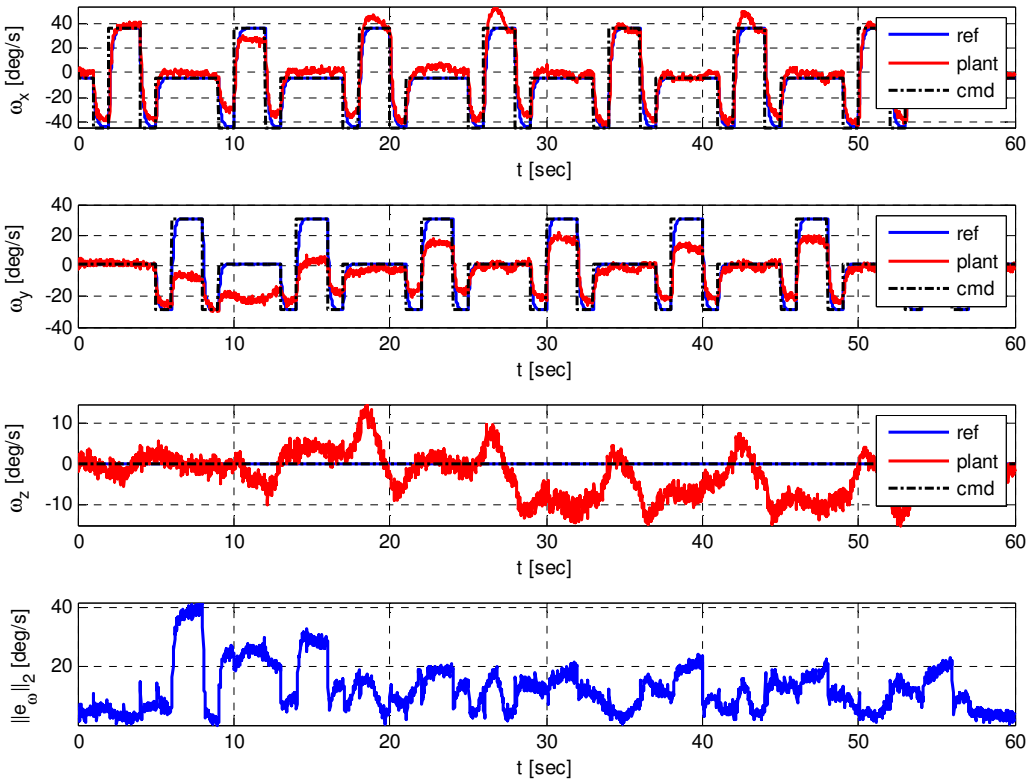


Figure 5.51 Simulation 4 – Tracking Error

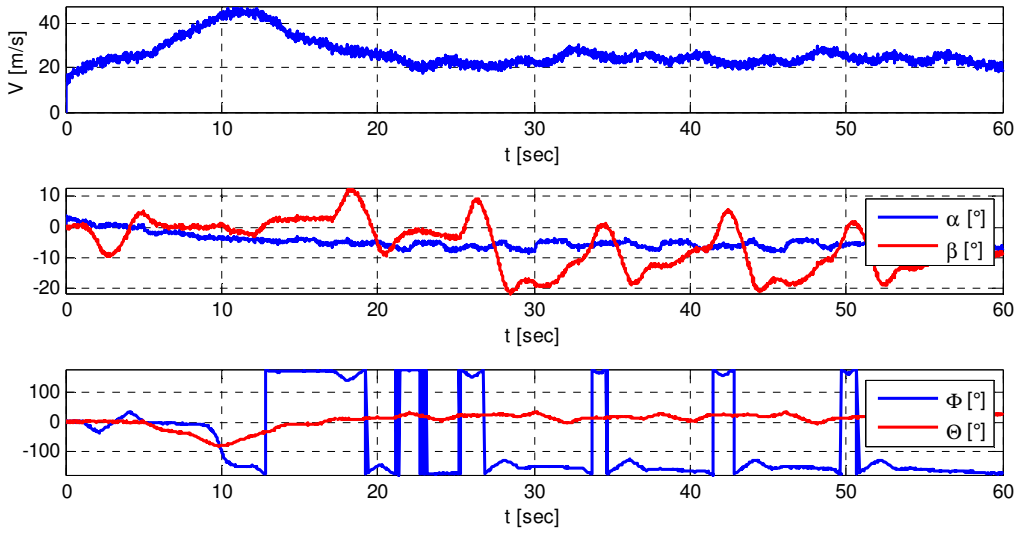


Figure 5.52 Simulation 4 – Internal States

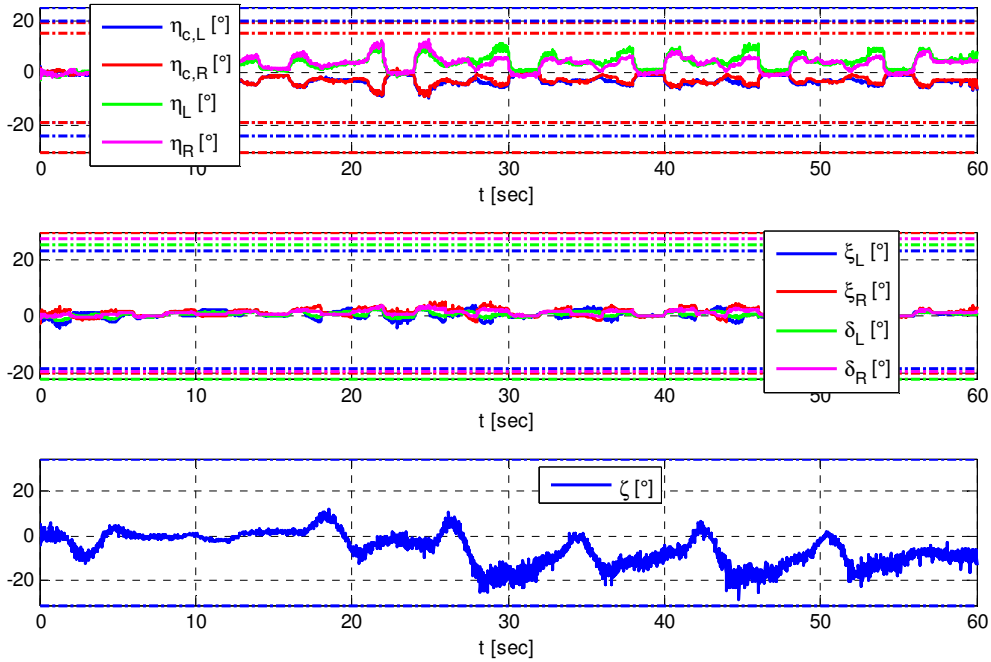


Figure 5.53 Simulation 4 – Control Surface Command

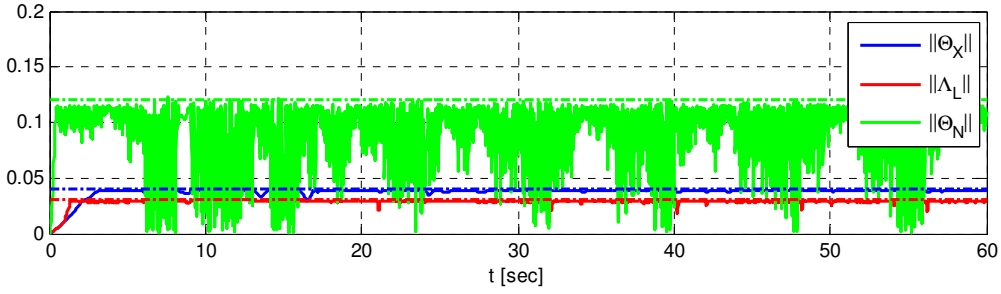


Figure 5.54 Simulation 4 – Weighted Frobenius Norms of Adaptive Parameters

Simulation Scenario 5: Damage Scenario 1, Emergency Button

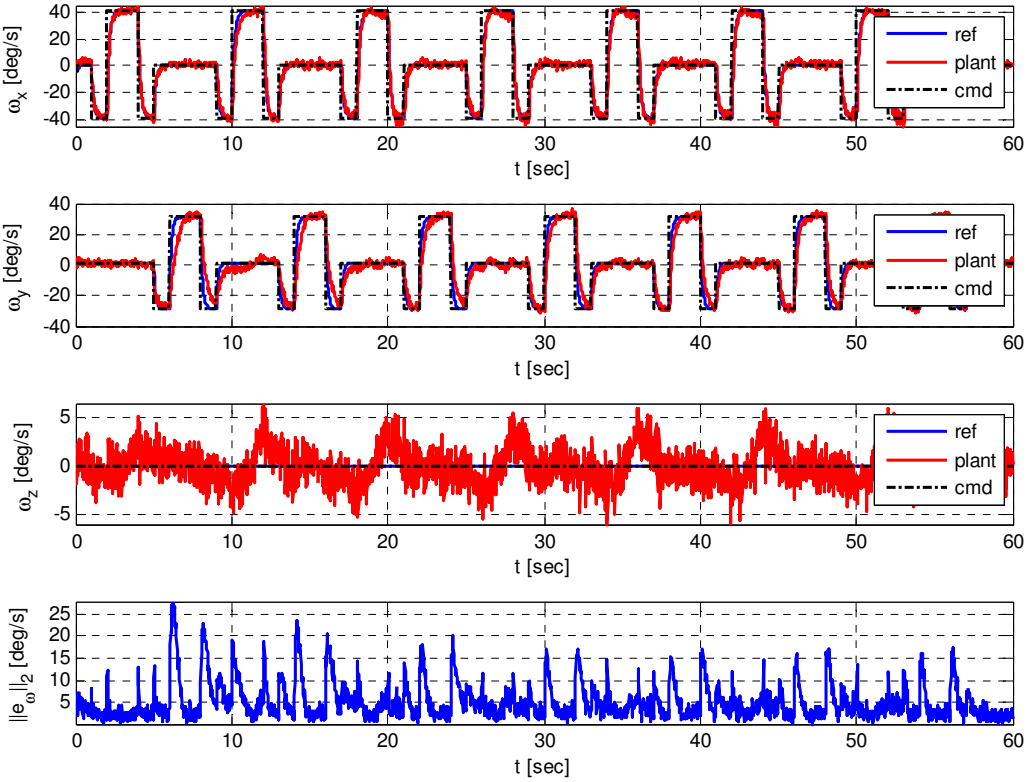


Figure 5.55 Simulation 5 – Tracking Error

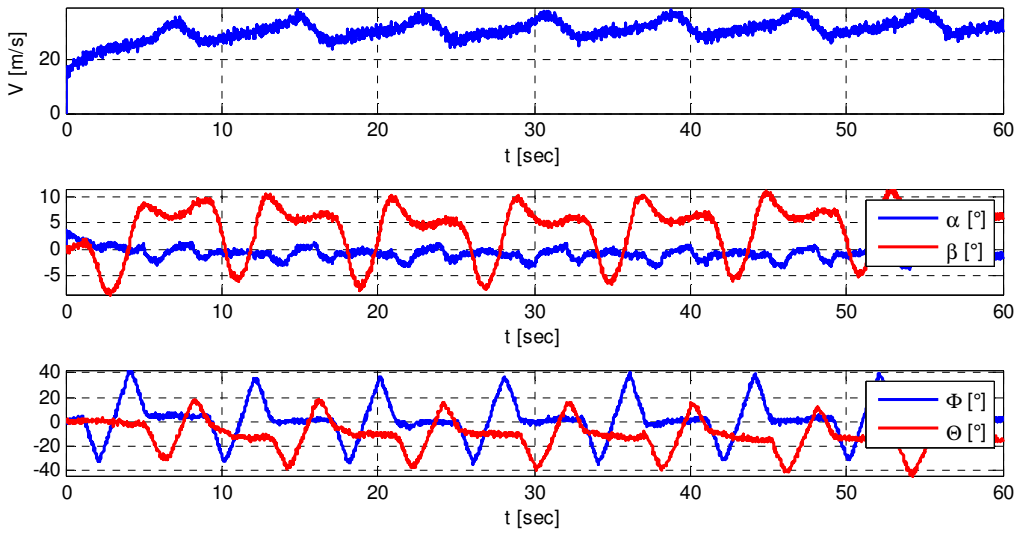


Figure 5.56 Simulation 5 – Internal States

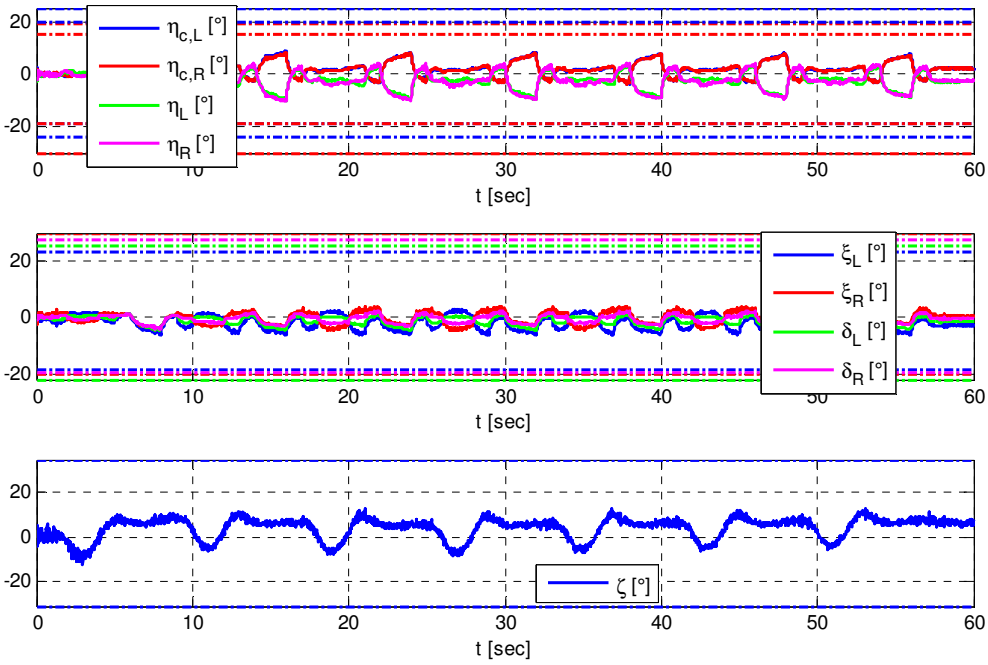


Figure 5.57 Simulation 5 – Control Surface Commands

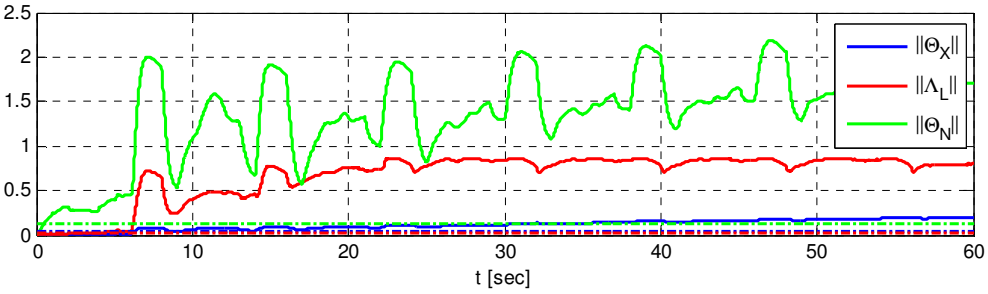


Figure 5.58 Simulation 5 – Weighted Frobenius Norms of Adaptive Parameters

Simulation Scenario 6: Damage Scenario 2, Emergency Button

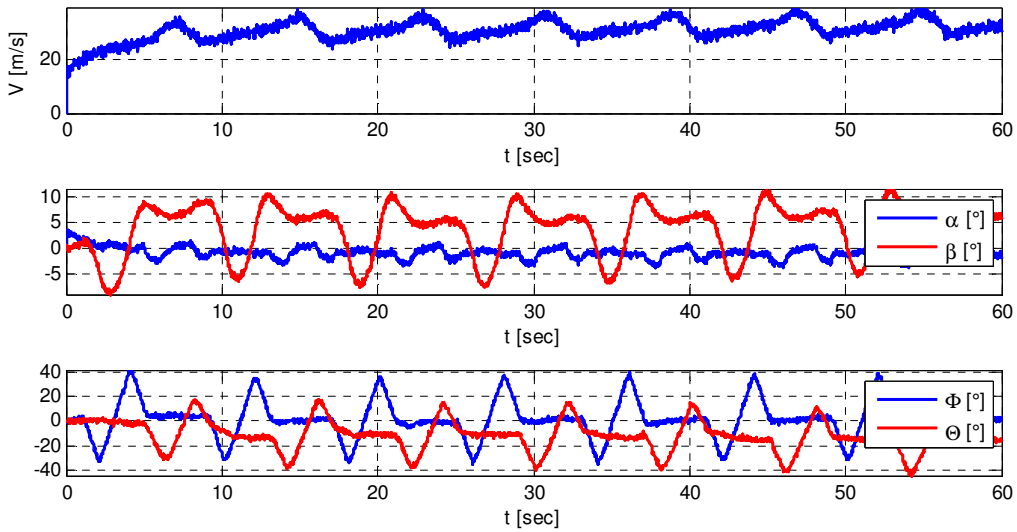


Figure 5.59 Simulation 6 – Internal States

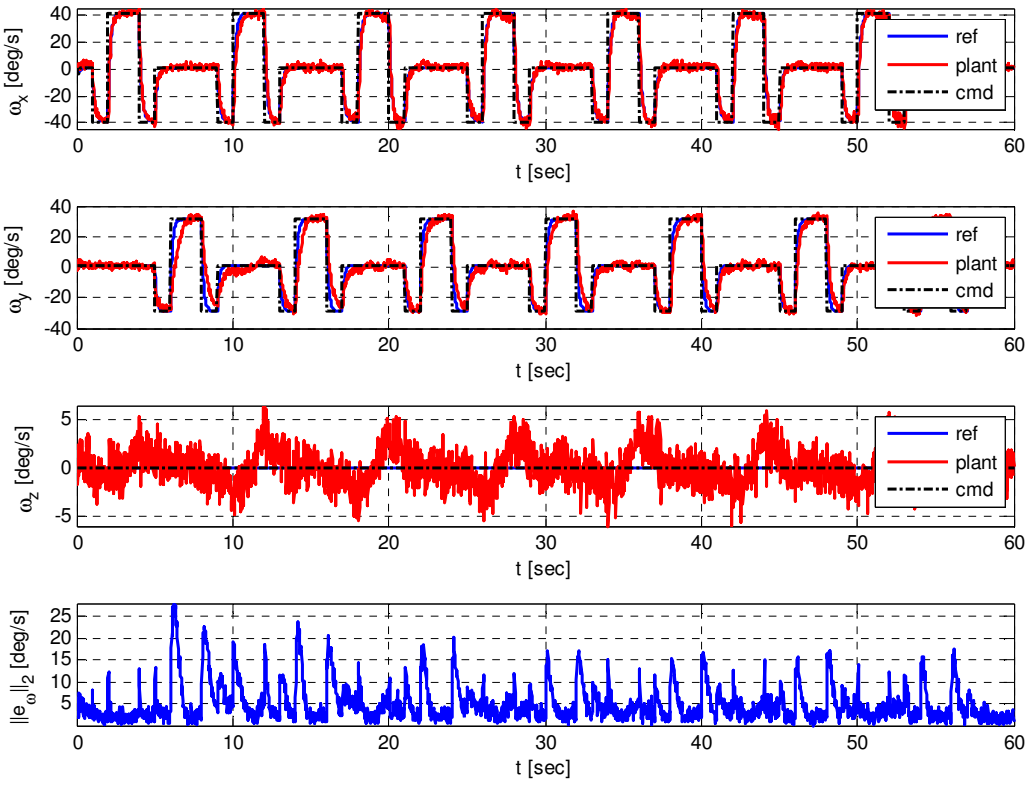


Figure 5.60 Simulation 6 – Tracking Error

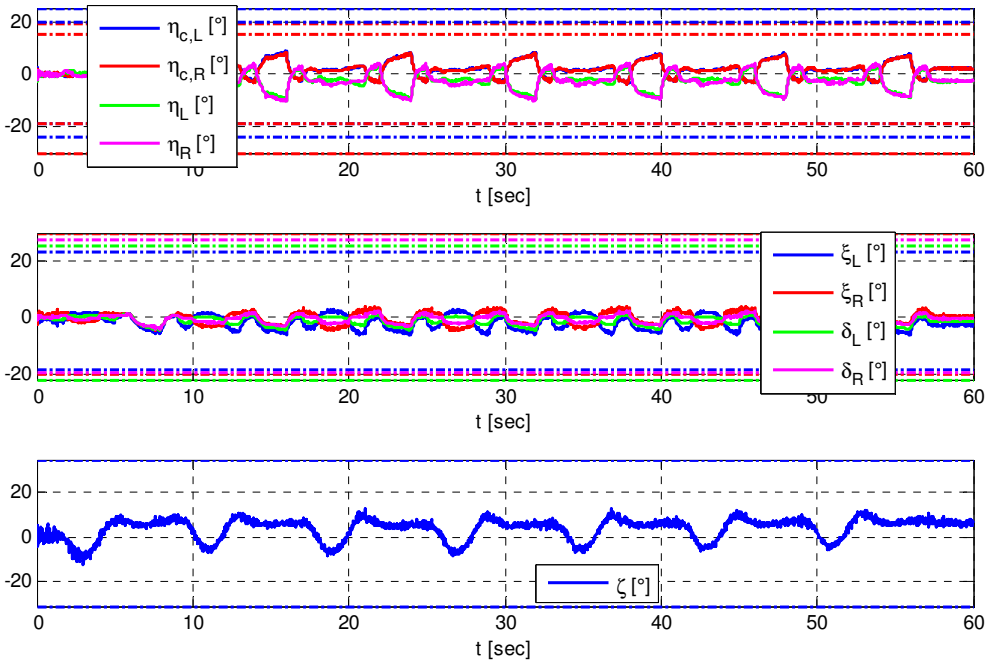


Figure 5.61 Simulation 6 – Control Surface Commands

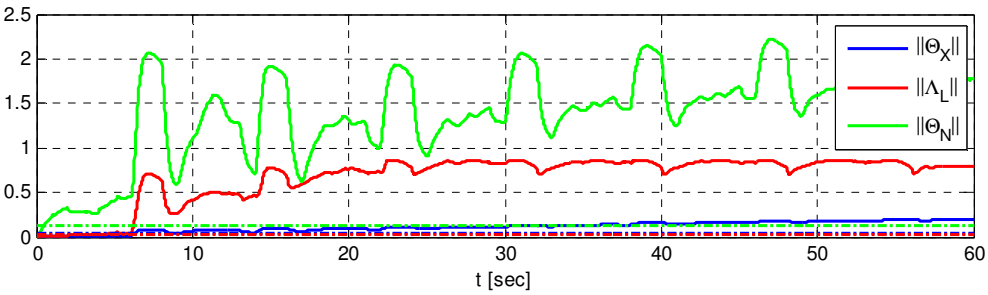


Figure 5.62 Simulation 6 – Weighted Frobenius Norms of Adaptive Parameters

Simulation Scenario 7: Damage Scenario 3, Emergency Button

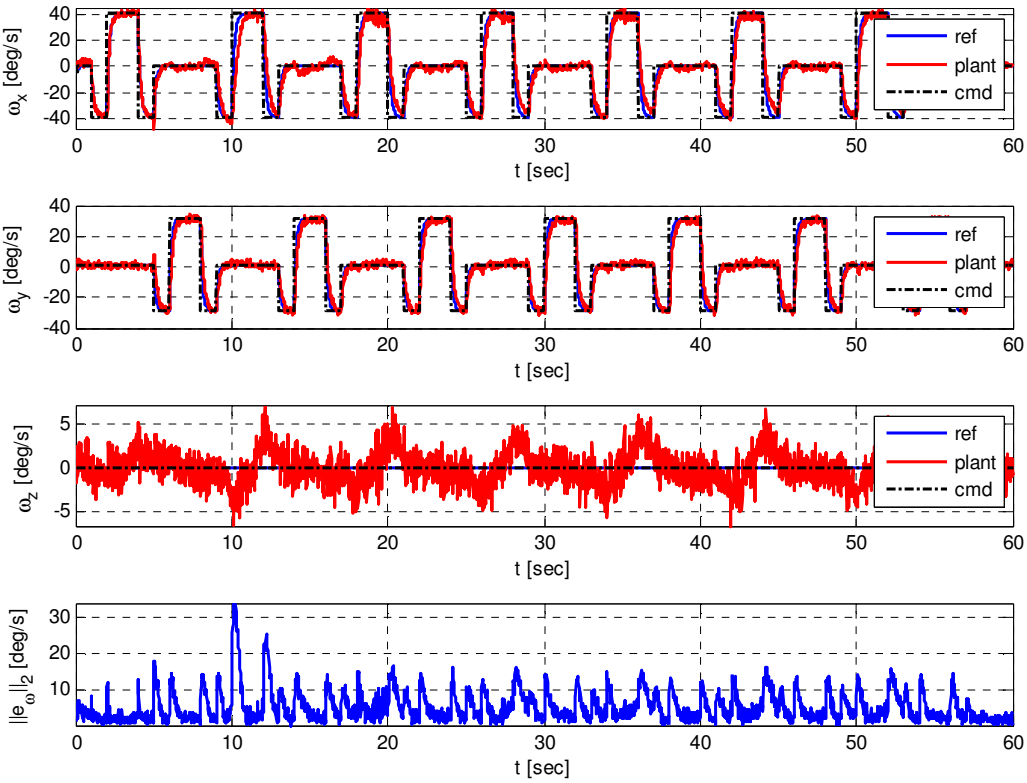


Figure 5.63 Simulation 7 – Tracking Error

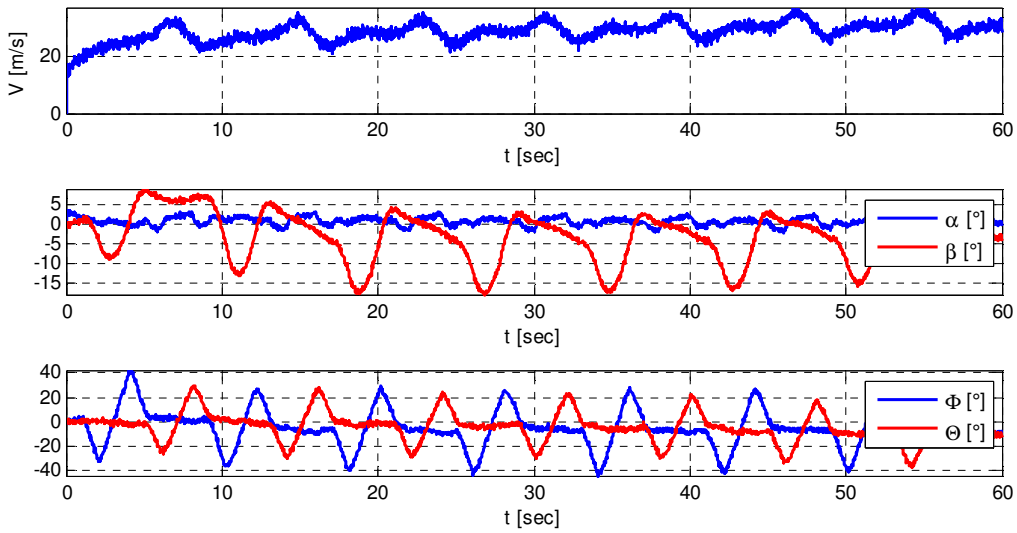


Figure 5.64 Simulation 7 – Internal States

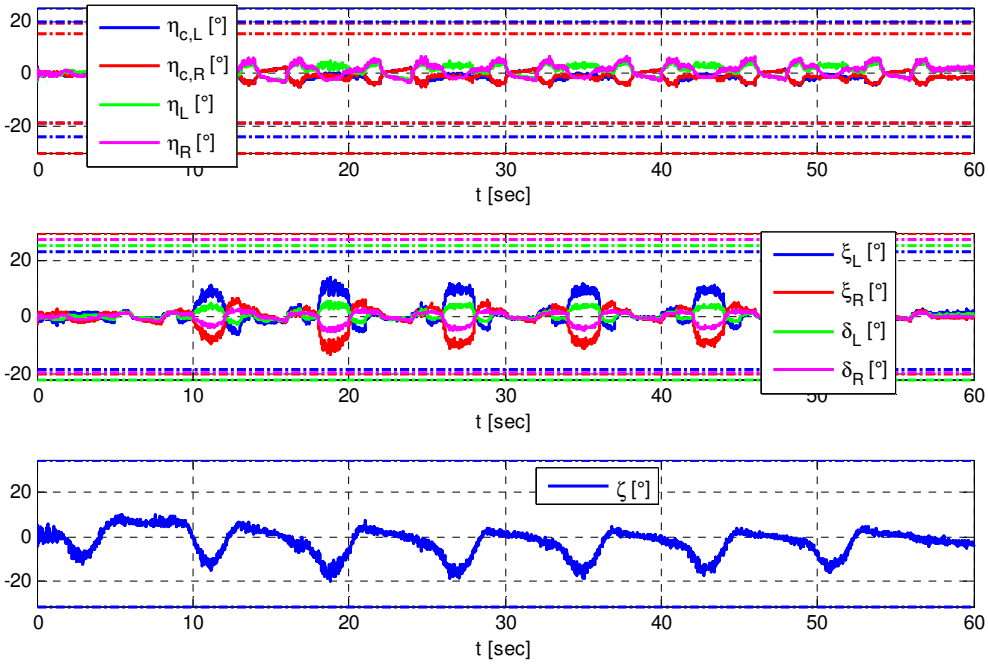


Figure 5.65 Simulation 7 – Control Surface Commands

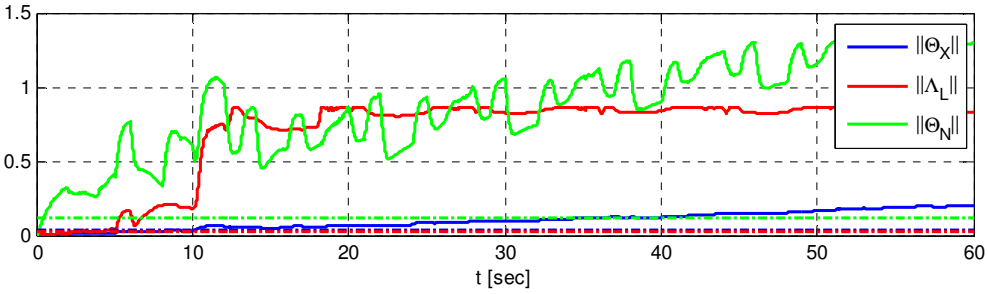


Figure 5.66 Simulation 7 – Weighted Frobenius Norms of Adaptive Parameters

Chapter 6

Conclusion and Outlook

In the context of the dissertation an MRAC system with an NDI baseline controller has been designed and implemented for a small scale unmanned aircraft testbed and the effectiveness of the developed control systems and algorithms have been shown by simulation scenarios where several actuator failures and even partial loss of the wing occurred during controlled flight. As it is indispensable to provide guaranteed bounds for the system states, if a flight control system is to be certified by the official authorities, a main emphasis throughout this work is the computation of explicit values within stability considerations, since it is felt by the author that this aspect is quite often neglected, if novel adaptive control concepts are developed.

Control System Design

The control system is designed, based on a detailed derivation of the involved equations of motion. After an introduction of the basic form of NDI, some extensions are derived, which are problem specific for the aircraft testbed, such as redundant control surfaces and the incorporation of nonaffine controls. Two variants for the utilization of the nonaffine controls are proposed, which are both augmented by an adaptive MRAC scheme. In the first variant, the nonaffine controls are rather considered as disturbance, which is cancelled by the linearizing state feedback and in the second variant, the nonaffine controls are used for feedback linearization of the plant dynamics, using nonlinear-in-control (NIC) design. Unfortunately, simulation revealed that the latter approach is not suitable for the aircraft, since a minimum control authority of the nonaffine thrust vector controls cannot be guaranteed uniformly within the whole flight envelope, but this is not necessarily the case for all types of aircraft. Nevertheless, the necessary equations have been derived, based on a precise statement on the preconditions for the system to work properly and a proof of ultimate boundedness in presence of unmatched uncertainties has been presented. The state of the technology for the NIC scheme has been cited but, moreover, an additional condition on the existence of the control map inverse, which is easier to verify in practice, has been derived. Up to now, stability analysis of the NIC scheme has been done, invoking the results of Thikonov's theorem. Yet, the latter does not provide an explicit value for the time scale separating factor ε of the gradient based online tuning algorithm, since the theorem is formulated for a generic SP model. The proof,

presented within the thesis, takes advantage of the special structure of the NIC equations to compute an explicit upper bound on ε, using the reformulated ultimate boundedness theorem in Chapter 3.

The internal dynamics have not been incorporated into the stability consideration, but were assumed to be bounded a priori. In case of the aircraft testbed, the assumption is justified, since the aircraft configuration is stable and AoA, AoS stay bounded for bounded inputs. Although some facts on stability of internal dynamics are cited, which are derived from the assumption of the system to be minimum phase, the results are rather of theoretic interest, since they do not provide explicit bounds. Theorem 4.3, which is based on the reformulated boundedness theorem, delivers explicit bounds, for the NDI tracking system, however, is does not incorporate the adaptive part. Summing up, an incorporation of the internal dynamics into stability analysis of the adaptive closed loop system remains an open item for future work.

Also, the actuators have not been considered within the stability analysis. PCH is an existing measure to account for actuator dynamics implicitly by a feedback of the expected reaction deficit to the reference model, which hides the influence of actuators from the tracking error. However, PCH destroys the a priori stability of the reference model and in turn, the arguments of the boundedness proof, and an alternative stability analysis does currently not exist. Hence, incorporation of actuators into the proofs remains open.

Achievement: Reduction of Conservativeness

Lyapunov stability analysis for MRAC systems suffer from the fact that the results are rather conservative, leading to bounds on the system states that are too large to be considered relevant within a physically meaningful range. Hence, reduction in conservativeness is challenging in adaptive control and several approaches have been derived within the thesis. It is physically clear that one cannot demand an infinite adaptation capability of the adaptive system. In descriptive words, a control system, designed for an aircraft cannot adapt to the dynamics of a stone while keeping the tracking error within tight bounds. Consequently, the more capabilities are demanded from the adaptive scheme, the larger the bounds. In our case, the linearizing state feedback cancels the known part of the dynamics, while the remaining uncertainties are cancelled by the adaptive part. A more accurate model allows reduction of the bound on the adaptive parameters – which indeed reduces the demand on the adaptive system – such that the bound on the tracking error, at least, reaches a physically relevant value. Nevertheless, this is not the end of the story. In fact, it is necessary to reduce bounds further.

The projection operator that is described in Appendix D, is actually an extension of the one, presented in [Hov10], but here it is formulated for parameter matrices rather than

for vectors and it allows the incorporation of a weighting matrix for computation of the Frobenius norms of the parameter matrices. This enables a scaling (deformation) of the allowable parameter space and hence a reduction of the parameter bounds. In the simulation examples, the weighting matrices have been set before design of the adaptive gains and computation of the bound on the tracking error. Hence, a possible measure to reduce the bounds further is an incorporation of the weighting matrices into the gain design procedure.

Particularly in case of the body-fixed angular rates controller, the maximum rotation speed depends on the respective axes, where roll axis is the fastest and yaw axis is the slowest. Now, the bound on the tracking error is computed in terms of the Euclidean vector norm, leading to a ball in $\mathbb{R}^3$ of allowable tracking errors. In other words, all three axes have the same amplitude. A weighting, analogously to the parameters could fit the allowable set to practical requirements, leading to less conservative results.

Another possibility for reduction of conservativeness is given by Theorem 3.1 for ultimate boundedness. Therein, upper bounds on the initial conditions are computed by the constants ρ_1, $\rho_2>0$, which have to fulfill

$$\beta_1(\rho_1)+\beta_2(\rho_2)\leq \underline{\underline{u}}$$

and the available “capacity” is distributed equally on both partitions, i.e.

$$\rho_1=\beta_1^{-1}\left(\frac{\underline{\underline{u}}}{2}\right) \quad , \quad \rho_2=\beta_2^{-1}\left(\frac{\underline{\underline{u}}}{2}\right).$$

Potentially, the state vector partitions reside in different scales, which motivates tailoring the distribution specifically to the demands of the state vector partitions by taking k_1, k_2 such that $k_1+k_2=1$ and defining

$$\rho_1=\beta_1^{-1}(k_1\underline{\underline{u}}) \quad , \quad \rho_2=\beta_2^{-1}(k_2\underline{\underline{u}}).$$

Novelty: MMQ Modification

A novel modification term for the adaptive laws, referred to as multi model Q modification, has been derived. It is actually an extension of q modification, which gains an expression for the uncertainty by utilizing the known terms within the plant dynamics. Effectiveness of q modification has already been proved in various publications, but simulations revealed that MMQ modification achieves an additional boost in adaptation performance, which is, of course, not guaranteed as for concurrent learning, but the probability increases with the number of employed filters. Unfortunately, MMQ modification is only effective in absence of unmatched uncertainties. The latter inhibits an exact identification of the matched uncertainty, using the plant equations, since it is not possible to distinguish between matched and

unmatched part. Hence, the modification term tries to compensate for matched and unmatched uncertainty simultaneously, which effects a contradiction, since the adaptive parameters can only compensate for the matched part.

Novelty: SVD Update Algorithm

The development of the SVD update algorithm is motivated by the constraint of the control effectiveness estimate to diagonal matrices with positive diagonal entries in predictor based MRAC systems. Although, this goal has been achieved, there are still some open questions to be clarified. At first, the SVD update needs computation of the eigenvectors of a matrix, which requires for an algorithm involving QR iteration. Although QR iteration is very efficient, an upper bound on number of computation steps cannot be guaranteed which renders the real-time capability of the algorithm problematic. However, the good news is that eigenvectors can, of course, be computed analytically, for small matrices up to dimension 3, which is sufficient for standard aircraft configuration with three primary controls – elevator, aileron, and rudder.

Another critical point is the numerical handling for the case of equal singular values (SV). If SVs approach each other, the time derivatives grow unbounded, which causes numerical problems. The open topic is a specification of the minimum distance for singular values that are considered equal. This distance should not be too large, since this increases the systematic error and it should also nut bee too small, since this leads to large time derivatives and, in consequence, to numerical problems using discrete time integration. Hence, further investigations are necessary to obtain a trade-off between the conflictive requirements. The last topic is the incorporation of the SVD update into the stability analysis. Contrary to projection operator, which can be included into stability analysis due to its convexity and the related properties, derived in Appendix D.1, an analog for the SVD update has to be established.

Perspectives towards Certifiable Adaptive Flight Control Systems

The reformulation of the ultimate boundedness theorem, which allows explicit computation of the bounds, is a main contribution of the thesis. The motivation behind is the hope to gain guaranteed results which are accepted by certification authorities. Although it is tailored to MRAC systems, its formulation is such generic, that is could also be used for analysis of other dynamic systems, which comprise a partition of the state vector. Nevertheless, the MRAC specific formulation is intended to reduce conservativeness of Lyapunov proofs within MRAC systems.

The main challenge towards certifiable adaptive flight control systems is finding robustness metrics that provide equivalent confidence for nonlinear systems as phase

and gain margin provide for linear systems. On the one hand, adaptive systems have to be robust against unmatched uncertainties and the ultimate boundedness theorem, stated here, is suitable to account for this. Within the proposed flight control systems, it provides guaranteed bounds, if the unmatched uncertainty does not exceed the assumed upper bound. On the other hand, the adaptive system also has to be robust against unmodeled dynamics. Such as phase margin accounts for unmodeled dynamics for linear system, time delay margin is an analog metric for nonlinear systems. In order to analyze time delay margins of nonlinear systems, an extension of Lyapunov's methods to functional differential equations (i.e. differential equations, whose time derivative does not only depend on the current but also on past states), also referred to as Lyapunov-Krasovskii ([GuK03]) functional, is a suitable tool. In [GuK09] a theorem for asymptotic stability, based on a Lyapunov-Krasovskii functional for a type of systems, referred to as "coupled differential difference equations", is presented. The considered system promises to fit the structure of adaptive control systems with time delay. It is a conjecture that the theorem therein only allows consideration of time-delay but not time-delay and unmatched uncertainty simultaneously for reasons, stated in the following. The theorem in [GuK09] guarantees asymptotic stability, if the time derivative of the Lyapunov Krasovskii functional is negative. It appears to be the analog to Lyapunov's direct method for asymptotic stability. However, in presence of unmatched uncertainties, negative definiteness of the Lyapunov function derivative is only assured outside a bounded set, which allows concluding boundedness of the system states as done in Theorem 3.1. Necessary for the simultaneous consideration of unmatched uncertainty and time-delay is the analog of the Lyapunov ultimate boundedness theorem for Lyapunov-Krasovskii functionals. If such a theorem exists, adaptive control systems could be analyzed for simultaneous robustness against unmatched uncertainty and time-delay and those two metrics could be established as a nonlinear analog for gain and phase margin.

Bibliography

[Aka12] *AkaModell München*. (n.d.). Retrieved November 07, 2012, from www.akamodell.vo.tum.de/

[And09] Anderson, R. T., Chowdhary, G., & Johnson, E. N. (2009). Comparison of RBF and SHL Neural Network Based Adaptive Control. *Journal of Intelligent and Robotic Systems* .

[Xse09] B.V., Xsens Technology;. (2009, May 27). *MTi-G User Manual and Technical Documentation.* Retrieved November 21, 2012, from http://amtechs.co.jp/2_gps/pdf/MTi-G/User/Manual/and/Technical/Documentation.pdf

[Bau12] Bauhaus Luftfhart. (n.d.). Retrieved November 08, 2012, from http://www.bauhaus-luftfahrt.net/

[Bau10] Baur, S. (2010). *Simulation and Adaptive Control of a High Agility Airplane in the Presence of Severe Structural Damage and Failures.* Diplomathesis, Technical University of Munich, Institut of Flight System Dynamics, Munich.

[Ben10] Bendat, J., & Piersol, A. (2010). *Random Data - Analysis and Measurement Procedures.* John Wiley & Sons.

[Smi10] Betts, J. T. (2010). Practical Methods for Optimal Control and Estimation Using Nonlinear Programming. In R. C. Smith, *Advances in Design and Control* (2nd ed.). Philadelphia, PA: Society for Industrial and Applied Mathmeatics (SIAM).

[Bie10] Bierling, T., Höcht, L., Holzapfel, F., Meier, R., & Wildschek, A. (2010). Comparative Analysis of MRAC Architectures in a Unified Framework. *Proceedings of the Guidance Navigation and Control Conference and Exhibit.* AIAA.

[Bie13] Bierling, T., Höcht, L., Merkl, C., Holzapfel, F., & Meier, R. (2013). Similarities of Hedging and L1 Adaptive Control. *European Aerospace GCN Conference.* Delft, Netherlands: CEAS.

[Boy09] Boyd, S., & Vandenberghe, L. (2009). *Convex Optimization.* New York: Cambridge University Press.

[Bri01] Brinker, J. S., & Wise, K. A. (2001, September-October). Flight Testing of Reconfigurable Control Law on the X-36 Tailless Aircraft. *Journal of Guidance Control and Dynamics , Vol.24* (No.5), pp. 903-909.

[Bri98] Brinker, J. S., & Wise, K. A. (1998). Reconfigurable Flight Control for a Tailless Advanced Fighter Aircraft. *Proceedings of the Guidance Navigation and Control Conference and Exhibit.* AIAA.

[Byr91] Byrnes, C., & Isidori, A. (1991, October). Asymptotic Stabilization of Minimum Phase Nonlinear Systems. *Transactions on Automatic Control , Vol. 36* (No. 10).

[Cho102] Chowdhary, G. (2010). *Concurrent Learning for Convergence in Adaptive Control without Persistency of Excitation.* Dissertation, Georgia Intstitute of Technology, Daniel Guggenheim School of Aerospace Engineering, Atlanta, Georgia.

[Cho10] Chowdhary, G., & Johnson, E. (2010). Concurrent Learning for Convergence in Adaptive Control without Persistency of Excitation. *Conference on Decision and Control.* Atlanta, GA: IEEE.

[Cho101] Chowdhary, G., & Johnson, E. N. (2010). Concurrent Learning for Improved Convergence in Adaptive Flight Control. *Guidance. Navigation and Control Conference.* Toronoto, Ontario, Canada: AIAA.

[Cho09] Chowdhary, G., & Johnson, E. N. (2009). Flight Test Validation of a Neural Network based Long Term Learning Adaptive Flight Controller. *Guidance, Navigation and Control Conference.* Chigago, Illinois: AIAA.

[Dör08] Döring, F., & Döring, P. (2008). *Elektrische Maschinen und Antriebe.* Wiesbaden: Vieweg.

[EUR98] EUROCONTROL, European Organization for the Safety of Air Navigation, Brussels, Belgium; Institute of Geodesy and Navigation, University FAF Munich, Germany. (1998, Feb 12). *WGS 84 IMPLEMENTATION.* Retrieved May 29, 2013, from http://www.dqts.net/files/wgsman24.pdf

[Far99] Farrel, J., & Barth, M. (1999). *The Global Positioning System & Inertial Navigation.* New York: McGraw-Hill.

[Fun89] Funahashi, K.-I. (1989, September). On the Approximate Realization of Continuous Mappings by Neural Networks. *Neural Networks* , pp. 183-192.

[Gel01] Gelb, A., Kaspar, J. F., Nash, R. A., Price, C. F., & Sutherland, A. (2001). *Applied Optimal Estimation* (16. ed.). (A. Gelb, Ed.) Cambridge, Massachusetts, London: The M.I.T. Press.

[Gol01] Goldstein, H., Poole, C. P., & Safko, J. L. (2001). *Classical mechanics* (3rd ed.). Addison Wesley.

[Gol85] Golub, G. H., & Loan, C. F. (1985). *Matrix Computations* (3rd ed.). Baltimore, Maryland, USA: The John Hopkins University Press.

[GuK09] Gu, K., & Liu, Y. (2009). LyapunovKrasovskii functional for uniform stability of coupled differential-functional equations. *Automatica , Vol. 45* (No. 3).

[GuK03] Gu, K., Kharitonov, V. L., & Chen, J. (2003). *Stability of Time-Delay Systems.* Boston: Birkhäuser Boston.

[Had08] Haddad, W. M., & Cellaboina, V. (2008). *Nonlinear dynamical systems and control. A Lyapunov-based approach.* New Jersey: Princeton University Press.

[Hal93] Hale, J. K., & Lunel, S. M. (1993). Introduction to Functional Differential Equations. In F. John, J. Mardsen, & L. Sirovich (Eds.), *Applied Mathematical Sciences.* New York: Springer-Verlag Inc.

[Has10] Hassel, R. v. (2010, 08 27). Retrieved 12 07, 2011, from http://www.win.tue.nl/~rvhassel/Onderwijs/2DE08-1011/ConTeXt-OWN-FA-201209-Bib/OwnLectureNotesFA.pdf

[Hei82] Heikinen, M. (1982). Geschlossene Formeln zur Berechnung räumlicher geodätischer Koordinaten aus rechtwinkligen Koordinaten. *Zeitschrift für Vermessungswesen , Vol. 5*, S. pp. 207-211.

[Hen90] Henson, M. A., & Seborg, D. E. (1990, November). Input-Output Linearization of General Nonlinear Processes. *AIChE Journal , Vol.36* (No.11), pp. 1753-1757.

[Hen96] Henson, M. A., & Seborg, D. E. (1996). *Nonlinear Process Control.* Englewood Cliffs, New Jersey.

[Hol04] Holzapfel, F. (2004). *Nichtlineare adaptive Regelung eines unbemannten Fluggerätes.* Dissertation, Technical University of Munich, Lehrstuhl für Flugmechanik und Flugregelung, Munich.

[Hol08] Holzapfel, F., & Höcht, L. (2008). *Flight Controls II.* Lecture Notes, Technical University of Munich, Institute of Flight System Dynamics, Munich.

[Hol12] Holzapfel, F., & Weingartner, M. (2012). *Flight System Dynamics II.* Lecture Notes, Technical University of Munich, Institute of Flight System Dynamics, Munich.

[Hor89] Hornik, K., Stinchcomb, M., & White, H. (1989, March). Multilayer Feedforward Networks are Universal Approximators. *Neural Networks* , pp. 359-366.

[Hov10] Hovakimyian, N., & Cao, C. (2010). *L1 Adaptive Control Theory - Gauranteed Robustness with Fast Adaptation.* Philadelphia, USA: Society for Industrial and Applied Mathematics.

[Hov06] Hovakimyian, N., Lavretsky, E., & Chengyu, C. (2006). Dynamic Inversion of Multi-input Nonaffine Systems via Time-scale Separation. *Proceedings of the 2006 American Control Conference.* Minneapolis, Minnesota.

[Ioa06] Ioannou, P., & Fidan, B. (2006). Adaptive Control Tutorial. In *Advances in Design and Control.* Philadelphia: Society for Industrial and Applied Mathematics.

[Isi95] Isidori, A. (1995). Nonlinear Control Systems. In B. Dickinson, A. Fettweis, J. Massey, & J. Modestino (Eds.), *Communications and Control Engineering Series* (3rd ed.). Springer.

[Jac92] Jack, C., & Chou. (1992). Quaternion kinematic and dynamic differential equations. *Transactions on Robotics and Automation , Vol. 8* (No. 1).

[Jat06] Jategaoncar, R. V. (2006). Flight Vehicle System Identification: A time Domain Methodology. In *Progress in Astronautics and Aeronautics: An American Institute of Aeronautics and Astronautics Series* (1st ed., Vol. 216). Academic Press.

[Joh00a] Johnson, E. N. (2000). *Limited Authority Adaptive Flight Control.* Dissertation, Georgia Institute of Technology, Atlanta, Georgia.

[Joh00b] Johnson, E. N. (2000). Pseudo Control Hedging: A new Method for Adaptive Control. *Advances in Navigation Guidance and Control Technology Workshop.* Redstone Arsenal, Alabama.

[Joh01] Johnson, E. N., & Calise, A. J. (2001). Neural Network Adaptive Control of Systems with Input saturation. *Proceedings of the Americal Control Conference.* Arlington, Virginia: IEEE.

[Joh04] Johnson, E. N., & Oh, S.-M. (2004). Adaptive Control Using Combined Online and Background Learning Neural Network. *Conference on Decision and Control.* Atlantis, Paradise Island, Bahamas: IEEE.

[Joh00c] Johnson, E. N., Calise, A. J., El-Shirbiny, H. A., & Rysdyk, R. T. (2000). Feedback Linearization with Neural Network Augmentation Applied to the X-33 Attitude Control. *Guidance Navigation and Control Conference and Exhibit.* Denver, CO: AIAA.

[Kap96] Kaplan, E., & Hegarty, C. (Eds.). (1996). *Understanding GPS: Principles and Applications* (2nd ed.). Norwood, MA: Artech House, Inc.

[Kha02] Khalil, H. K. (2002). *Nonlinear Systems* (3rd ed.). Upper Saddle River: Prentice Hall, Inc.

[Kle06] Klein, V., & Morelli, E. A. (2006). Aircraft System Identification: Theory and Practice. In J. A. Schetz (Ed.), *AIAA Education Series.* American Institute of Aeronautics and Astronautics, Inc.

[Lav08] Lavretsky, E., & Hovakimyian, N. (2008). Control of Uncertain Systems with Nonaffine Dynamics using Hamilton's Principle and Time Scale Separation. *Proceedings of the Guidance Navigation and Control Conference and Exhibit.* Honolulu, Hawaii: AIAA.

[Lav07b] Lavretsky, E., Chengyu, C., & Hovakimyian, N. (2007). Exponential Stability of Gradient Systems with Application to Nonlinear-in-Control Design Methods. *Proceedings of the American Control Conference.* New York City: AIAA.

[Lav09] Lavretsky, E., Gadient, R., & Gregory, I. M. (2009). Predictor-Based Model Reference Adaptive Control. *Proceedings of the Guidance, Navigation and Control Conference and Exhibit.* Chicago, Illinois: AIAA.

[Lav07a] Lavretsky, E., Hovakimyian, N., & Chengyu, C. (2007). Adaptive Design for Uncertain Systems with Nonlinear-in-Control Dynamics. *Proceedings of the Guidance Navigation and Control Conference and Exhibit.* Hilton Head, South Carolina: AIAA.

[Lüt96] Lütkepohl, H. (1996). *Handbook of Matrices.* England: John Wiley & Sohns Ltd.

[McS47] McShane, E. (1947). *Integration.* USA: Princeton University Press.

[Mic90] Michel, F. (1990). Generalized Controller Canonical Forms for Linear and Nonlinear Dynamics. *Transactions on Automatic Control , Vol. 35* (No. 9).

[Mul12] *Multiplex Homepage.* (kein Datum). Abgerufen am 07. November 2012 von http://www.multiplex-rc.de/fileadmin/content/dateien/downloads/md_twinstar_ii_5sp.pdf

[Nar05] Narandra, K. S., & Annaswamy, A. M. (2005). *Stable Adaptive Systems.* New York: Dover Publications Inc.

[Nar80] Narendra, K. S., Lin, Y.-H., & Valavani, L. S. (1980, June). Stable Adaptive Controller Design, Part II: Proof of Stability. *Transactions on Automatic Control , Vol. 25* (No. 3), pp. 440-448.

[Ngu08] Nguyen, N. T., Krishnakumar, K., & Boskovic, J. (2008). An Optimal Control Modification to Model-Reference Adaptive Control for Fast

Adaptation. *Guidance, Navigation and Control Conference and Exhibit.* Honolulu, Hawaii: AIAA.

[Nie60] Nielsen, J. N., Bussard, DeLauer, & Newell. (1960). Missile Aerodynamic. In H. G. Stever (Ed.), *Series in Missile and Space Technology.* New York, Totonto, London: McGraw-Hill Book Company.

[Nij90] Nijmejer, H., & van der Schaft, A. (1990). *Nonlinear Dynamical Control Systems* (3rd Ed. Ausg.). New York: Springer-Verlag.

[Oli13] Olivera da, O. (10. Dec. 2012). *Cornell University Library.* Abgerufen am 4. April 2013 von Cornell University Library: http://arxiv.org/abs/1212.2066

[Pre07] Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipies - The Art of Scientific Computing.* New York: Cambridge University Press.

[Rao09] Rao, S. S. (2009). *Engineering Optimization - Theory and Practice* (4th ed.). New Yersey: John Wiley & Sons, Inc.

[Röß12] Rößler, C. (2012). *Conceptual Design of Unmanned Aircraft with Fuel Cell Propulsion.* Dissertation, Technical University of Munich, Lehrstuhl für Luftfahrtsysteme, München.

[San80] Sandberg, I. W. (1980, November). Global Inverse Funcion Theorems. *Transactions on Automatic Control , Vol.27* (No.11), pp. 998-1004.

[Shi05] Shin, Y. (2005). *Neural Network Based Adaptive Control for Nonlinear Dynamic Regimes.* Dissertation, Georgia Institute of Technology, George W. Woodruff School of Mechanical Engineering.

[Sim06] Simon, D. (2006). *Optimal State Estimation.* New Jersey: John Wiley & Sons, Inc.

[Slo90] Slotine, J.-J., & Li, W. (1990). *Applied Nonlinear Control.* USA: Prentice Hall.

[Ste] Steiner, H.-J. (2010). *Prliminary Design Tool for Propeller-Wing Aerodynamics.* Technical University of Munich, Bauhaus Luftfahrt, Institut of Flight System Dynamics, Munich.

[Ste03] Stevens, B., & Lewis, F. (2003). *Aircraft Control and Simulation.* New Jersey: John Wiley & Sons, Inc.

[Sva06] Svaricek, F. (2006, July). *Nulldynamik linearer und nichtlinearer Systeme: Definitionen, Eigenschaften und Anwendungen.* Retrieved October 2012, from

http://forschung.unibw.de/papers/qqqrxvh9npz6f4dwfozyhyjsf24cj9.pdf

[Tao03] Tao, G. (2003). *Adaptive Contol Design and Analysis.* New Yersey, USA: John Wiley & Sons, Inc.

[Tre09] Treil, S. (2009). *Linear Algebra Done Wrong.* Abgerufen am 24. 05 2013 von http://www.math.brown.edu/~treil/papers/LADW/LADW.pdf

[Vol061] Volyanskyy, K. Y., Calise, A. J., & Yang, B.-J. (2006). A novel Q-Modification Term for Adaptive Control. *Proceedings of the 2006 American Control Conference.* Minneapolis, Minnesota: IEEE.

[Vol06] Volyanskyy, K. Y., Calise, A. J., Yang, B.-J., & Lavretsky, E. (2006). An Error Minimization Method in Adaptive Control. *Guidance, Navigation and Control Conference and Exhibit.* Keystone, Colorado: AIAA.

[Vol09] Volyanskyy, K. Y., Haddad, W. M., & Calise, A. J. (2009, Nov). A New Neuroadaptive Control Architecture for Nonlinear Uncertain Dynamical Systems: Beyond sigma- and e-Modifications. *Transactions on Neural Networks , Vol. 20* (No. 11), pp. 1707-1723.

[Wen07] Wendel, J. (2007). *Integrierte Navigationssysteme.* München: Oldenbourg Wissenschaftsverlag GmbH.

[Wol74] Wolovich, W. A. (1974). *Linear Multivariable Systems.* New York, USA: Springer.

[WuF72] Wu, F., & Desoer, C. (Mar 1972). Global inverse function theorem. *IEEE Transactions on Circuit Theory , Vol. 19* (No. 2).

[Yuc09] Yucelen, T. (2009). A Kalmann Filter Optimization Approach to Direct Adaptive Control. *Proceedings of the Guidance Navigation and Control Conference and Exhibit.* Chicago, Illinois: AIAA.

[Yuc091] Yucelen, T. (2009). A Loop Recovery Method for Adaptive Control. *Proceedings of the Guidance Navigation and Control Conference and Exhibit.* Chicago, Illinois: AIAA.

[Yuc092] Yucelen, T., Calise, A. J., Haddad, W. M., & Volyanskyy, K. Y. (2009). A Comparison of a New Neuroadaptive Controller Architecture with sigma- and e- Modification Architectures. *Proceedings of the Guidance Navigation and Control Conference and Exhibit.* Chicago, Illinois: AIAA.

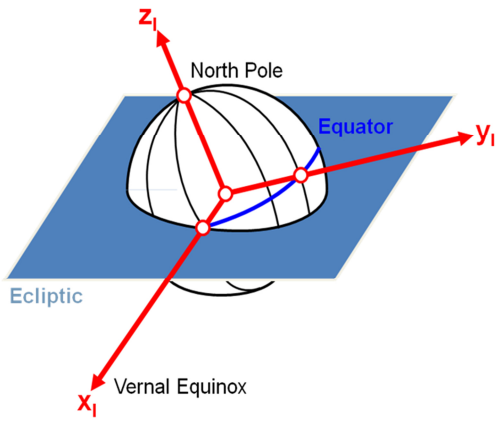
Appendix A

Frames and Transformations

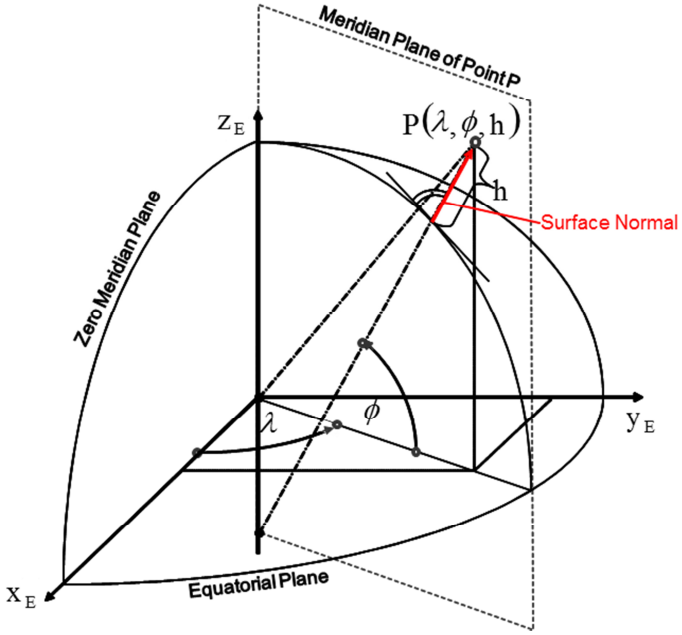
In the following, coordinate frames, relevant for aircraft modeling and control, are introduced. The nomenclature the one that is commonly used at the Institute of Flight System Dynamics but analogous frames can be found in corresponding aviation norms such as LN 9300. Every frame is associated with an index that is used throughout the thesis for nomenclature of vectors and transformation matrices. Additionally, if applicable, rotations between different frames are listed, too. E.g. rotation from frame I_1 to I_2, is denoted as $I_1 \rightarrow I_2$. Rotations are defined by a set of elementary rotations about coordinate axes, by some angle. Each rotation is associated with a positive or negative direction, indicating whether the rotation is done in a right hand (positive) of heft hand (negative) sense. The complete rotation is specified as a sequence of the elementary rotations.

The pictures within this chapter are taken from internal documents by courtesy of Institute of Flight System Dynamics.

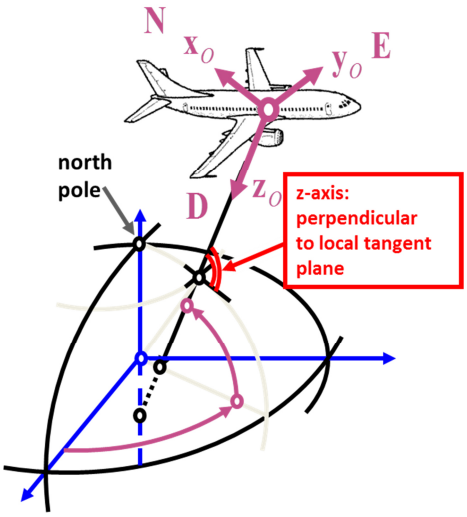
Earth Centered Inertial (ECI) frame	
frame index	I
origin	earth's center
origin index	0
x-Axis	in equatorial plane, pointing to vernal equinox
y-Axis	in equatorial plane, completing an orthogonal right hand system
z-Axis	earth rotation axis



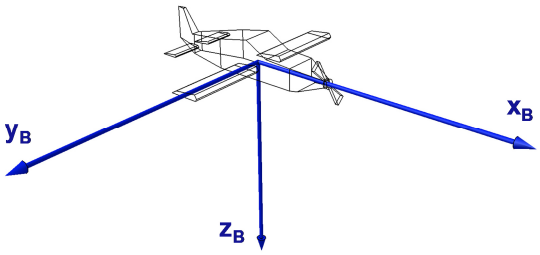
Earth Centered Earth Fixed (ECEF) frame	
frame index	E
origin	earth's center
origin index	0
x-axis	in equatorial plane, pointing to Greenwich meridian
y-axis	in equatorial plane, completing an orthogonal right hand system
z-axis	earth rotation axis



North-East-Down (NED) frame			
frame index	O		
origin	aircraft reference point		
origin index	R		
x-axis	parallel to local geoid surface, pointing to geographic north pole		
y-axis	parallel to local geoid surface, pointing eastward		
z-axis	pointing downwards, perpendicular to local geoid surface		
rotation: $E \rightarrow O$			
angle	symbol	axis of rotation	direction
$90°$	-	y-axis	positive
geodetic longitude	λ	x-axis	positive
geodetic latitude	ϕ	y-axis	negative
sequence		$90° - \lambda - \phi$	



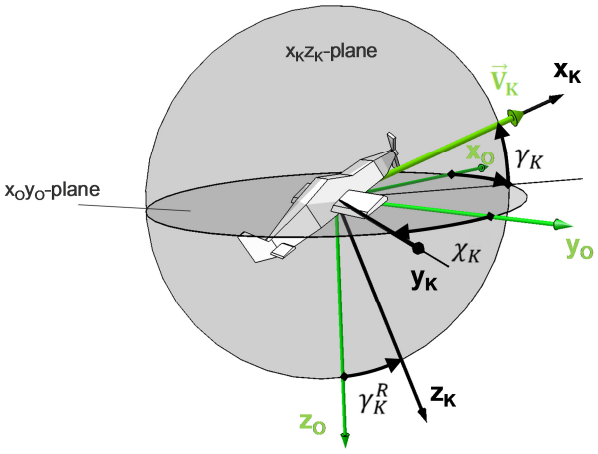
body-fixed frame			
frame index	B		
origin	aircraft reference point		
origin index	R		
x-axis	points to aircraft nose with in xz-symmetry plane		
y-axis	points to right wing, completing an orthogonal right hand system		
z-axis	points downwards within symmetry plane of aircraft, perpendicular to x and y axes		
rotation: $O \rightarrow B$			
angle	symbol	axis of rotation	direction
Euler azimuth	Ψ	z-axis	positive
Euler pitch angle	Θ	y-axis	positive
Euler bank angle	Φ	x-axis	positive
sequence		$\Psi - \Theta - \Phi$	



kinematic Frame	
frame index	K
origin	aircraft reference point
origin index	R
x-axis	pointing in direction of *kinematic* velocity
y-axis	pointing to the right, perpendicular to x and x axes, completing an orthogonal right hand system
z-axis	points downwards, parallel to projection of the local surface normal of the WGS-84 reference ellipsoid into the plane, perpendicular to the x-axis

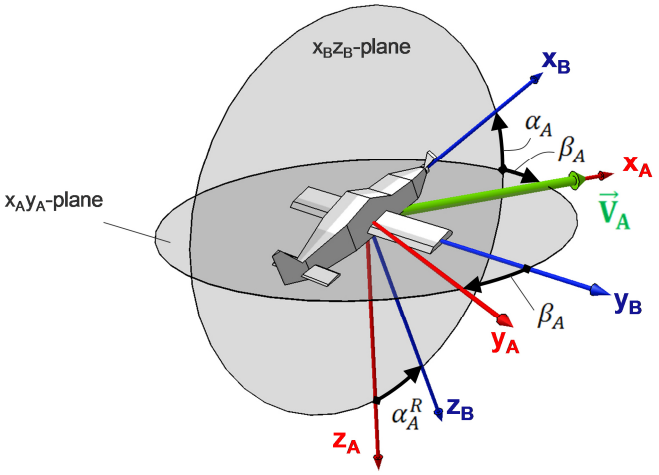
rotation: $O \rightarrow K$			
angle	symbol	axis of rotation	direction
kinematic course angle	χ_K	z-axis	positive
kinematic inclination angle	γ_K	y-axis	positive
sequence		χ_K - γ_K	

rotation: $K \rightarrow B$			
angle	symbol	axis of rotation	direction
kinematic bank angle	μ_K	z-axis	positive
kinematic angle of side-slip	β_K	y-axis	negative
kinematic angle of attack	α_K	x-axis	positive
sequence		μ_K - β_K - α_K	



modified kinematic frame	
frame index	$\overline{K}$
origin	aircraft reference point
origin index	R
orientation	equal to kinematic frame, but rotated about kinematic bank angle μ_K

aerodynamic frame			
frame index	A		
origin	aircraft reference point		
origin index	A		
x-axis	pointing in direction of *aerodynamic* velocity		
y-axis	pointing to the right, completing an orthogonal right hand system		
z-axis	points downwards within the symmetry plane of the aircraft, perpendicular to x and y axes		
rotation: $A \rightarrow B$			
angle	symbol	axis of rotation	direction
aerodynamic angle of side-slip (AoS)	β_A	z-axis	negative
aerodynamic angle of attack (AoA)	α_A	y-axis	positive
sequence		β_A - α_A	
rotation: $O \rightarrow A$			
angle	symbol	axis of rotation	direction
aerodynamic course angle	χ_A	z-axis	positive
aerodynamic inclination angle	γ_A	y-axis	positive
aerodynamic bank angle	μ_A	x-axis	positive
sequence		χ_A - γ_A - μ_A	



Appendix B

Mathematics

B.1 Analysis Basics

This section gives an overview over some basic facts from functional analysis, which might not be known by an engineer in-depth. Often such facts are often rather anticipated on an intuitive manner or simply they are simply read over. Many topics, pertaining to design and analysis of nonlinear systems, are mathematically quite involved as they e.g. consider abstract Banach spaces of functions and an in-depth understanding of the fundamentals, presented in the following, helps for a better understanding of the conclusions that are drawn there.

The ideas are taken from [Has10]. The formulations are kept as mathematically exact as necessary, but if it supports understanding, intuitive formulations are preferred.

B.1.1 Linear Spaces

Definition B.1 Linear Space

A set $\mathcal{X}$ is a *linear space* over the a field $\mathbb{K}$ if, for $\mathbf{x},\mathbf{y},\mathbf{z}\in\mathcal{X}$, $\alpha,\beta\in\mathbb{K}$

1. Addition: $\mathbf{x}+\mathbf{y}\in\mathcal{X}$ is defined with following properties
   a. commutativity: $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$
   b. associativity: $(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z})$
   c. there is a neutral element $\in\mathcal{X}$ such that: $\mathbf{x}+\mathbf{0}=\mathbf{x}$
   d. for every $\mathbf{x}$ there is an inverse element $(-\mathbf{x})$ such that $\mathbf{x}+(-\mathbf{x})=\mathbf{0}$
2. Scalar Multiplication: $\alpha\cdot\mathbf{x}\in\mathcal{X}$ is defined with the following properties
   a. associativity: $\alpha\cdot(\beta\cdot\mathbf{x})$
   b. for the neutral element 1 of the field $\mathbb{K}$: $1\cdot\mathbf{x}=\mathbf{x}$
   c. $\mathcal{X}$ distributes over $\mathbb{K}$ $(\alpha+\beta)\cdot\mathbf{x}=\alpha\cdot\mathbf{x}+\beta\cdot\mathbf{x}$
   d. $\mathbb{K}$ distributes over $\mathcal{X}$: $\alpha(\mathbf{x}+\mathbf{y})=\alpha\cdot\mathbf{x}+\alpha\cdot\mathbf{y}$

Definition B.2 Topological Space

A set $\mathcal{X}$ called a *topological space*, if a collection of subsets S on $\mathcal{X}$ is defined, which has the following properties. Let $s_i \in S$:

1. Basic and empty set belong to S : $\mathcal{X}, \varnothing \in S$
2. The intersection of a finite collection of sets s_i belongs to S : $\bigcap_{i=1}^{N}(s_i) \in S$
3. The union of any collection of sets in S belongs to S.

The elements of the set S are sets by themselves, and are called *open sets* (by definition). S is called *topology* on $\mathcal{X}$.

Remark

The important fact is that an *intersection* of an only *finite* number of open sets is again an open set, while the *union* of an *infinite* number of open sets in again an open set.

If we intersect an infinite number of open sets, then the result is possibly not an open set. The definition of open sets for a topological space conforms to our intuition for open sets. Let e.g. for the linear space of real numbers $\mathcal{X} = \mathbb{R}$, if we define the topology S as set of all open intervals

$$S = \{(a,b) \mid a,b \in \mathbb{R},\ a < b\}$$

Then it fulfills properties 1-3. The next example gives an illustrative understanding, why the definition of "*open set*" (which is supposed to be a generalization of our intuitive open set) for a topological space is reasonable.

Example

- Let an open interval: $s_i = \left(\frac{i}{2(i+1)}, \frac{2(i+1)}{i}\right),\ i \in \mathbb{N}$
- If we intersect a finite number of s_i, lets' say $i = 1,\ldots,N$, then the result is:
  $$\bigcap_{i=1}^{N}\left(\frac{i}{2(i+1)}, \frac{2(i+1)}{i}\right)$$
  $$= \left(\frac{1}{4},4\right) \cap \left(\frac{1}{3},3\right) \cap \ldots \cap \left(\frac{N}{2(N+1)}, \frac{2(N+1)}{N}\right) = \left(\frac{N}{2(N+1)}, \frac{2(N+1)}{N}\right) \in S$$
  clearly an open set
- If we unite all s_i, $i = 1,\ldots,\infty$, then the result is: $\bigcup_{i=1}^{\infty}\left(\frac{i}{2(i+1)}, \frac{2(i+1)}{i}\right) = \left(\frac{1}{4},4\right) \in S$ an open interval
- If we, however intersect an infinite number of s_i, $i = 1,\ldots,\infty$, then the result is: $\bigcap_{i=1}^{\infty}\left(\frac{i}{2(i+1)}, \frac{2(i+1)}{i}\right) = \left(\frac{1}{4},4\right) \cap \left(\frac{1}{3},3\right) \cap \ldots = \left[\frac{1}{2},2\right]$ NOT an open interval.

Definition B.3 Topological Linear Space

A set *linear space* $\mathcal{X}$ is a *topological linear space*, if a *topology* $\mathcal{S}$ is defined on $\mathcal{X}$

Definition B.4 Metric

A *metric* or *distance function* is defined on a *topological space* $\mathcal{X}$ and represents the distance of two elements $\mathbf{x},\mathbf{y}\in\mathcal{X}$.

The distance function $d(\mathbf{x},\mathbf{y}):\mathcal{X}\times\mathcal{X}\to\mathbb{R}$ has to fulfill the following properties. Let $\mathbf{x},\mathbf{y},\mathbf{z}\in\mathcal{X}$:

1. positivity: $0\le d(\mathbf{x},\mathbf{y})<\infty,\ d(\mathbf{x},\mathbf{y})=0\Leftrightarrow\mathbf{x}=\mathbf{y}$
2. symmetry $d(\mathbf{x},\mathbf{y})=d(\mathbf{y},\mathbf{x})$
3. triangle inequality: $d(\mathbf{x},\mathbf{y})\le d(\mathbf{x},\mathbf{z})+d(\mathbf{z},\mathbf{y})$

Definition B.5 Metric Space

A *metric space* is a *topological linear space* $\mathcal{X}$ where a metric $d(\cdot,\cdot)$ is defined on it and the topology is induced by the metric such that the *open sets* are formed by *open balls* of some $\mathbf{x}_0\in\mathcal{X}$ with radius $r>0$.

$$\mathcal{S}=\{\mathcal{B}_r(\mathbf{x}_0)\},\ \mathbf{x}_0\in\mathcal{X},\ r>0$$

Remarks

Since the topology is induced by the metric, the rather abstract topological definition of an open subset substantiates to, what one intuitively associates with an open set:

- In a metric space, a subset $\mathcal{M}$ is *open* if and only if, for every $\mathbf{x}\in\mathcal{M}$, a sufficiently small open ball $\mathcal{B}_\varepsilon(\mathbf{x})$ entirely lies in $\mathcal{M}$. An open subset thus consists only of interior points (see Definition B.14).
- All open balls $\mathcal{B}_r(\mathbf{x})$ (refer to Definition B.6) are open sets in the topological definition, since intersection of a finite number and unity of any open balls is again an open set, that does not have any boundary points.

Definition B.6 Open Ball

For a metric space $\mathcal{X}$ and some $\mathbf{x}_0\in\mathcal{X}$, an open ball around $\mathbf{x}_0$ with radius $r>0$ is defined as: $\mathcal{B}_r(\mathbf{x}_0)=\{\mathbf{x}\in\mathcal{X}\,|\,d(\mathbf{x},\mathbf{x}_0)<r\}$

Definition B.7 Closed Ball

For a metric space $\mathcal{X}$ and some $\mathbf{x}_0 \in \mathcal{X}$, a closed ball around $\mathbf{x}_0$ with radius $r>0$ is defined as: $\overline{\mathcal{B}}_r(\mathbf{x}_0) = \{\mathbf{x} \in \mathcal{X} \mid d(\mathbf{x}, \mathbf{x}_0) \leq r\}$

Definition B.8 Norm

A *norm* is defined on a *topological linear space* $\mathcal{X}$ and represents the *length* of its elements $\mathbf{x} \in \mathcal{X}$. It is a function $\|\mathbf{x}\| : \mathcal{X} \to \mathbb{R}$ and has to fulfill the following properties. Let $\mathbf{x}, \mathbf{y} \in \mathcal{X}$, $\alpha \in \mathbb{K}$

1. positivity: $0 \leq \|\mathbf{x}\| < \infty$, $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
2. homogeneity: $\|\alpha \cdot \mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$
3. triangle inequality: $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

Remark: An indirect result of the triangle inequality is: $\|\mathbf{x} + \mathbf{y}\| \geq \big|\, \|\mathbf{x}\| - \|\mathbf{y}\| \,\big|$

Definition B.9 Normed Space

A *normed space* is a *topological vector space* $\mathcal{X}$, where a norm $\|\cdot\|$ is defined on it

Remark

- A normed space is also a metric space, where the metric is induced by the norm such that: $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$
- As in the metric space:
  - The open sets consists of open balls $\mathcal{B}_r(\mathbf{x})$, where $r>0$ and $\mathbf{x} \in \mathcal{X}$.
  - A subset $\mathcal{M}$ is open if and only if, for every $\mathbf{x} \in \mathcal{M}$ for a sufficiently small ε, there is an open ball $\mathcal{B}_\varepsilon(\mathbf{x}) \subset \mathcal{M}$ and thus consists only of interior points (see Definition B.14).

Example

For $\mathcal{X} = \mathbb{R}^2$, a disc with radius r around the origin without its boundary is an open subset: $\mathcal{M} = \{\mathbf{x} \in \mathcal{X} \mid \|\mathbf{x}\| < r\}$

Definition B.10 Banach Space

Let $\mathcal{X}$ be a *normed linear space*. Then $\mathcal{X}$ is a *Banach space*, if it is *complete.*

Remarks

- Completeness is defined in Definition B.20
- Thus in a Banach space, all Cauchy sequences converge by definition, but this is not necessarily true for normed spaces.

The following picture is inspired by [Has10] and illustrates the relation between the spaces, introduced so far.

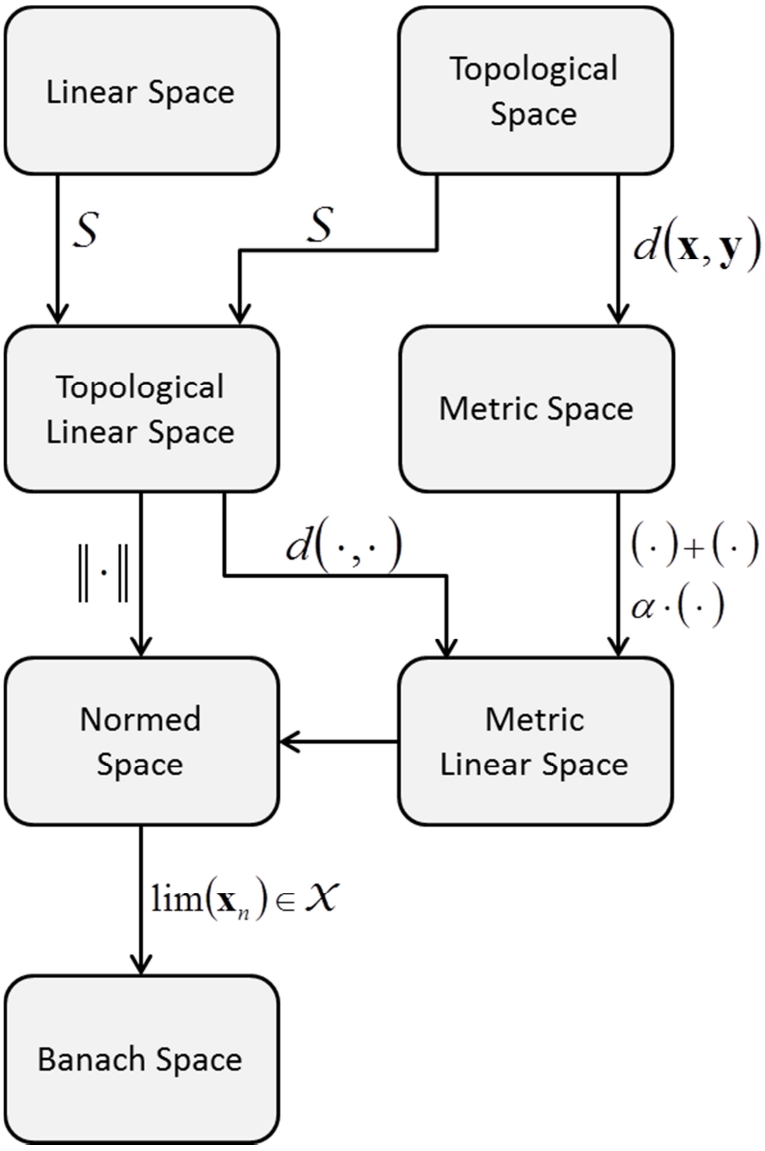


Figure B.1 Relationships between Spaces

B.1.2 Subsets of Spaces

DefinitionB.11 Complement

Let $\mathcal{M}$ be a subset of a space $\mathcal{X}$ then the complement of $\mathcal{M}$ is defined as:

$$\mathcal{M}^c = \mathcal{X} \setminus \mathcal{M}$$

Definition B.12 Bounded Subset

A subset $\mathcal{M}$ of a metric space $\mathcal{X}$ is *bounded*,
if for some $k > 0$ and some $\mathbf{x}_0 \in \mathcal{X}$: $d(\mathbf{x}, \mathbf{x}_0) \leq k$ for all $\mathbf{x} \in \mathcal{M}$

Definition B.13 Totally Bounded Subset

A subset $\mathcal{M}$ of a metric space $\mathcal{X}$ is *totally bounded*, if for every $\varepsilon > 0$, there is a finite collection of open balls $\{\mathcal{B}_\varepsilon(\mathbf{x}_i)\}_i$, $i = 1, \ldots, N$, $N \in \mathbb{N}$, $\mathbf{x}_i \in \mathcal{X}$, that completely covers $\mathcal{M}$.

Remark

The difference between bounded and totally bounded is secondary for our purposes. Important to know is, that in a Banach space of finite dimension, *boundedness* and *total boundedness* are equivalent.

Definition B.14 Interior and Boundary

Let $\mathcal{M}$ be a subset of a metric space $\mathcal{X}$. Some $\mathbf{x} \in \mathcal{X}$ is

1. *interior point* of $\mathcal{M}$, $\mathcal{B}_r(\mathbf{x}) \subset \mathcal{M}$ for some $r>0$
2. *boundary point* of $\mathcal{M}$, if every $\mathcal{B}_r(\mathbf{x})$ has elements inside and outside of $\mathcal{M}$

Remark

An inner point of a set $\mathcal{M}$ always belongs to $\mathcal{M}$, while a boundary point does not necessarily belong to $\mathcal{M}$.

Definition B.15 Closed Subset

A subset $\mathcal{M}$ of a topological space $\mathcal{X}$ is *closed*,

if and only if its complement $\mathcal{M}$ is *open*.

Remark

Since $\mathcal{X}$ and $\varnothing$ are both open sets, in the sense of the topological definition, both sets are closed as well by Definition B.15. In literature, such sets are denoted as “clopen”.

B.1.3 Sequences and Convergence

Considering sequences, their behavior as n tends to ∞ is of special interest and the question whether the sequence converges or not. In order to talk about convergence of sequences, we need at least a metric space, since convergence is defined by means of the distance of the sequence elements $\{\mathbf{x}_n\}$ and its limit point $\mathbf{x}_0$. Of course, convergence is also meaningful for a normed space since it is also a metric space.

Definitions

Definition B.16 Sequence

Given set $\mathcal{X}$:

An *unlimited ordered subset* set $\{\mathbf{x}_n\}$ in $\mathcal{X}$ is called *sequence*, where the indices $n \in \mathbb{N}$ determine the order.

Definition B.17 Convergent Sequence

For a metric space $\mathcal{X}$:

A sequence $\{\mathbf{x}_n\}$ in $\mathcal{X}$ is *convergent* with limit $\mathbf{y} \in \mathcal{X}$ if, for every $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that $d(\mathbf{x}_n, \mathbf{y}) < \varepsilon$ for all $n \geq N$

If the limit element $\mathbf{y}$ is unknown, the definition of convergence can also be done without $\mathbf{y}$, which leads to the definition of Cauchy sequence (Definition B.19).

Definition B.18 Subsequence

Let $\{\mathbf{x}_n\}$, $n \in \mathbb{N}$ be a sequence in a set $\mathcal{X}$. Then a *subsequence* is obtained by choosing an infinite number of elements n_i , $i \in \mathbb{N}$ from $\{\mathbf{x}_n\}$ without changing the order.

Definition B.19 Cauchy Sequence

A sequence $\{\mathbf{x}_n\}$ in $\mathcal{X}$ is said to be a Cauchy sequence, if for every $\varepsilon > 0$, there is a $N \in \mathbb{N}$, such that $d(\mathbf{x}_n, \mathbf{x}_m) < \varepsilon$ for all $n, m \geq N$.

Remark

Every convergent sequence is a Cauchy sequence, but not every Cauchy sequence is convergent. The reason is, that the limit element $\mathbf{y}$ is potentially not an element of the basic set $\mathcal{X}$.

$$\{\mathbf{x}_n\} \text{ convergent} \Rightarrow \{\mathbf{x}_n\} \text{ Cauchy sequence}$$

The question that naturally arises is the following. Under which conditions do all Cauchy sequences converge? Another issue is the convergence of Cauchy sequences in a subset. The following definition gives a part of the answer.

Definition B.20 Completeness

A metric space $\mathcal{X}$ is said to be complete,
if every Cauchy sequence $\{\mathbf{x}_n\}$ in $\mathcal{X}$ is convergent.

Remark

The completeness property can also be applied to subsets of metric spaces. A Subset $\mathcal{M}$ to a metric space $\mathcal{X}$ is said to be complete, if every Cauchy sequence in $\mathcal{M}$ converges in $\mathcal{M}$.

Definition B.21 Accumulation point

For a subset $\mathcal{M}$ of a metric space $\mathcal{X}$, an element $\mathbf{x}_0 \in \mathcal{X}$ is said to be an *accumulation point* of $\mathcal{M}$, if every open ball of $\mathbf{x}_0$: $\mathcal{B}_r(\mathbf{x}_0)$ with $r>0$

contains at least one element in $\mathcal{M}$ except for $\mathbf{x}_0$

Remark: If $\mathbf{x} \notin \mathcal{M}$, it must be on the boundary of $\mathcal{M}$ to be an accumulation point.

DefinitionB.22 Closure

Let $\mathcal{M}$ be a subset of a metric space. The *closure* of the subset, denoted as $\overline{\mathcal{M}}$ is defined as the union of $\mathcal{M}$ with all its *accumulation points*.

Definition B.23 Dense Subset

Let $\mathcal{M}$ and $\mathcal{N} \subset \mathcal{M}$ be subsets of a metric space. $\mathcal{N}$ is a dense subset of $\mathcal{M}$ if, for every $\mathbf{x} \in \mathcal{M}$, there is a sequence in $\mathcal{N}$, $\{\mathbf{y}_n\}$, $n \in \mathbb{N}$, such that $\lim_{n\to\infty}(\mathbf{y}_n) = \mathbf{x}$.

Example

The rational numbers $\mathbb{Q}$ are a dense subset of $\mathbb{R}$. E.g. for $\sqrt{2}$ which is in $\mathbb{R}$ but not in $\mathbb{Q}$, there is a sequence in $\mathbb{Q}$, which converges to $\sqrt{2}$.

Definition B.24 Compact Subset

Let $\mathcal{M}$ be a subset of a metric space. $\mathcal{M}$ is *compact* if every sequence $\{\mathbf{x}_n\}$, $n \in \mathbb{N}$ in $\mathcal{M}$ has a subsequence $\{\mathbf{x}_m\}$, $\{m\} \subset \{n\}$ that converges to an element in $\mathcal{M}$

Theorems

Theorem B.1 Accumulation Point and Convergence

Let $\mathcal{M}$ be a subset of a normed space.

Then $\mathbf{x}$ is an accumulation point of $\mathcal{M}$,

if and only if there is a sequence $\{\mathbf{x}_n\}$ in $\mathcal{M}$ which converges to $\mathbf{x}$.

Proof:

SUFFICIENT PART: If $\lim_{n\to\infty}(\mathbf{x}_n)=\mathbf{x}$ *then* $\mathbf{x}$ *is an accumulation point*

Suppose there is a sequence in $\mathcal{M}$ that converges to $\mathbf{x}$, then for every $\varepsilon>0$ there is an $N(\varepsilon)\in\mathbb{N}$ such that $\|\mathbf{x}_N-\mathbf{x}\|<\varepsilon$. Therefore, every open ball $\mathcal{B}_\varepsilon(\mathbf{x})$ contains an element $\mathbf{x}_N\in\mathcal{M}$. Thus $\mathbf{x}$ is an *accumulation point.*

NECESSARY PART: If $\mathbf{x}$ *is an accumulation point, then there is a sequence in* $\mathcal{M}$ *that converges to* $\mathbf{x}$*.*

Suppose that $\mathbf{x}$ is an accumulation point of $\mathcal{M}$. Then every open ball $\mathcal{B}_{1/n}(\mathbf{x})$ contains an element $\mathbf{x}_n$ in $\mathcal{M}$, which also lies in $\mathcal{B}_{1/n}(\mathbf{x})$. Thus we have constructed a sequence $\{\mathbf{x}_n\}$ in $\mathcal{M}$ which converges to $\mathbf{x}$.

□

Theorem B.2 Closed Subset and Convergence

Let $\mathcal{M}$ be a subset of a metric space $\mathcal{X}$

$\mathcal{M}$ is closed if and only if

every sequence $\{\mathbf{x}_n\}$ in $\mathcal{M}$, that converges in $\mathcal{X}$, also converges in $\mathcal{M}$.

Proof:

SUFFICIENT PART: If every convergent sequence $\{\mathbf{x}_n\}$ *in* $\mathcal{M}$ *has a limit* $\mathbf{x}\in\mathcal{M}$ *then* $\mathcal{M}$ *is a closed subset.*

The proof is done by contradiction. If $\mathcal{M}$ is not closed, then $\mathcal{M}^c$ is not open. Therefore there is an element $\mathbf{x}\in\mathcal{M}^c$ such that every open ball $\mathcal{B}_{1/n}(\mathbf{x})$, $n\in\mathbb{N}$, contains at least one element in $\mathbf{x}_n\in\mathcal{M}$ (see remark to Definition B.5). So the sequence $\{\mathbf{x}_n\}$ converges to $\mathbf{x}\notin\mathcal{M}$ which is a contradiction and thus $\mathcal{M}$ is a closed subset.

NECESSARY PART: If $\mathcal{M}$ is a closed set then every sequence $\{\mathbf{x}_n\}$ in $\mathcal{M}$ converges in $\mathcal{M}$

If $\mathcal{M}$ is closed then $\mathcal{M}^c$ is open by definition. Thus, if there is a sequence $\{\mathbf{x}_n\}$ in $\mathcal{M}$ that converges to $\mathbf{x}\in\mathcal{M}^c$, then there is an $\varepsilon>0$ such that the open ball $\mathcal{B}_\varepsilon(\mathbf{x})$: lies entirely in $\mathcal{M}^c$. This implies that $\|\mathbf{x}-\mathbf{x}_n\|\geq\varepsilon$ for all $n\in\mathbb{N}$ which contradicts $\lim_{n\to\infty}(\mathbf{x}_n)=\mathbf{x}$ and thus $\{\mathbf{x}_n\}$ converges in $\mathcal{M}$.

□

Theorem B.3 Closure and Closed Subset

Let $\mathcal{M}$ be a subset of a metric space $\mathcal{X}$

The closure $\overline{\mathcal{M}}$ is always a closed subset.

Proof

The proof is done by contradiction. If $\overline{\mathcal{M}}$ is not *closed*, then $\overline{\mathcal{M}}^c$ is not *open*. Thus there is a point $\mathbf{x}\in\overline{\mathcal{M}}^c$, such that every open ball $\mathcal{B}_{1/n}(\mathbf{x})$, $n\in\mathbb{N}$, has at least one element $\mathbf{x}_n\in\overline{\mathcal{M}}$ (see remark to Definition B.5). Thus, the sequence $\{\mathbf{x}_n\}$ in $\mathcal{M}$ converges to $\mathbf{x}$. This implies that $\mathbf{x}$ is an *accumulation point* of $\mathcal{M}$ (Theorem B.1) and thus belongs to $\overline{\mathcal{M}}$ by definition. This contradicts the $\mathbf{x}\in\overline{\mathcal{M}}^c$ and therefore $\overline{\mathcal{M}}$ is closed.

□

Theorem B.4 Closed and Complete Subset

Let $\mathcal{M}$ be a subset of a complete metric space $\mathcal{X}$.

Then $\mathcal{M}$ is complete, if and only if $\mathcal{M}$ is closed.

Remark

For the subset $\mathcal{M}$ of a general metric space $\mathcal{X}$, the property of being complete is not equivalent to the property of being closed. Nevertheless, a *complete* subset is *closed*, but not vice versa.

In a closed subset $\mathcal{M}$, every convergent sequence with elements in $\mathcal{M}$ converges to some $\mathbf{x}\in\mathcal{M}$ (Theorem B.2). However even in a closed subset $\mathcal{M}$ there could be a *Cauchy sequence*, which does not even converge in $\mathcal{X}$ (hence it is not convergent), if $\mathcal{X}$ is not complete. Consequently, this sequence does not

converge in $\mathcal{M}$, too. *Completeness* requires that every Cauchy sequence in $\mathcal{M}$ converges in $\mathcal{M}$. Thus, $\mathcal{M}$ is not complete.

Proof

Since $\mathcal{X}$ is complete, a sequence is convergent if and only if it is a Cauchy sequence (Definition B.20), i.e. Cauchy and convergent sequence in $\mathcal{X}$ are equivalent terms.

NECESSARY PART: *If $\mathcal{M}$ is closed then $\mathcal{M}$ is complete.*

If $\mathcal{M}$ is not complete, there is a Cauchy sequence in $\mathcal{M}$ that converges to some $\mathbf{x} \in \mathcal{X}$ which does not lie in $\mathcal{M}$ (Definition B.20). This however contradicts the assumption that $\mathcal{M}$ is closed (Theorem B.2). Hence $\mathcal{M}$ is complete.

SUFFICIENT PART: *If $\mathcal{M}$ is complete, then $\mathcal{M}$ is closed.*

Since $\mathcal{M}$ is compete, every Cauchy sequence in $\mathcal{M}$ converges in $\mathcal{M}$ (Definition B.20). Since Cauchy and convergent sequences are equivalent terms in $\mathcal{X}$, every convergent sequence with elements in $\mathcal{M}$, converges in $\mathcal{M}$. Therefore, $\mathcal{M}$ is closed.

□

Theorem B.5 Convergence of Subsequences

Let $\mathcal{X}$ be a metric space.

Further let $\{\mathbf{x}_n\}$ be a sequence that converges to some $\mathbf{x} \in \mathcal{X}$

Then also every subsequence converges to $\mathbf{x}$.

Proof

Since $\{\mathbf{x}_n\}$ is convergent, for every $\varepsilon > 0$ there is some $N \in \mathbb{N}$ such that $d(\mathbf{x}_n, \mathbf{x}) < \varepsilon$ for all $n \geq N$. Hence for every subsequence $\{\mathbf{x}_m\}$, there is some $M \in \mathbb{N}$ such that $d(\mathbf{x}_m, \mathbf{x}) < \varepsilon$ for all $m \geq M$ and the subsequence also converges to $\mathbf{x}$.

□

Theorem B.6 Equivalence of Compactness and Closed / Boundedness

Let $\mathcal{M}$ be a subset of a finite dimensional Banach space $\mathcal{X}$.

$\mathcal{M}$ is *compact* if and only if it is *closed* and *bounded.*

Remark

The necessary part: "If $\mathcal{M}$ is compact, then it is closed and bounded" is also true for the wider class of metric spaces, while the sufficient part needs the

properties of a finite dimensional Banach space such as real vector spaces $\mathbb{R}^n$. Contrary, the normed (metric) space of time domain signals with convergent $\mathcal{L}_1$-norm has an infinite dimension.

Proof

NECESSARY PART: If $\mathcal{M}$ is compact, then it is closed and bounded.

$\mathcal{M}$ IS CLOSED:

For every $\mathbf{x} \in \overline{\mathcal{M}}$ there is a sequence $\{\mathbf{x}_n\}$ in $\mathcal{M}$ which converges to $\mathbf{x}$ (Definition B.21, DefinitionB.22). Hence, also every subsequence $\{\mathbf{x}_m\}$ converges to $\mathbf{x}$ (Theorem B.5). But, as $\mathcal{M}$ is compact, $\mathbf{x} \in \mathcal{M}$ (Definition B.24). Since the argument holds for any $\mathbf{x} \in \overline{\mathcal{M}}$, we have $\mathcal{M} = \overline{\mathcal{M}}$ and thus $\mathcal{M}$ is closed (Theorem B.3).

$\mathcal{M}$ IS BOUNDED:

The proof is done by contradiction. If $\mathcal{M}$ is unbounded, then there is a sequence $\{\mathbf{x}_n\}$, such that $d(\mathbf{x}_{n+1}, \mathbf{x}_n) \geq n$ and there is no subsequence that converges in $\mathcal{M}$. This contradicts the assumption of $\mathcal{M}$ being compact (Definition B.24), thus $\mathcal{M}$ is bounded.

So far, we only employed properties of a general metric space. The following part however needs the properties of a finite dimensional Banach space.

SUFFICIENT PART: If $\mathcal{M}$ is closed and bounded, then it is compact.

Since $\mathcal{M}$ is a closed subset of a complete space, it is also complete (Theorem B.4). Moreover, since $\mathcal{M}$ is a bounded subset of a finite dimensional Banach space, it is also totally bounded (remark to Definition B.13).

For compactness, it has to be shown, that every sequence $\{\mathbf{x}_n\}$ in $\mathcal{M}$ has a subsequence that converges in $\mathcal{M}$. To this end notice that, since $\mathcal{M}$ is totally bounded, there is a finite collection of open unit balls $\mathcal{B}_1(\mathbf{y}_m^{(1)})$, $\mathbf{y}_m^{(1)} \in \mathcal{X}$, $m = 1, \ldots, M_1$ which completely covers $\mathcal{M}$. Since $\{\mathbf{x}_n\}$ has an infinite number of elements and there is only a finite number of unit balls, consequently there is at least one $\mathcal{B}_1(\mathbf{y}_m)$, let us say $\mathcal{B}_1(\mathbf{y}_{\mu_1}^{(1)})$, that contains an infinite number of $\{\mathbf{x}_n\}$ and hence there is a subsequence $\{\mathbf{x}_{n1}\}$ whose elements lie in $\mathcal{B}_1(\mathbf{y}_{\mu_1}^{(1)}) \cap \mathcal{M}$.

Analogously, there is finite collection of $\mathcal{B}_{1/2}(\mathbf{y}_m^{(2)})$, $\mathbf{y}_m^{(2)} \in \mathcal{X}$ $m = 1, \ldots, M_2$, that covers $\mathcal{M}$. Furthermore, a sub collection of it pairwise intersects with and completely covers

$\mathcal{B}_1(\mu_1)\cap\mathcal{M}$. The sub collection of $\mathcal{B}_{1/2}\left(\mathbf{y}_m^{(2)}\right)$ is finite and there is an infinite number of elements of $\{\mathbf{x}_{n1}\}$ in $\mathcal{B}_1(\mu_1)\cap\mathcal{M}$. For this reason, there is at least one $\mathcal{B}_{1/2}\left(\mathbf{y}_{\mu_2}^{(2)}\right)$ that contains an infinite number of elements of $\{\mathbf{x}_{n1}\}$ and hence there is a subsequence $\{\mathbf{x}_{n2}\}$ whose elements lie in $\mathcal{B}_{1/2}\left(\mathbf{y}_{\mu_2}^{(2)}\right)\cap\mathcal{B}_1\left(\mathbf{y}_{\mu_1}^{(1)}\right)\cap\mathcal{M}$. This procedure is repeated infinite times and for the k^{th} iteration step, we arrive at a subsequence $\{\mathbf{x}_{nk}\}$, whose elements lie in $\mathcal{B}_{1/k}\left(\mathbf{y}_{\mu_k}^{(k)}\right)\cap\ldots\cap\mathcal{B}_1\left(\mathbf{y}_{\mu_1}^{(1)}\right)\cap M$.

Now, out of every subsequence, we chose an element $\mathbf{x}_{\nu_K}$ such that $\nu_{k+1}>\nu_k$ and we obtain a sequence$\{\mathbf{x}_{\nu k}\}$. This sequence is a subsequence of $\{\mathbf{x}_n\}$ and a Cauchy sequence since, for every $K\in\mathbb{N}$, all $\mathbf{x}_{\nu k}$ lie within $\mathcal{B}_{1/K}\left(\mathbf{y}_{\mu_K}^{(K)}\right)$ for all $k\geq K$ (Definition B.19). Since the Banach space $\mathcal{X}$ is complete and $\mathcal{M}$ is closed, $\mathcal{M}$ it is also complete (Theorem B.4) and hence the Cauchy sequence converges to some element in $\mathcal{M}$ (Definition B.20).

Summing up, to an arbitrary sequence, whose elements are in $\mathcal{M}$, we have constructed a subsequence that converges in $\mathcal{M}$, hence $\mathcal{M}$ is compact (Definition B.24).

□

Figure B.2 gives an overview for the relationships between the terms “complete”, “closed” and “compact” of subsets that reside in different spaces. There are two interesting results of the figure:

1. In a complete metric space “closed” and “complete” are equivalent properties for subsets
2. In a finite dimensional Banach space, “compact” and “closed and bounded” are equivalent properties for subsets.

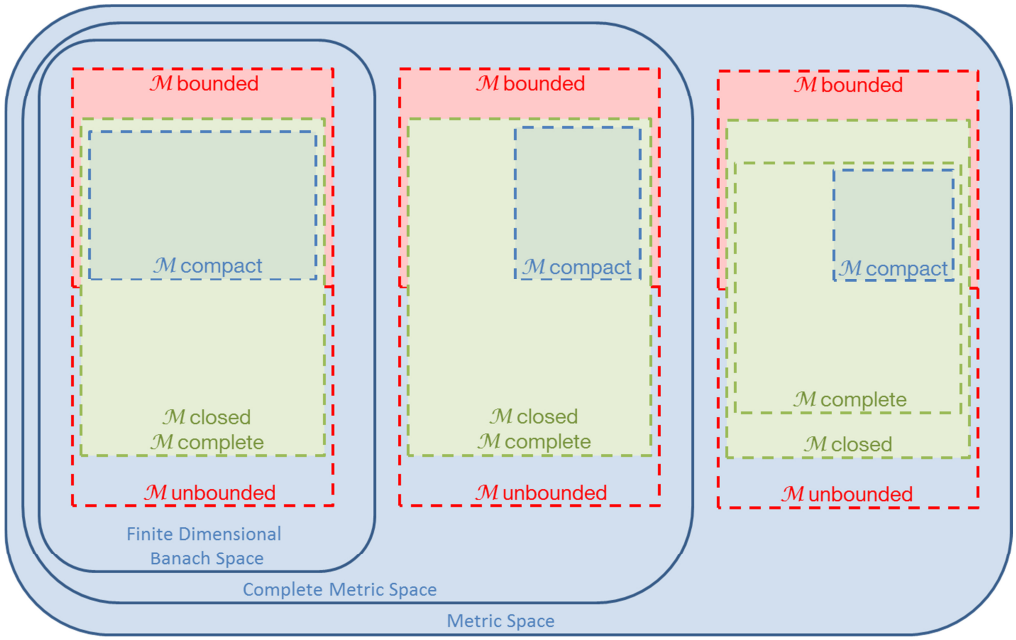


Figure B.2 Relationships between the Terms "Complete", "Closed", "Compact" and "Bounded"

So in the well-known Euclidean vector space $\mathbb{R}^n$, the terms "compact" and "closed and bounded" can be used equivalently. This is not necessarily true for Banach spaces, whose elements are real valued functions (as e.g. the $\mathcal{L}_p$ spaces). These are infinite dimensional and thus a subset is possibly *closed and bounded* but not *compact*.

The next theorem is useful for the proof of Lyapunov's asymptotic stability and is restricted to the Banach space of real numbers $\mathbb{R}$ whose norm is given by the absolute value.

Theorem B.7 Convergence of Monotone Sequence

Let $\{x_n\}$ be some sequence in $\mathbb{R}$

Let further $x_{n+1} \le x_n$ and $x_n \ge c$, $n \in \mathbb{N}$ and some $c \in \mathbb{R}$

Then $\{x_n\}$ converges to some $x_0 \ge c$, i.e.

$$\lim_{n\to\infty}(x_n) = x_0$$

Proof

The proof is done by contradiction. Assume that $\{x_n\}$ does not converge to x_0. Then $\{x_n\}$ is not a Cauchy sequence by Definition B.20 since $\mathbb{R}$ is complete. Moreover, there exists some $\varepsilon > 0$ such that, for every $N \in \mathbb{N}$, there are some integers n_N and m_N such that

$$\left| x_{m_N} - x_{n_N} \right| \ge \varepsilon$$

Assume without loss of generality that $m_N > n_N$. Then , since $\{x_n\}$ is monotonically decreasing, we have

$$x_{m_N} \leq x_{n_N} - \varepsilon .$$

From these considerations, a subsequence is constructed as described in the following. For the first pair of the sequence, we choose $N_1=1$ and obtain some n_{N1} and m_{N1} such that $x_{m_{N1}} \leq x_{n_{N1}} - \varepsilon$. In the second iteration step, N_2 is chosen such that $n_{N2} > m_{N1}$ which implies that $x_{n_{N2}} \leq x_{m_{N1}}$ since the $\{x_n\}$ monotonically decreasing. Consequently, for the k^{th} iteration step, we chose N_k such that $n_{Nk}>m_{Nk-1}$ and obtain the k^{th} pair of sequence elements such that $x_{m_{Nk}} \leq x_{n_{Nk}} - \varepsilon$ and $x_{n_{Nk}} \leq x_{m_{Nk-1}}$. Summing up, the subsequence reads as

$$\{y_k\}=\{x_{n_{N1}}, x_{m_{N1}}, x_{n_{N2}}, x_{m_{N2}}, \ldots\}, k \in \mathbb{N},$$

its elements are monotonically decreasing and fulfill the inequality $y_{k+2} \leq y_k - \varepsilon$, which implies $\lim_{k\to\infty}(y_k) = -\infty$ which, in turn, contradicts the assumption that $\{x_n\}$ is bounded from below by some c. Therefore $\{x_n\}$ converges to x_0.

□

B.1.4 Functions and Continuity

Definitions

Definition B.25 Function

Given some sets $\mathcal{X}$, $\mathcal{Y}$ and a subset $\mathcal{A} \subset \mathcal{X}$.

Then a *function* $T: \mathcal{A} \to \mathcal{Y}$ allocates to every $\mathbf{x} \in \mathcal{A}$ an element $\mathbf{y} \in \mathcal{Y}$:

$$\mathbf{y} = T(\mathbf{x})$$

Definition B.26 Pointwise Continuous Function

Given some metric spaces $\mathcal{X}$ and $\mathcal{Y}$ with metrics $d_x(\cdot,\cdot)$ and $d_y(\cdot,\cdot)$ and a subset $\mathcal{A} \subset \mathcal{X}$.

Then a function $f: \mathcal{A} \to \mathcal{Y}$ is *pointwise continuous* on $\mathcal{A}$,

if for every $\mathbf{x}_0 \in \mathcal{A}$ and every $\varepsilon > 0$ there is a $\delta(\varepsilon, \mathbf{x}_0) > 0$

such that $d_y(\mathbf{y}, \mathbf{y}_0) < \varepsilon$ if $d_x(\mathbf{x}, \mathbf{x}_0) < \delta$

where $\mathbf{x} \in \mathcal{X}$, $\mathbf{y} = f(\mathbf{x})$ and $\mathbf{y}_0 = f(\mathbf{x}_0)$

Definition B.27 Uniform Continuous Function

Given some metric spaces $\mathcal{X}$ and $\mathcal{Y}$ with metrics $d_x(\cdot,\cdot)$ and $d_y(\cdot,\cdot)$ and a subset $\mathcal{A} \subset \mathcal{X}$.

Then a function $f:\mathcal{A}\to\mathcal{Y}$ is *uniformly continuous* on $\mathcal{A}$,

if for every $\varepsilon > 0$ there is a $\delta(\varepsilon)>0$

such that $d_y(\mathbf{y},\mathbf{y}_0)<\varepsilon$ if $d_x(\mathbf{x},\mathbf{x}_0)<\delta$

where $\mathbf{x},\mathbf{x}_0 \in \mathcal{A}$, $\mathbf{y}=f(\mathbf{x})$ and $\mathbf{y}_0=f(\mathbf{x}_0)$

Remarks

- While in pointwise continuity, every δ that has to be found for some ε might be different for different $\mathbf{x}_0$, in uniform continuity, δ is equal for all $\mathbf{x}_0 \in \mathcal{A}$.
- *Uniformly continuous* is a stronger condition than *pointwise continuous*, i.e. every uniformly continuous function is pointwise continuous but not vice versa.

Theorems

Theorem B.8 Continuity and Compact Set

Let $(\mathcal{X},d_x)$ and $(\mathcal{Y},d_y)$ be metric spaces, $\mathcal{A} \subset \mathcal{X}$ a compact subspace.

Let there be further a pointwise continuous function $\mathbf{f}:\mathcal{A}\to\mathcal{Y}$.

Then $\mathcal{B}=\mathbf{f}(\mathcal{A})$ is also compact.

Remark

If $\mathcal{X}=\mathbb{R}^n$ and $\mathcal{Y}=\mathbb{R}^m$ with some respective p-norm defined on it, he theorem implies that a function $\mathbf{f}$ that is continuous on a compact set (which is equivalent to being closed and bounded - see Theorem B.6, e.g. a closed ball $\overline{\mathcal{B}_r}(\mathbf{0})$ of radius $r>0$) has an image that is closed and particularly bounded, i.e. $\|\mathbf{f}(\mathbf{x})\| \le M$ on $\overline{\mathcal{B}_r}$ for some $M \ge 0$. For scalar real valued continuous functions, this particularly excludes the case that it has a vertical asymptote at the boundaries of a compact set.

Proof

For every sequence $\{\mathbf{y}_n\}$ in $\mathcal{B}$ there is a sequence $\{\mathbf{x}_n\}$ in $\mathcal{A}$, such that $\mathbf{y}_n = f(\mathbf{x}_n)$. Since $\mathcal{A}$ is compact, there is a sub-sequence $\{\mathbf{x}_m\}$, which converges to some $\mathbf{x} \in \mathcal{A}$ (Definition B.24). Thus for every $\delta > 0$, there is an $M(\delta) \in \mathbb{N}$ such that $\|\mathbf{x}_m - \mathbf{x}\| < \delta$ for all $m \geq M$. Further, we define $\{\mathbf{y}_m\} = \{f(\mathbf{x}_m)\}$, in fact a subsequence of $\{\mathbf{y}_n\}$ and $\mathbf{y} = f(\mathbf{x})$, the image of the limit point. Since f is continuous, for every $\varepsilon > 0$ there is a $\delta(\varepsilon, \mathbf{x})$ such that $\|\mathbf{y}_m - \mathbf{y}\| < \varepsilon$ for all $m \geq M(\delta(\varepsilon, \mathbf{x}))$. Thus, every sequence in $\mathcal{B}$ has a subsequence, which converges in B, and thus $\mathcal{B}$ is compact.

□

Theorem B.9 Uniform Continuity and Compact Set

Let $(\mathcal{X}, d_x)$ and $(\mathcal{Y}, d_y)$ some metric spaces,

$\mathcal{A} \subseteq \mathcal{X}$ a compact subspace.

Every function $f : \mathcal{A} \to \mathcal{Y}$, that is continuous, is uniformly continuous.

Proof

Assume that f is not uniformly continuous. Then, for some $\varepsilon_0 > 0$ and every $\delta > 0$, there exists a pair $\alpha, \beta \in \mathcal{A}$ with $d_x(\alpha, \beta) < \delta$ such that $d_y(f(\alpha), f(\beta)) \geq \varepsilon_0$ (refer to Definition B.27). Therefore we can construct sequences $\{\alpha_n\}$ and $\{\beta_n\}$ in $\mathcal{A}$, such that

$$d_x(\alpha_n, \beta_n) < 1/n \ , \ d_y(f(\alpha_n), f(\beta_n)) \geq \varepsilon_0 \text{ for all } n \in \mathbb{N}.$$

Since $\mathcal{A}$ is compact, there is a subsequence $\{\alpha_{ni}\}$, $i \in \mathbb{N}$ which converges to some $c \in \mathcal{A}$. It follows that the analog subsequence $\{\beta_{ni}\}$ converges to c, too. To this end notice that, for every $\delta > 0$, there exists some

$$I_1(\delta) \in \mathbb{N} \text{ such that } d_x(\alpha_{ni}, c) < \delta/2 \text{ for } i \geq I_1(\delta).$$

Moreover, there is some

$$I_2(\delta) \in \mathbb{N} \text{ such that } n_{I_2}^{-1} < \delta/2 \text{ which implies that } d_x(\beta_{ni}, \alpha_{ni}) < \delta/2 \text{ for } i \geq I_2(\delta).$$

From the triangle inequality, we get

$$d_x(\beta_{ni}, c) \leq d_x(\beta_{ni}, \alpha_{ni}) + d_x(\alpha_{ni}, c) < \delta \text{ for all } i \geq I_\beta(\delta) := \max(I_1(\delta), I_2(\delta)). \quad \text{(B.1)}$$

On the other hand, since $\{\alpha_{ni}\}$ converges to c, there exists an $I_\alpha \in \mathbb{N}$ such that

$$d_x(\alpha_{ni}, c) < \delta \text{ for all } i \geq I_\alpha(\delta). \quad \text{(B.2)}$$

Since $\{\alpha_{ni}\}$ and $\{\beta_{ni}\}$ are subsequences of $\{\alpha_n\}$ and $\{\beta_n\}$ respectively, we have

$$d_y(f(\alpha_{ni}), f(\beta_{ni})) \geq \varepsilon_0 \quad \text{(B.3)}$$

for all $i \in \mathbb{N}$. However, since f is pointwise continuous and by equations (B.1) and (B.2), for every $\varepsilon>0$, there are some $\delta(\varepsilon)$, $I_\alpha(\delta)$ and $I_\beta(\delta)$ such that

$$d_y(f(\alpha_{ni}), f(c)) < \varepsilon/2 \quad , \quad d_y(f(\beta_{ni}), f(c)) < \varepsilon/2 \text{ for } i \geq I_0(\delta) := \max\left(I_\alpha(\delta), I_\beta(\delta)\right).$$

By application of triangle inequality, we get

$$d_y(f(\alpha_{ni}), f(\beta_{ni})) < d_y(f(\alpha_{ni}), f(c)) + d_y(f(\beta_{ni}), f(c)) < \varepsilon \text{ for } i \geq I_0(\delta)$$

which contradicts (B.3) if $\varepsilon \leq \varepsilon_0$. Hence, f is uniformly continuous.

□

B.1.5 Limits

This section defines some special limits that are useful in stability analysis of adaptive control systems. The ordinary limit is assumed to be known to the reader, but these special definitions are often not known precisely. The exact knowledge of these definitions, however, helps to understand the stability proves for adaptive systems in detail.

Definition B.28 Limit Superior

Let $\{x_n\}$ be a sequence in $\mathbb{R}$. Then the *limit superior* is defined as

$$\limsup_{n\to\infty}(x_n) = \lim_{n\to\infty}(s_n) \text{ where } s_n = \sup\{x_k \mid k \geq n\}$$

Definition B.29 Limit Inferior

Let $\{x_n\}$ be a sequence in $\mathbb{R}$. Then the *limit inferior* is defined as

$$\liminf_{n\to\infty}(x_n) = \lim_{n\to\infty}(s_n) \text{ where } s_n = \inf\{x_k \mid k \geq n\}$$

Remarks

- In this case $\limsup$ and $\liminf$ are defined for sequences in the real numbers.
- A sequence in the real numbers that has convergent $\limsup$ or $\liminf$, does not necessarily have a convergent $\lim_{n\to\infty}(x_n)$.
- However, $\limsup_{n\to\infty}(x_n)$ and $\lim_{n\to\infty}(x_n)$ converge to the same value c if and only if $\lim_{n\to\infty}(x_n)$ converges to c.

Convergence properties for $\limsup$ *and* $\liminf$

Without loss of generality consider limit superior. Cleary, the sequence $\{s_n\}$ is monotonically decreasing. Hence, there are only two possibilities for its behavior for large n. Either $\{s_n\}$ is bounded from below by some constant c and, by Theorem B.7, converges

$$\lim_{n\to\infty}(s_n) = c(s_n)\ \lim_{n\to\infty}(s_n) = c$$

or it strives to $-\infty$ if it is not bounded from below. To this end, assume that $c = \inf(s_n)$, then for every $\varepsilon > 0$ there is some $N_0 \in \mathbb{N}$ such that $s_{N_0} < c + \varepsilon$. Since $\{s_n\}$ is monotonically decreasing, also $s_N < c + \varepsilon$ for all $N \geq N_0$, which shows convergence of $\{s_n\}$ to c. If it is not bounded from below, then for every $d \in \mathbb{R}$ there is some M_0 such that $s_{M_0} < d$ and since $\{s_n\}$ is monotonically decreasing $s_M < d$ for all $M \geq M_0$ and hence $\lim_{n\to\infty}(s_n) = -\infty$. Intuitively, the reason for restrictive behavior of monotonic sequences is the fact that oscillating sequences such as $x_n = \sin(\pi \cdot n)$ are excluded.

Example

Consider the sequence: $x_n = \left(1 + e^{-\frac{n}{500}}\right)\sin(0.026 \cdot \pi \cdot n)$

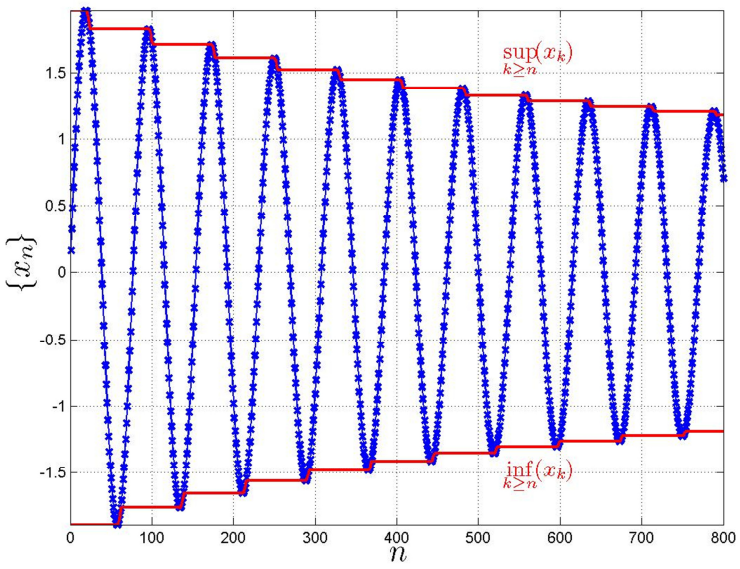


Figure B.3 Example limit superior and limit inferior

Figure B.3 shows that the sequence has no limit, as $n \to \infty$. However $\limsup$ and $\liminf$ have a limit very well, namely

- $\limsup_{n\to\infty}(x_n)=1$
- $\liminf_{n\to\infty}(x_n)=-1$

Dini's Derivatives

An application of limit superior and limit inferior is given in a generalized definition for derivatives. For scalar real valued functions $f(t)$, it is well known that the derivative is defined as the limit of the difference quotient

$$\dot{f}(t)=\lim_{h\to 0}\left(\frac{f(t+h)-f(t)}{h}\right)$$

Thereby h can go to zero from the left AND the right side and the limit exist, only if it also exists for every sequence $\{h_n\}$ that fulfills $\lim_{n\to\infty}(h_n)=0$. In some cases it might occur that the derivative as commonly defined does not exist but, if h is restricted to be strictly positive (or strictly negative) the limit still could exist. A prominent example is the absolute value function $f(t)=|t|$. At point $t=0$, the derivative in the conventional sense does not exist, but if one restricts h to strictly positive (strictly negative) values, the derivative evaluates to $+1$ or -1 respectively. For such cases, *Dini's derivatives* [McS47] generalize the common definition of derivative.

Definition B.30 Dini's Derivatives

Let $f(t)$ be defined on an interval $[\alpha,\beta]$ and t_0 some point within the interval. Then

- The upper derivative of $f(t)$ at t_0 is $\overline{D}f(t_0)=\limsup_{h\to 0}\frac{f(t_0+h)-f(t_0)}{h}$
- The lower derivative of $f(t)$ at t_0 is $\underline{D}f(t_0)=\liminf_{h\to 0}\frac{f(t_0+h)-f(t_0)}{h}$
- The upper right derivative of $f(t)$ at t_0 is $D^{+}f(t_0)=\limsup_{h\to 0^{+}}\frac{f(t_0+h)-f(t_0)}{h}$
- The lower right derivative of $f(t)$ at t_0 is $D_{+}f(t_0)=\liminf_{h\to 0^{+}}\frac{f(t_0+h)-f(t_0)}{h}$
- The upper left derivative of $f(t)$ at t_0 is $D^{-}f(t_0)=\limsup_{h\to 0^{-}}\frac{f(t_0+h)-f(t_0)}{h}$
- The lower left derivative of $f(t)$ at t_0 is $D_{-}f(t_0)=\liminf_{h\to 0^{-}}\frac{f(t_0+h)-f(t_0)}{h}$

B.2 Lie Derivative and Lie Product

The section presents an introduction to Lie derivative and Lie product, which are important within the derivation of the nonlinear dynamic inversion (NDI) algorithm. For more details, the reader is referred to [Isi95] and [Had08].

Definition B.31 Lie Derivative [Isi95]

Let $\mathcal{U}$ be an open subspace of $\mathbb{R}^n$, a real valued function $h:\mathcal{U}\to\mathbb{R}$ and a real valued vector field $\mathbf{f}:\mathcal{U}\to\mathbb{R}^n$.

The Lie derivative of h along $\mathbf{f}$ at some $\mathbf{x}\in\mathbb{U}$ is defined as:

$$L_{\mathbf{f}}h(\mathbf{x})=\frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}\cdot\mathbf{f}(\mathbf{x})$$

Remarks:

1. Note that $\partial h/\partial \mathbf{x}$ is the Jacobi matrix of a scalar function and thus a row vector
2. Iterated application of the Lie derivative is defined as

$$L_{\mathbf{f}}^{k}h(\mathbf{x})=L_{\mathbf{f}}\left(L_{\mathbf{f}}^{k-1}(\mathbf{x})\right)=\frac{\partial L_{\mathbf{f}}^{k-1}(\mathbf{x})}{\partial \mathbf{x}}\cdot\mathbf{f}(\mathbf{x}) \qquad \text{(B.4)}$$

with $L_{\mathbf{f}}^{0}h(\mathbf{x})=h(\mathbf{x})$

3. If the vector field is replaced by a matrix $\mathbf{G}(\mathbf{x})=\left[\mathbf{g}_1(\mathbf{x})\ \cdots\ \mathbf{g}_n(\mathbf{x})\right]$, where each columns $\mathbf{g}_i:\mathcal{U}\to\mathbb{R}^n$ is a smooth vector field, the Lie derivative is defined accordingly:

$$L_{\mathbf{G}}h=\left[L_{\mathbf{g}_1}h\ \cdots\ L_{\mathbf{g}_n}h\right] \qquad \text{(B.5)}$$

Definition B.32 Lie Product [Isi95]

Let $\mathcal{U}$ be an open subspace of $\mathbb{R}^n$

and two real valued vector fields $\mathbf{f}:\mathcal{U}\to\mathbb{R}^n$, $\mathbf{g}:\mathcal{U}\to\mathbb{R}^n$.

The Lie bracket at some $\mathbf{x}\in\mathcal{U}$ is defined as

$$[\mathbf{f},\mathbf{g}](\mathbf{x})=\frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x})-\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}}\mathbf{g}(\mathbf{x})$$

Remark: Repeated application of the Lie product is defined recursively

$$\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}(\mathbf{x})=\left[\mathbf{f}\ \ \mathrm{ad}_{\mathbf{f}}^{k-1}\mathbf{g}\right](\mathbf{x}) \qquad \text{(B.6)}$$

with: $\mathrm{ad}_{\mathbf{f}}^{0}\mathbf{g}(\mathbf{x})=\mathbf{g}(\mathbf{x})$

The Lie product has some basic properties, which appear to be useful and are easy to proof [Isi95].

Basic Properties

1. bilinear over $\mathbb{R}$, i.e. if $\mathbf{f_1}$, $\mathbf{f_2}$, $\mathbf{g_1}$, $\mathbf{g_2}$ are vector fields and r_1, r_2 are real numbers, then

$$\begin{aligned}[r_1\mathbf{f}_1 + r_2\mathbf{f}_2, \mathbf{g}_1] &= r_1[\mathbf{f}_1, \mathbf{g}_1] + r_2[\mathbf{f}_2, \mathbf{g}_1] \\ [\mathbf{f}_1, r_1\mathbf{g}_1 + r_2\mathbf{g}_2] &= r_1[\mathbf{f}_1, \mathbf{g}_1] + r_2[\mathbf{f}_1, \mathbf{g}_2]\end{aligned} \tag{B.7}$$

2. skew commutative, i.e.

$$[\mathbf{f}, \mathbf{g}](\mathbf{x}) = -[\mathbf{g}, \mathbf{f}](\mathbf{x}) \tag{B.8}$$

3. satisfies the Jacobi identity, i.e. if $\mathbf{f}$, $\mathbf{g}$, $\mathbf{p}$ are vector fields, then

$$[\mathbf{f}, [\mathbf{g}, \mathbf{p}]] + [\mathbf{p}, [\mathbf{f}, \mathbf{g}]] + [\mathbf{g}, [\mathbf{p}, \mathbf{f}]] = \mathbf{0} \tag{B.9}$$

4. Let α, β be some scalar real valued functions and $\mathbf{f}, \mathbf{g}$ smooth vector fields, then

$$[\alpha\mathbf{f}, \beta\mathbf{g}] = \alpha\beta[\mathbf{f}, \mathbf{g}] + (L_{\mathbf{f}}\beta)\cdot\alpha\cdot\mathbf{g} - (L_{\mathbf{g}}\alpha)\cdot\beta\cdot\mathbf{f} \tag{B.10}$$

5. Let $\mathbf{f}$, $\mathbf{g}$, be vector fields and h a real valued function, then

$$L_{[\mathbf{f},\mathbf{g}]}h = L_{\mathbf{g}}L_{\mathbf{f}}h - L_{\mathbf{f}}L_{\mathbf{g}}h \tag{B.11}$$

Proof

Expanding the left hand side (B.11) of yields

$$L_{\mathbf{g}}L_{\mathbf{f}}h - L_{\mathbf{f}}L_{\mathbf{g}}h = \frac{\partial}{\partial\mathbf{x}}\left(\frac{\partial h}{\partial\mathbf{x}}\cdot\mathbf{f}\right)\cdot\mathbf{g} - \frac{\partial}{\partial\mathbf{x}}\left(\frac{\partial h}{\partial\mathbf{x}}\cdot\mathbf{g}\right)\cdot\mathbf{f}$$

$$L_{\mathbf{g}}L_{\mathbf{f}}h - L_{\mathbf{f}}L_{\mathbf{g}}h = \mathbf{f}^T\cdot\frac{\partial^2 h}{\partial\mathbf{x}^2}\cdot\mathbf{g} + \frac{\partial h}{\partial\mathbf{x}}\cdot\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\cdot\mathbf{g} - \mathbf{g}^T\cdot\frac{\partial^2 h}{\partial\mathbf{x}^2}\cdot\mathbf{f} - \frac{\partial h}{\partial\mathbf{x}}\cdot\frac{\partial\mathbf{g}}{\partial\mathbf{x}}\cdot\mathbf{f}$$

Note that $\partial^2\mathbf{h}/\partial\mathbf{x}^2$ is the Hessian matrix and therefore symmetric. Hence the first and third summand cancel each other and

$$L_{\mathbf{g}}L_{\mathbf{f}}h - L_{\mathbf{f}}L_{\mathbf{g}}h = \frac{\partial h}{\partial\mathbf{x}}\cdot\left(\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\cdot\mathbf{g} - \frac{\partial\mathbf{g}}{\partial\mathbf{x}}\cdot\mathbf{f}\right) = L_{[\mathbf{f},\mathbf{g}]}h\,.$$

□

Identity (B.11) can further be extended to the following expression.

$$L_{\mathrm{ad}_{\mathbf{f}}^k\mathbf{g}}h = L_{\mathrm{ad}_{\mathbf{f}}^k\mathbf{g}}L_{\mathbf{f}}h - L_{\mathbf{f}}L_{\mathrm{ad}_{\mathbf{f}}^k\mathbf{g}}h \tag{B.12}$$

The proof follows along the same lines as for (B.11). Successive application of (B.12) yields the following identity.

Lemma B.1 Lie Identity 1 [Isi95]

Let $\mathcal{U}$ be an open subset of $\mathbb{R}^n$, and let $h:\mathcal{U}\to\mathbb{R}$ be a real valued function and $\mathbf{f}:\mathcal{U}\to\mathbb{R}^n$, $\mathbf{g}:\mathcal{U}\to\mathbb{R}^n$ vector fields. Let further be integers $k,r\geq 0$. Then

$$L_{\mathrm{ad}_{\mathbf{f}}^{k+r}\mathbf{g}}h=\sum_{i=0}^{r}(-1)^i\cdot\binom{n}{i}\cdot L_{\mathbf{f}}^{n-i}L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{i}h$$

Proof:

Applying (B.12) to the right hand side of Lemma B.1 successively yields (note that the coefficients comply with Pascal's triangle):

$$\boxed{1}\cdot L_{ad_{\mathbf{f}}^{k+r}\mathbf{g}}h=$$

$$=\boxed{1}\cdot L_{\mathbf{f}}^{0}L_{ad_{\mathbf{f}}^{k+r-1}\mathbf{g}}L_{\mathbf{f}}^{1}h-\boxed{1}\cdot L_{\mathbf{f}}^{1}L_{ad_{\mathbf{f}}^{k+r-1}\mathbf{g}}L_{\mathbf{f}}^{0}h$$

$$=\boxed{1}\cdot L_{\mathbf{f}}^{0}L_{ad_{\mathbf{f}}^{k+r-2}\mathbf{g}}L_{\mathbf{f}}^{2}h-\boxed{2}\cdot L_{\mathbf{f}}^{1}L_{ad_{\mathbf{f}}^{k+r-2}\mathbf{g}}L_{\mathbf{f}}^{1}h+\boxed{1}\cdot L_{\mathbf{f}}^{2}L_{ad_{\mathbf{f}}^{k+r-2}\mathbf{g}}L_{\mathbf{f}}^{0}h$$

$$=\boxed{1}\cdot L_{\mathbf{f}}^{0}L_{ad_{\mathbf{f}}^{k+r-3}\mathbf{g}}L_{\mathbf{f}}^{3}h-\boxed{3}\cdot L_{\mathbf{f}}^{1}L_{ad_{\mathbf{f}}^{k+r-3}\mathbf{g}}L_{\mathbf{f}}^{2}h+\boxed{3}\cdot L_{\mathbf{f}}^{2}L_{ad_{\mathbf{f}}^{k+r-3}\mathbf{g}}L_{\mathbf{f}}^{1}h-\boxed{1}\cdot L_{\mathbf{f}}^{3}L_{ad_{\mathbf{f}}^{k+r-3}\mathbf{g}}L_{\mathbf{f}}^{0}h$$

$$\vdots$$

$$=(-1)^0\binom{r}{0}L_{\mathbf{f}}^{0}L_{ad_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{r}h+(-1)^1\binom{r}{1}L_{f}^{1}L_{ad_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{r-1}h+\ldots+(-1)^{r-1}\binom{r}{r-1}L_{f}^{r-1}L_{ad_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{1}h+(-1)^r\binom{r}{r}L_{f}^{r}L_{ad_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{0}h$$

□

The next lemma plays an important role in the proof for the existence of a local transformation to Byrnes- Isidori normal form.

Lemma B.2 Lie Identity 2 [Isi95]

Let $\mathcal{U}$ be an open subset of $\mathbb{R}^n$, and let $h:\mathcal{U}\to\mathbb{R}$ be a smooth function, $\mathbf{f}:\mathcal{U}\to\mathbb{R}^n$, $\mathbf{g}:\mathcal{U}\to\mathbb{R}^n$ smooth vector fields. Let further k,n,r be integers ≥ 0.

Then the conditions

$$L_{\mathbf{g}}L_{\mathbf{f}}^{0}h=L_{\mathbf{g}}L_{\mathbf{f}}^{1}h=L_{\mathbf{g}}L_{\mathbf{f}}^{2}h=\ldots=L_{\mathbf{g}}L_{\mathbf{f}}^{r-2}h=0 \text{ and } L_{\mathbf{g}}L_{\mathbf{f}}^{r-1}\neq 0 \text{ imply}$$

1. $L_{\mathrm{ad}_{\mathbf{f}}^{0}\mathbf{g}}h=L_{\mathrm{ad}_{\mathbf{f}}^{1}\mathbf{g}}h=L_{\mathrm{ad}_{\mathbf{f}}^{2}\mathbf{g}}h=\ldots L_{\mathrm{ad}_{\mathbf{f}}^{r-2}\mathbf{g}}h=0$ and $L_{\mathrm{ad}_{\mathbf{f}}^{r-1}\mathbf{g}}h=(-1)^{r-1}L_{\mathbf{g}}L_{\mathbf{f}}^{r-1}h\neq 0$
2. $L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{n}h=0$ for $k+n<r-1$ and $L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{n}h=L_{\mathrm{ad}_{\mathbf{f}}^{r-1}\mathbf{g}}h\neq 0$ for $k+n=r-1$

Proof:

FIRST PART

The first part of is proved by induction. For $j=0$, the statement $L_{\mathbf{g}}L_{\mathbf{f}}^{0}h=L_{\mathbf{g}}h=0$ implies $L_{ad_{\mathbf{f}}^{0}\mathbf{g}}h=L_{\mathbf{g}}h=0$ is obviously true. Now assume that the statement is true for some $j<r-2$, it can be shown that $L_{\mathrm{ad}_{\mathbf{f}}^{j+1}\mathbf{g}}h=0$. To that end, employ Lemma B.2 and set $k=0$.

$$L_{\mathrm{ad}_{\mathbf{f}}^{j+1}\mathbf{g}}h=(-1)^{0}\binom{j+1}{0}L_{\mathbf{f}}^{j+1}\underbrace{L_{\mathbf{g}}L_{\mathbf{f}}^{0}h}_{=0}+(-1)^{1}\binom{j+1}{1}L_{\mathbf{f}}^{j}\underbrace{L_{\mathbf{g}}L_{\mathbf{f}}^{1}h}_{=0}+\ldots+(-1)^{j+1}\binom{j+1}{j+1}L_{\mathbf{f}}^{0}\underbrace{L_{\mathbf{g}}L_{\mathbf{f}}^{j+1}h}_{=0}$$

Since, by assumption $L_{\mathbf{g}}L_{\mathbf{f}}^{0}h=\ldots=L_{\mathbf{g}}L_{\mathbf{f}}^{j+1}h=0$, clearly also $L_{\mathrm{ad}_{\mathbf{f}}^{j+1}\mathbf{g}}h=0$. If however $j=r-2$, we get

$$L_{\mathrm{ad}_{\mathbf{f}}^{r-1}\mathbf{g}}h=(-1)^{0}\binom{r-1}{0}L_{\mathbf{f}}^{r-1}\underbrace{L_{\mathbf{g}}L_{\mathbf{f}}^{0}h}_{=0}+(-1)^{1}\binom{r-1}{1}L_{\mathbf{f}}^{r-2}\underbrace{L_{\mathbf{g}}L_{\mathbf{f}}^{1}h}_{=0}+\ldots+(-1)^{r-1}\binom{r-1}{r-1}\underbrace{L_{\mathbf{g}}L_{\mathbf{f}}^{r-1}h}_{\neq 0}$$

which proves the first part.

SECOND PART

For the second part, again Lemma B.1 is employed, while n is set to 1 and k is chosen such that $k+1<r-1$.

$$\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k+1}\mathbf{g}}h}_{=0\ (\text{part1})}=L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}h-L_{\mathbf{f}}\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}h}_{=0\ (\text{part1})}$$

Therefore one can conclude that $L_{\mathrm{ad}_{\mathbf{f}}^{k}}L_{\mathbf{f}}h=0$, since it was shown in the first part that $L_{\mathrm{ad}_{\mathbf{f}}^{k+1}\mathbf{g}}h=L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}h=0$. Applying Lemma B.1 with $n=2$ and assuming that $k+2<r-1$, one obtains.

$$\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k+2}\mathbf{g}}h}_{=0\ (\text{part1})}=L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{2}h-3L_{\mathbf{f}}\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}h}_{=0}+L_{\mathbf{f}}^{2}\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}h}_{=0\ (\text{part1})}$$

Again, part 1 assures that $L_{\mathrm{ad}_{\mathbf{f}}^{k+2}\mathbf{g}}h=L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}h=0$, and the first iteration of part 2 assures that $L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}h=0$ from which can be concluded that $L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{2}h=0$. Repeating the procedure successively until n (still under the proposition that $k+n<r-1$) yields

$$\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k+n}\mathbf{g}}h}_{=0\,(\text{part1})}=(-1)^{0}\binom{n}{0}L_{\mathbf{f}}^{0}L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{n}h+(-1)^{1}\binom{n}{1}L_{\mathbf{f}}^{1}\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{n-1}h}_{=0}+\ldots+(-1)^{n-1}\binom{n}{n-1}L_{\mathbf{f}}^{n-1}\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{1}h}_{=0}+(-1)^{n}\binom{n}{n}L_{\mathbf{f}}^{n}\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k}\mathbf{g}}L_{\mathbf{f}}^{0}h}_{=0\ (\text{part1})}$$

Hence it can be concluded that $L_{\mathrm{ad}_{\mathbf{f}}^k \mathbf{g}} L_{\mathbf{f}}^n h = 0$ for $n+k<r-1$. If $n+k=r-1$, we have

$$\underbrace{L_{\mathrm{ad}_{\mathbf{f}}^{k+n} \mathbf{g}} h}_{\substack{\neq 0 \\ (\text{since } k+n=r-1)}} = (-1)^0 \binom{n}{0} L_{\mathbf{f}}^0 L_{\mathrm{ad}_{\mathbf{f}}^k \mathbf{g}} L_{\mathbf{f}}^n h + (-1)^1 \binom{n}{1} L_{\mathbf{f}}^1 \underbrace{L_{\mathrm{ad}_{\mathbf{f}}^k \mathbf{g}} L_{\mathbf{f}}^{n-1} h}_{=0} + \ldots + (-1)^{n-1} \binom{n}{n-1} L_{\mathbf{f}}^{n-1} \underbrace{L_{\mathrm{ad}_{\mathbf{f}}^k \mathbf{g}} L_{\mathbf{f}}^1 h}_{=0} + (-1)^n \binom{n}{n} L_{\mathbf{f}}^n \underbrace{L_{\mathrm{ad}_{\mathbf{f}}^k \mathbf{g}} L_{\mathbf{f}}^0 h}_{=0 \text{ (part 1)}}$$

and obviously $L_{\mathrm{ad}_{\mathbf{f}}^k \mathbf{g}} L_{\mathbf{f}}^n h = L_{\mathrm{ad}_{\mathbf{f}}^{r-1} \mathbf{g}} h = (-1)^{r-1} L_{\mathbf{g}} L_{\mathbf{f}}^{r-1} h \neq 0$ which completes the proof.

□

B.3 Distributions

Definition B.33 Distribution [Isi95]

Let $\mathcal{U}$ be an open subset of $\mathbb{R}^n$, vector fields $\mathbf{f}_i : \mathcal{U} \to \mathbb{R}^n$, $i = 1, \ldots, d$ (composed of column vectors).

A distribution $F(\mathbf{x})$ is defined as an assignment of a subspace of $\mathbb{R}^n$, spanned by the vector fields $\mathbf{f}_i(\mathbf{x})$, to every $\mathbf{x} \in \mathcal{U}$, i.e.

$$F(\mathbf{x}) = \mathrm{span}\{\mathbf{f}_1(\mathbf{x}), \ldots, \mathbf{f}_d(\mathbf{x})\}$$

Remarks:

- At each point $\mathbf{x}_0 \in \mathcal{U}$, a distribution is a subspace of $\mathbb{R}^n$.
- A vector field $\boldsymbol{\tau}(\mathbf{x})$ is said to belong to the distribution F, if $\boldsymbol{\tau}(\mathbf{x}) \in F(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{U}$. This implies basically that $\boldsymbol{\tau}(\mathbf{x})$ can be expressed as linear combination of

$$\mathbf{f}_i(\mathbf{x}) : \ \boldsymbol{\tau}(\mathbf{x}) = \sum_{i=1}^{d} c_i(\mathbf{x}) \mathbf{f}_i(\mathbf{x})$$

  where $c_i(\mathbf{x})$ are real valued scalar functions on $\mathcal{U}$.

- The matrix-function $\mathbf{F} = [\mathbf{f}_1(\mathbf{x}) \ \cdots \ \mathbf{f}_d(\mathbf{x})]$ is associated with the distribution such that the distribution is spanned by the columns of $\mathbf{F}$. Note that bold notation $\mathbf{F}$ is used for the matrix representation while an italic capital letter F is used if the distribution itself is addressed.
- If the vector fields $\mathbf{f}_i(\mathbf{x})$ are smooth (partial derivatives of any order are continuous), the distribution is said to be smooth.
- The *dimension* of a distribution F at some point $\mathbf{x} \in \mathcal{U}$ is defined as the dimension of the subspace of the distribution at $\mathbf{x}$ and is denoted as $\dim(F(x))$.

There are two bivalent operations, defined for distributions

Sum

For two distributions F_1 and F_2, the sum is defined pointwise as the sum of the subspaces of both distributions at some point $\mathbf{x}$, preconditioned both distributions are defined at the considered point

$$F_1(\mathbf{x}) + F_2(\mathbf{x}) = \left\{ \mathbf{z} \in \mathbb{R}^n \;\middle|\; \mathbf{z} = \boldsymbol{\sigma} + \boldsymbol{\tau}\,, \boldsymbol{\sigma} \in F_1(\mathbf{x}), \boldsymbol{\tau} \in F_2(\mathbf{x}) \right\}$$

Intersection

For two distributions F_1 and F_2, the intersection is defined pointwise for each $\mathbf{x}$, preconditioned both distributions are defined at the considered point.

$$F_1(\mathbf{x}) \cap F_2(\mathbf{x}) = \left\{ \mathbf{z} \in \mathbb{R}^n \;\middle|\; \mathbf{z} \in F_1(\mathbf{x}) \wedge \mathbf{z} \in F_2(\mathbf{x}) \right\}$$

Considering its dimension, another property can be assigned to a distribution.

Nonsingular Distribution

A distribution is called *nonsingular*, if $\dim(F(\mathbf{x})) = \text{const}$ for all $\mathbf{x} \in \mathcal{U}$. Otherwise, it is a *distribution of variable dimension.*

Regular point

A point $\mathbf{x}_0 \in \mathcal{U}$ is a *regular point* of the distribution, if there is a neighborhood $\mathcal{U}_0$ of $\mathbf{x}_0$, such that F is nonsingular on $\mathcal{U}_0$. A point, which is not regular, is said to be a point of singularity.

The following definition plays an important role in the derivation of NDI.

Definition B.34 Involutive Distribution [Isi95]

A distribution F is involutive, if the Lie bracket $[\tau_1, \tau_2]$ of any pair of vector fields τ_1 and τ_2, belonging to F, is a vector field which belongs to F:

$$\tau_1 \in F\,, \boldsymbol{\tau}_2 \in F \rightarrow [\boldsymbol{\tau}_1\,, \boldsymbol{\tau}_2] \in F$$

In order to test, whether a distribution is involutive or not, the following lemma is useful

Lemma B.3 Condition for Involutive Distributions [Isi95]

Let a distribution $F(\mathbf{x}) = \text{span}\{\mathbf{f}_1(\mathbf{x}), \ldots, \mathbf{f}_d(\mathbf{x})\}$, $\mathbf{x} \in \mathcal{U} \subset \mathbb{R}^n$.

$F(\mathbf{x})$ is involutive if and only if $[\mathbf{f}_i(\mathbf{x}), \mathbf{f}_j(\mathbf{x})] \in F(\mathbf{x})$ for all $i, j = 1, \ldots, d$

Proof:

NECESSARY PART: *If F is involutive, then* $[\mathbf{f}_i(\mathbf{x}),\mathbf{f}_j(\mathbf{x})]\in F(\mathbf{x})$.

The necessary part is a direct consequence of Definition B.34, since $\mathbf{f}_i(\mathbf{x})$ and $\mathbf{f}_j(\mathbf{x})$ are vector fields in $F(\mathbf{x})$.

SUFFICIENT PART: *If* $[\mathbf{f}_i(\mathbf{x}),\mathbf{f}_j(\mathbf{x})]\in F(\mathbf{x})$*, then F is involutive.*

Suppose vector fields

$$\boldsymbol{\tau}_1(\mathbf{x})=\sum_{i=1}^{d} c_i(\mathbf{x})\mathbf{f}_i(\mathbf{x}) \text{ and } \boldsymbol{\tau}_2(\mathbf{x})=\sum_{i=1}^{d} d_i(\mathbf{x})\mathbf{f}_i(\mathbf{x})$$

where $c_i(\mathbf{x}),d_i(\mathbf{x})$ are real valued scalar functions. With basic property 4 in Appendix B.2, we get (the argument $\mathbf{x}$ is dropped for readability)

$$[\boldsymbol{\tau}_1,\boldsymbol{\tau}_2]=\sum_{j=1}^{d}\sum_{i=1}^{d} c_i\cdot d_j\cdot[\mathbf{f}_i,\mathbf{f}_j]+\left(L_{\mathbf{f}_i}d_j\right)\cdot c_i\cdot\mathbf{f}_j-\left(L_{\mathbf{f}_j}c_i\right)\cdot d_j\cdot\mathbf{f}_i$$

In order for the distribution to be involutive, the right hand side of the above equation has to belong to F. The first term belongs to F by assumption, the second and third terms also belong to F, since it points into the directions $\mathbf{f}_i$ and $\mathbf{f}_j$. Hence, F is involutive.

□

Associated with the vector space $\mathbb{R}^n$ is a so-called *dual space*, which is defined as the set of all linear maps from $\mathbb{R}^n$ to $\mathbb{R}$. It is well-known that linear maps on a vector $\mathbf{x}^T=(x_1\ \cdots\ x_n)$ have the common form $z(\mathbf{x})=y_1x_1+\ldots y_nx_n$, where y_i are some real constants. By defining $\mathbf{y}^T=(y_1\ \cdots\ y_n)$, we also write $z=\mathbf{y}^T\mathbf{x}$. It is therefore reasonable to associate each linear map on $\mathbb{R}^n$ with a real row vector $\mathbf{y}$, also referred to as *covector*, where $\mathbf{y}^T$ is considered as an element of the dual space. Based on the definition of dual space and covectors, it is nearby to define *codistributions* accordingly.

Definition B.35 Codistribution [Isi95]

Let $\mathcal{U}$ be an open subset of $\mathbb{R}^n$, covector fields $\mathbf{g}_i^T;\mathcal{U}\to(\mathbb{R}^n)^*$, $i=1,\ldots,d$ (composed of row vectors).

A codistribution $G(\mathbf{x})$ is defined as the assignment of a subspace of $(\mathbb{R}^n)^*$, spanned by the covector fields $\mathbf{g}_i^T(\mathbf{x})$, to every $\mathbf{x}\in\mathcal{U}$, i.e.

$$G(\mathbf{x})=\operatorname{span}\{\mathbf{g}_1^T(\mathbf{x}),\ldots,\mathbf{g}_d^T(\mathbf{x})\}$$

Remarks:

- At each point $\mathbf{x}_0 \in \mathcal{U}$, a codistribution is a subspace of $\left(\mathbb{R}^n\right)^*$.
- A covector field $\boldsymbol{\sigma}^T(\mathbf{x})$ is said to belong to the codistribution G, if $\boldsymbol{\sigma}^T(\mathbf{x}) \in G(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{U}$. This implies that $\boldsymbol{\sigma}^T(\mathbf{x})$ can be expressed as linear combination of $\mathbf{g}_i^T(\mathbf{x})$:

$$\boldsymbol{\sigma}^T(\mathbf{x}) = \sum_{i=1}^{d} c_i(\mathbf{x}) \mathbf{g}_i^T(\mathbf{x})$$

where $c_i(\mathbf{x})$ are real valued scalar functions on $\mathcal{U}$.

- A matrix-function

$$\mathbf{G}^T = \begin{bmatrix} \mathbf{g}_1^T \\ \vdots \\ \mathbf{g}_d^T \end{bmatrix}$$

- can be associated with a codistribution in a way that the codistribution is spanned by the rows of $\mathbf{G}^T$. Note that bold notation $\mathbf{G}$ is used for the matrix representation while an italic capital letter G is used if the distribution itself is addressed.
- The smoothness property, the terms “dimension”, “nonsingular distribution” as well as “regular points” are defined analogously for codistributions.
- The operations of addition and intersection are defined analogously to distributions

The following definition introduces a relationship between a distribution (codistribution) and an associated codistribution (distribution), called *annihilator*.

Definition B.36 Annihilator [Isi95]

Let $\mathcal{U}$ be an open subset of $\mathbb{R}^n$ and F a distribution (codistribution) on $\mathcal{U}$. The *annihilator* of F, denoted as $F^\perp$, is defined as the codistribution (distribution) whose elements annihilate every element of F, i.e.

$$F^\perp(\mathbf{x}) = \left\{ \boldsymbol{\tau}^T(\mathbf{x}) \in \left(\mathbb{R}^n\right)^* \,\middle|\, \boldsymbol{\tau}^T(\mathbf{x}) \boldsymbol{\sigma}(\mathbf{x}) = 0 \;\forall\, \boldsymbol{\sigma}(\mathbf{x}) \in F(\mathbf{x}), \forall \mathbf{x} \in \mathcal{U} \right\}$$

$$\left(F^\perp(\mathbf{x}) = \left\{ \boldsymbol{\sigma}(\mathbf{x}) \in \mathbb{R}^n \,\middle|\, \boldsymbol{\tau}^T(\mathbf{x}) \boldsymbol{\sigma}(\mathbf{x}) = 0 \;\forall\, \boldsymbol{\tau}^T(\mathbf{x}) \in F(\mathbf{x}), \forall \mathbf{x} \in \mathcal{U} \right\} \right)$$

Remarks

There are some useful properties concerning the relationship between a distribution (codistribution) and its annihilator

1. Let a distribution (codistribution) F and its annihilator $F^\perp$ be defined on a subset of $\mathbb{R}^n$, then

$$\dim(F)+\dim(F^{\perp})=n$$

Let F_1 and F_2 be distributions or codistributions, defined on a subset of $\mathbb{R}$, then

2. the annihilator of an intersection of distributions (codistributions) is equal to the sum of the annihilators of the separate distributions (codistributions)

$$(F_1 \cap F_2)^{\perp} = F_1^{\perp} + F_2^{\perp}$$

3. the annihilator of a sum of distributions (codistributions) is equal to the intersection of the annihilators of the separate distributions (codistributions)

$$(F_1 + F_2)^{\perp} = F_1^{\perp} \cap F_2^{\perp}$$

Proofs:
The proofs are presented for the case that F, F_1, F_2 are distributions. Analogous proofs for codistribution F, F_1, F_2 follow analogous arguments.

FIRST PART
The first part is proved by using the matrix notation of distributions. Let, for $\mathbf{x} \in \mathcal{U}$, F have dimension d. Then, there are vector fields $\mathbf{f}_i \in \mathbb{R}^n, i=1,\ldots,d$, which are linearly independent. Therefore the associated matrix in $\mathbb{R}^{n\times d}$, $\mathbf{F}=[\mathbf{f}_1 \quad \cdots \quad \mathbf{f}_d]$, has rank d and an element of F is obtained by $\boldsymbol{\tau}=c_1\mathbf{f}_1+\ldots+c_d\mathbf{f}_d=\mathbf{F}\cdot\mathbf{c}$, where $\mathbf{c}^T=(c_1 \quad \cdots \quad c_d)$ and c_i are real scalar fields.
All covectors $\boldsymbol{\sigma}^T \in F^{\perp}$ annihilate all vectors $\boldsymbol{\tau} \in F$ and therefore satisfy $\boldsymbol{\sigma}^T \cdot \mathbf{F} \cdot \mathbf{c} = 0$ for any $\mathbf{c} \in \mathbb{R}^d$. This, however, is possible if and only, if $\mathbf{F}^T\boldsymbol{\sigma}=0$. This equation is satisfied exactly by the kernel of $\mathbf{F}^T$, whose dimension is n-d, which follows from $\operatorname{rank}(\mathbf{F})=\operatorname{rank}(\mathbf{F}^T)=d$ (Corollary B.3) and Theorem B.15. Hence, one can find an n-d dimensional subspace of $(\mathbb{R}^n)^*$ which annihilates any $\boldsymbol{\tau} \in F$. Therefore $\dim(F^{\perp})=n-d$, which proves the claim.

SECOND PART
With $\widehat{F}=F_1 \cap F_2$, let $\dim(F_1)=d_1$, $\dim(F_2)=d_2$ and $\dim(\widehat{F})=d$, for $\mathbf{x} \in \mathcal{U}$. Moreover, there are matrices $\mathbf{F}_1 \in \mathbb{R}^{n\times d_1}$ and $\mathbf{F}_2=\mathbb{R}^{n\times d_2}$, associated with F_1 and F_2 respectively, whose columns are linearly independent and span the subspaces of F_1 and F_2. Consequently, some element $\boldsymbol{\tau}$, belonging to the intersection $\widehat{F}$ has to fulfill simultaneously $\boldsymbol{\tau}=\mathbf{F}_1\cdot\mathbf{c}$ and $\boldsymbol{\tau}=\mathbf{F}_2\cdot\mathbf{d}$ for some $\mathbf{c}\in\mathbb{R}^{d_1}$ and $\mathbf{d}\in\mathbb{R}^{d_2}$. This is equivalently expressed as.

$$[\mathbf{F}_1 \quad -\mathbf{F}_2]\begin{pmatrix}\mathbf{c}\\ \mathbf{d}\end{pmatrix}=\mathbf{0} \qquad \text{(B.13)}$$

Then, $\dim(\widehat{F}) = d$ if and only if exactly d linearly independent vectors $\widehat{\mathbf{f}}_i = \mathbf{F}_1 \cdot \mathbf{c}_i = \mathbf{F}_2 \cdot \mathbf{d}_i$ are obtained where the compound vectors $\begin{pmatrix} \mathbf{c}_i^T & \mathbf{d}_i^T \end{pmatrix}$ solve (B.13). To this end, it will be shown next that the existence of exactly d linearly independent $\widehat{\mathbf{f}}_i$ is tied to the condition that the kernel of $[\mathbf{F}_1 \quad -\mathbf{F}_2]$ has dimension d.

For the sufficient part, assume that $\ker([\mathbf{F}_1 \quad -\mathbf{F}_2])$ has dimension d. Then, there are exactly d linearly independent compound vectors $\begin{pmatrix} \mathbf{c}_i^T & \mathbf{d}_i^T \end{pmatrix}$, which fulfill (B.13) and the single vectors $\mathbf{c}_i$ and $\mathbf{d}_i$ are linearly independent on its own, which is shown in the following. Linear independence of the compound vectors implies that

$$\begin{bmatrix} \mathbf{C} \\ \mathbf{D} \end{bmatrix} \cdot \boldsymbol{\alpha} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix} \tag{B.14}$$

$$\text{where } \mathbf{C} = [\mathbf{c}_1 \quad \cdots \quad \mathbf{c}_d],\ \mathbf{D} = [\mathbf{d}_1 \quad \cdots \quad \mathbf{d}_d],\ \boldsymbol{\alpha} \in \mathbb{R}^d$$

if and only if $\boldsymbol{\alpha} = \mathbf{0}$. Assume, without loss of generality that $\mathbf{c}_i$ are linearly dependent. Then, there is some $\boldsymbol{\alpha} \neq \mathbf{0}$ such that $\mathbf{C} \cdot \boldsymbol{\alpha} = \mathbf{0}$ and (B.14) implies that $\mathbf{D}\boldsymbol{\alpha} = \mathbf{d}_\alpha \neq \mathbf{0}$. Applying this result to (B.13), we get

$$[\mathbf{F}_1 \quad -\mathbf{F}_2] \cdot \begin{bmatrix} \mathbf{C} \\ \mathbf{D} \end{bmatrix} \boldsymbol{\alpha} = -\mathbf{F}_2 \mathbf{d}_\alpha . \tag{B.15}$$

Since the columns of $\mathbf{F}_2$ are linearly independent and $\mathbf{d}_\alpha \neq \mathbf{0}$, consequently $\mathbf{F}_2 \mathbf{d}_\alpha \neq \mathbf{0}$ which is a contradiction to (B.13). Therefore, the single vectors $\mathbf{c}_i$ and $\mathbf{d}_i$ are linearly independent. Furthermore, linear independence of the $\widehat{\mathbf{f}}_i$ implies that

$$\mathbf{F}_1 \cdot \mathbf{C} \cdot \boldsymbol{\beta}_1 = \mathbf{0} \quad , \quad \mathbf{F}_2 \cdot \mathbf{D} \cdot \boldsymbol{\beta}_2 = \mathbf{0}$$

$$\text{where } \boldsymbol{\beta}_1, \boldsymbol{\beta}_2 \in \mathbb{R}^d$$

if and only if $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \mathbf{0}$. However, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \mathbf{0}$ follows from the fact that $\mathbf{F}_1$, $\mathbf{F}_2$, $\mathbf{C}$, $\mathbf{D}$ have full columns rank and Theorem B.15 (the dimension of the kernel is zero in each case).

The necessary part is done by contradiction. Assume that $\dim(\ker([\mathbf{F}_1 \quad -\mathbf{F}_2])) = d^* \neq d$. Then, there exist d^* linearly independent compound vectors $\begin{pmatrix} \mathbf{c}_i & \mathbf{d}_i \end{pmatrix}$ which satisfy (B.13). Following the arguments of the sufficient part, there are d^* linearly independent vectors $\widehat{\mathbf{f}}_i$. This contradicts the assumed existence of exactly d linearly independent $\widehat{\mathbf{f}}_i$, satisfying (B.13).

Summing up the last section, it can be stated that $\dim(\widehat{F}) = d$ if and only if $\dim(\ker([\mathbf{F}_1 \quad -\mathbf{F}_2])) = d$.

On the one hand, from the proof of part 1, there are matrices $\mathbf{G}_1^T \in \mathbb{R}^{(n-d_1)\times n}$ and $\mathbf{G}_2^T \in \mathbb{R}^{(n-d_2)\times n}$, each with linearly independent rows (hence $\operatorname{rank}(\mathbf{G}_1) = n - d_1$ and $\operatorname{rank}(\mathbf{G}_2) = n - d_2$) and associated with $G_1 = F_1^{\perp}$, $G_2 = F_2^{\perp}$ which satisfy

$$\mathbf{G}_1^T \mathbf{F}_1 = \mathbf{0} \text{ and } \mathbf{G}_2^T \mathbf{F}_2 = \mathbf{0}. \tag{B.16}$$

On the other hand, from proof of part 1, there is a matrix $\widehat{\mathbf{G}} \in \mathbb{R}^{(n-d)\times n}$, associated with $\widehat{G} = \widehat{F}^{\perp}$ which has n-d linearly independent rows (hence $\operatorname{rank}(\widehat{\mathbf{G}}) = n - d$) and fulfills $\widehat{\mathbf{G}}^T \widehat{\mathbf{f}}_i = \mathbf{0}$ for $i = 1, \ldots, d$ or, equivalently

$$\widehat{\mathbf{G}}^T \cdot \widehat{\mathbf{F}} = \mathbf{0} \tag{B.17}$$

$$\text{where } \widehat{\mathbf{F}} = \begin{bmatrix} \widehat{\mathbf{f}}_1 & \cdots & \widehat{\mathbf{f}}_d \end{bmatrix} = \mathbf{F}_1 \mathbf{C} = \mathbf{F}_2 \mathbf{D}.$$

Conversely, if another matrix $\widehat{\mathbf{G}}^{*T}$ fulfills (B.17), the subspace that is spanned by its rows belongs to $\widehat{G}$. By equations (B.16), such a matrix is given by

$$\widehat{\mathbf{G}}^{*T} = \begin{bmatrix} \mathbf{G}_1^T \\ \mathbf{G}_2^T \end{bmatrix} \tag{B.18}$$

since $\mathbf{G}_1^T \widehat{\mathbf{F}} = \mathbf{G}_1^T \mathbf{F}_1 \mathbf{C} = \mathbf{0}$ and $\mathbf{G}_2^T \widehat{\mathbf{F}} = \mathbf{G}_2^T \mathbf{F}_2 \mathbf{D} = \mathbf{0}$.

As intermediate result, note that some vector $\mathbf{h}$, which is not in the span of $\mathbf{F}_1$ ($\mathbf{F}_2$), is not in $\ker(\mathbf{G}_1^T)$ ($\ker(\mathbf{G}_2^T)$). On the one hand, since $\operatorname{rank}(\mathbf{G}_1) = n - d_1$, ($\operatorname{rank}(\mathbf{G}_2^T) = n - d_2$), we have $\dim(\ker(\mathbf{G}_1^T)) = d_1$ ($\dim(\ker(\mathbf{G}_2^T)) = d_2$) by Theorem B.15. On the other hand $\mathbf{F}_1$ ($\mathbf{F}_2$) has d_1 (d_2) linearly independent columns which consequently, in light of equation (B.16), span the whole kernel of $\mathbf{G}_1^T$ ($\mathbf{G}_2^T$). Hence $\mathbf{h}$ is not in $\ker(\mathbf{G}_1^T)$ ($\ker(\mathbf{G}_2^T)$) and consequently $\mathbf{G}_1^T \mathbf{h} \neq \mathbf{0}$ ($\mathbf{G}_2^T \mathbf{h} \neq \mathbf{0}$).

Furthermore, a vector belongs to $\ker(\widehat{\mathbf{G}}^{*T})$, if it is in the span of both, $\mathbf{F}_1$ and $\mathbf{F}_2$, i.e. in the intersection of the subspaces, spanned by $\mathbf{F}_1$ and $\mathbf{F}_2$. This, however, is exactly the subspace that is spanned by $\widehat{\mathbf{F}}$. By the intermediate consideration stated recently, if some vector $\mathbf{h}$ is not in the intersection of the subspaces, spanned by $\mathbf{F}_1$ and $\mathbf{F}_2$, it is not in $\ker(\widehat{\mathbf{G}}^{*T})$ since then, either $\mathbf{G}_1^T \mathbf{h} \neq \mathbf{0}$ or $\mathbf{G}_2^T \mathbf{h} \neq \mathbf{0}$. Consequently $\widehat{\mathbf{F}}$ spans the whole kernel of $\widehat{\mathbf{G}}^{*T}$ and has dimension d. Hence $\dim(\ker(\widehat{\mathbf{G}}^{*T})) = d$ and, by Theorem B.15, we have $\operatorname{rank}(\widehat{\mathbf{G}}^{*T}) = n - d$.

Overall, it has been shown that the rows of $\widehat{\mathbf{G}}^{*T}$ belong to $\widehat{G}$ and span a subspace of dimension n-d, which is exactly the dimension of the subspace of $\widehat{G}$. Hence, the rows of $\widehat{\mathbf{G}}^{*T}$ span $\widehat{G}$ on the one hand. On the other hand, the rows of $\widehat{\mathbf{G}}^{*T}$ also span $G_1 + G_2$. Therefore the rows of $\widehat{\mathbf{G}}^{*T}$ and $\widehat{\mathbf{G}}^{T}$ span the same subspace, which finally proves that $G_1 + G_2 = \widehat{G}$.

The proof for the third part is done along similar lines as the second part and is therefore dropped.

□

The next theorem is central in the theory of distributions and plays an important role in the proof for the existence of a transformation to input-normalized Byrnes-Isidori normal form. The detailed proof is presented in [Isi95].

Theorem B.10 Frobenius [Isi95]

Let F be a smooth distribution on $\mathcal{U}$, an open subset of $\mathbb{R}^n$.

Let further F be nonsingular and $\dim(F)=d$

Then there exist n-d functions $\lambda_i : \mathcal{U} \to \mathbb{R}$, $i=1,\ldots,n-d$, such that

$$F^{\perp} = \operatorname{span}\left\{ \frac{\partial \lambda_1(\mathbf{x})}{\partial \mathbf{x}} \quad \cdots \quad \frac{\partial \lambda_{n-d}(\mathbf{x})}{\partial \mathbf{x}} \right\} \ \forall \mathbf{x} \in \mathcal{U}$$

if and only if

$F^{\perp}$ is involutive

B.4 Some Facts on Nonlinear Maps

Proper Map and C^k Diffeomorphism

Amongst others, Chapter 4 is concerned with nonaffine-in-control dynamical systems. It plays an important role for nonlinear-in-control design in section 4.1.7 under which conditions a nonlinear map has a unique inverse. Therefore nonlinear maps $\mathbf{f} : \mathcal{X} \to \mathcal{Y}$, $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^m$, i.e. vector fields in $\mathbb{R}^m$ are investigated.

It is a well-known fact that in $\mathbb{R}$, a continuously differentiable function is uniquely invertible if it is monotonic or, equivalently, if its derivative does not have any zeros on the whole definition space. Therefore, one could conjecture, that this condition can be transferred to $\mathbb{R}^m$, such that $\mathbf{f}$ is invertible if its Jacobian is regular.

$$\det\left(\frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right) \neq 0 \, \forall \mathbf{x} \in \mathcal{X}$$

However it turns out that this condition is not sufficient in $\mathbb{R}^m$, $m>1$. Additionally the map has to owe the property of being proper ([San80]).

Definition B.37 Proper Map

Let a map $\mathbf{f}(\mathcal{X})=\mathcal{Y},\ \mathcal{X},\mathcal{Y}\subset\mathbb{R}^m$

The mapping $\mathbf{f}$ is said to be proper, if the preimage of any compact subset $U\subset\mathcal{Y}$:

$V=\mathbf{f}^{-1}(U)=\{\mathbf{x}\in\mathcal{X}\,|\,\mathbf{f}(\mathbf{x})\in U\}$ is compact

Remark

Note that Theorem B.8 assures for continuous functions that the image of any compact set in the definition space is compact, i.e. exactly the other way round as required for a proper map. If we assume $V\in\mathcal{X}$ is compact, then $U=\mathbf{f}(V)$ is compact for a continuous mapping. However, for a proper map, we assume that U is compact and we require that V compact. Hence, continuity is not sufficient for a map to be proper.

Next, we define a C^k diffeomorphism, which is a map, k times continuously differentiable, that can be uniquely inverted and whose inverse map is k times continuously differentiable as well.

Definition B.38 C^k Diffeomorphism

Let $\mathbf{f}(\mathcal{X})=\mathcal{Y},\ \mathcal{X},\mathcal{Y}$ open subsets of $\mathbb{R}^m$. $\mathbf{f}$ is a C^k diffeomorphism of $\mathcal{X}$ onto $\mathcal{Y}$, if

- $\mathbf{f}\in C^k$ (k times continuously differentiable)
- $\mathbf{f}$ is a bijection
- $\mathbf{f}^{-1}\in C^k$ exists

The next theorem of Hadamard plays an important role as it provides necessary and sufficient conditions for a mapping to be a C^k diffeomorphism. A proof is given in [WuF72].

Theorem B.11 Hadamard

Let $\mathbf{f}(\mathcal{X})=\mathcal{Y},\ \mathcal{X},\mathcal{Y}$ open subsets of $\mathbb{R}^m$, $\mathbf{f}\in C^k$

$\mathbf{f}$ is a C^k-Diffeomorphism if and only if

$\det\left(\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\right)\neq 0$ for all $\mathbf{x}\in\mathcal{X}$ and $\mathbf{f}$ is proper

In practice, the verification of a nonsingular Jacobian is straightforward. Potentially more complicated will be the verification of the rather abstract quality of being proper. Though, Lemma B.5 presents a verifiable condition for proper maps. It is motivated by the corresponding lemma in [Lav08] but it uses a slightly modified condition involving the integral of the Jacobian.

In [Lav08] there is a little imprecision. The lemma is claimed to hold for any subset of $\mathbb{R}^m$, however the proof only allows to consider the lemma to be true for a map on the whole $\mathbb{R}^m$ as it requires the map to be radially unbounded, which can only be fulfilled, if the map is defined on the whole $\mathbb{R}^m$. The modified condition in Lemma B.5, however, allows at least an extension for maps that exist on open convex subsets of $\mathbb{R}^m$. But before, a fact is presented that appears to be useful in the proof of Lemma B.5.

Lemma B.4 Minimum and Maximum Distance for Nonlinear Maps

Let $\mathcal{X}$ be an open convex set in $\mathbb{R}^m$,

a continuously differentiable map $\mathbf{f}(\mathcal{X})=\mathcal{Y}\subset\mathbb{R}^m$ and $\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}\in\mathcal{X}$. Further

$$\mathbf{J}(\mathbf{x})=\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}}\ ,\ \mathbf{J}_s(\mathbf{x})=\frac{1}{2}\left[\mathbf{J}(\mathbf{x})+\mathbf{J}^T(\mathbf{x})\right]\ ,\ \mathbf{P}(\mathbf{x}_1,\mathbf{x}_2)=\int_0^1\mathbf{J}\left[\mathbf{x}_1+s(\mathbf{x}_2-\mathbf{x}_1)\right]ds\,,$$

$\underline{\sigma}_P,\overline{\sigma}_P$: minimum singular value of $\mathbf{P}(\mathbf{x}_1,\mathbf{x}_2)$, $\lambda_i[\mathbf{J}_s(\mathbf{x})]$: i^{th} eigenvalue of $\mathbf{J}_s(\mathbf{x})$

If, for some $k_1>0$ either

1. $\underline{\sigma}_P\geq k_1$ or
2. $\mathbf{J}_s(\mathbf{x})$ positive or negative definite such that

   either $\lambda_i[\mathbf{J}_s(\mathbf{x})]\geq k_1$ or $\lambda_i[\mathbf{J}_s(\mathbf{x})]\leq -k_1$ respectively for $i=1,\ldots,m$

then $\left\|\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1)\right\|_2\geq k_1\left\|\mathbf{x}_2-\mathbf{x}_1\right\|_2$.

If, for some $k_2>0$

3. $\overline{\sigma}_P\leq k_2$ then $\left\|\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1)\right\|_2\leq k_2\left\|\mathbf{x}_2-\mathbf{x}_1\right\|_2$

Proof

At first note that $\mathcal{X}$ has to be convex since it is necessary that the straight line connecting $\mathbf{x}_1$ and $\mathbf{x}_2$ has to belong to $\mathcal{X}$ such that $\mathbf{P}(\mathbf{x}_1,\mathbf{x}_2)$ exists. From the mean value theorem, the difference between 2 function values is obtained by integrating the Jacobian along a curve connecting $\mathbf{x}_1$ and $\mathbf{x}_2$

$$\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1)=\int_\gamma\mathbf{J}(\boldsymbol{\gamma})\cdot d\boldsymbol{\gamma}$$

where γ denotes a straight line, connecting $\mathbf{x}_1$ and $\mathbf{x}_2$. A possible parameterization is

$$\boldsymbol{\gamma}(s)=\mathbf{x}_1+(\mathbf{x}_2-\mathbf{x}_1)s$$

where $s\in[0,1]$ is the curve parameter. Thus the mean value theorem is

$$\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1)=\int_\gamma\mathbf{J}[\boldsymbol{\gamma}(s)]\cdot\frac{d\boldsymbol{\gamma}(s)}{ds}\cdot ds=\int_0^1\mathbf{J}[\mathbf{x}_1+s(\mathbf{x}_2-\mathbf{x}_1)]\cdot ds\cdot(\mathbf{x}_2-\mathbf{x}_1)$$

and thus

$$\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1)=\mathbf{P}(\mathbf{x}_1,\mathbf{x}_2)\cdot(\mathbf{x}_2-\mathbf{x}_1) \tag{B.19}$$

Taking the 2-norm on both sides, we obtain by application of Theorem B.18

$$\underline{\sigma}_P\cdot\|\mathbf{x}_2-\mathbf{x}_1\|_2\le\|\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1)\|_2\le\overline{\sigma}_P\cdot\|\mathbf{x}_2-\mathbf{x}_1\|_2.$$

and conditions 1 and 3 yield

$$k_1\cdot\|\mathbf{x}_2-\mathbf{x}_1\|_2\le\|\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1)\|_2\le k_2\cdot\|\mathbf{x}_2-\mathbf{x}_1\|_2. \tag{B.20}$$

Further, consider the following Cauchy-Schwartz inequality.

$$\|\mathbf{x}_2-\mathbf{x}_1\|_2\|\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1)\|_2\ge\left|(\mathbf{x}_2-\mathbf{x}_1)^T(\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1))\right|. \tag{B.21}$$

Then from result of Theorem B.19 and equation (B.19) we get

$$\begin{aligned}\left|(\mathbf{x}_2-\mathbf{x}_1)^T(\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1))\right|&=\left|\int_0^1(\mathbf{x}_2-\mathbf{x}_1)^T\mathbf{J}[\mathbf{x}_1+s(\mathbf{x}_2-\mathbf{x}_1)]\cdot(\mathbf{x}_2-\mathbf{x}_1)\cdot ds\right|\\&=\left|\int_0^1(\mathbf{x}_2-\mathbf{x}_1)^T\mathbf{J}_s[\mathbf{x}_1+s(\mathbf{x}_2-\mathbf{x}_1)]\cdot(\mathbf{x}_2-\mathbf{x}_1)\cdot ds\right|\end{aligned}$$

and application of condition 2 in connection with Theorem B.20 and equation (B.21) yields the left inequality of (B.20).

□

Lemma B.5 Verifiable Conditions for Proper Maps

Let $\mathcal{X}$ be an open convex set in $\mathbb{R}^m$, a continuously differentiable map $\mathbf{f}(\mathcal{X})=\mathcal{Y}\subset\mathbb{R}^m$ and $\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}\in\mathcal{X}$. Further

$$\mathbf{J}(\mathbf{x})=\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}},\ \mathbf{J}_s(\mathbf{x})=\frac{1}{2}\left[\mathbf{J}(\mathbf{x})+\mathbf{J}^T(\mathbf{x})\right]\text{ and }\mathbf{P}(\mathbf{x}_1,\mathbf{x}_2)=\int_0^1\mathbf{J}\left[(\mathbf{x}_2-\mathbf{x}_1)\cdot s+\mathbf{x}_1\right]\cdot ds$$

$\underline{\sigma}_P$ minimum singular value of $\mathbf{P}(\mathbf{x}_1,\mathbf{x}_2)$

$\lambda_i[\mathbf{J}_s(\mathbf{x})]$: i^{th} eigenvalue of $\mathbf{J}_s(\mathbf{x})$

If either of the conditions is fulfilled for some $k>0$

1. $\underline{\sigma}_P\ge k$
2. $\mathbf{J}_s(\mathbf{x})$ is either positive or negative definite
   such that either $\lambda_i[\mathbf{J}_s(\mathbf{x})]\ge k$ or $\lambda_i[\mathbf{J}_s(\mathbf{x})]\le -k$
   for $i=1,\ldots,m$ respectively

then $\mathbf{f}$ is proper.

Proof

By Lemma B.4 in connection with condition 2, we have

$$\left\|\mathbf{f}(\mathbf{x}_2)-\mathbf{f}(\mathbf{x}_1)\right\|_2 \geq k\cdot\left\|\mathbf{x}_2-\mathbf{x}_1\right\|_2 \tag{B.22}$$

Then, Definition B.24 says that compactness of a set $\mathcal{U}$ is tied to the property that every sequence with elements in $\mathcal{U}$ has a subsequence that converges in $\mathcal{U}$.

Therefore, let some set $\mathcal{V}\subset\mathcal{X}$, such that $\mathcal{U}=\mathbf{f}(\mathcal{V})$ is compact. Then, take a sequence $\{\mathbf{v}_i\}$ such that $\mathbf{v}_i\in\mathcal{V}$, $i\in\mathbb{N}$. Then the sequence $\mathbf{u}_i=\mathbf{f}(\mathbf{v}_i)$ has its elements in $\mathcal{U}$. Further, since $\mathcal{U}$ is compact by assumption, $\{\mathbf{u}_i\}$ has a subsequence $\{\overline{\mathbf{u}}_i\}$ that converges to some $\mathbf{u}_0\in\mathcal{U}$.

To $\{\mathbf{v}_i\}$, define a subsequence $\{\overline{\mathbf{v}}_i\}$, such that $\mathbf{f}(\overline{\mathbf{v}}_i)=\overline{\mathbf{u}}_i$ and $\mathbf{v}_0\in\mathcal{V}$ such that $\mathbf{f}(\mathbf{v}_0)=\mathbf{u}_0$. Since $\{\overline{\mathbf{u}}_i\}$ converges to $\mathbf{u}_0$, for every $\varepsilon_u>0$ there is some $N(\varepsilon_u)\in\mathbb{N}$ such that

$$\left\|\overline{\mathbf{u}}_i-\mathbf{u}_0\right\|<\varepsilon_u$$

for all $i\geq N(\varepsilon_u)$. Then, with (B.22) one can conclude, that

$$\left\|\overline{\mathbf{v}}_i-\overline{\mathbf{v}}_0\right\|\leq\frac{\left\|\overline{\mathbf{u}}_i-\mathbf{u}_0\right\|}{k}<\frac{\varepsilon_u}{k}=:\varepsilon_v$$

Hence for every $\varepsilon_v>0$, we have $\left\|\overline{\mathbf{v}}_i-\mathbf{v}_0\right\|<\varepsilon_v$ for all $i\geq N(k\cdot\varepsilon_v)$, in other words, we have constructed a subsequence $\{\overline{\mathbf{v}}_i\}$ that converges in $\mathcal{V}$. This procedure is possible for any sequence in $\mathcal{V}$, hence $\mathcal{V}$ is compact (by Definition B.24) and consequently $\mathbf{f}$ is a proper mapping.

□

The following corollary is a direct consequence of Lemma B.5 Hadamard's Theorem (Theorem B.11).

Corollary B.1 C^k **Diffeomorphism**

Let $\mathcal{X}$ be an open convex set in $\mathbb{R}^m$, a C^k map $\mathbf{f}(\mathcal{X})=\mathcal{Y}\subset\mathbb{R}^m$ and $\mathbf{x}_1,\mathbf{x}_2,\mathbf{x}\in\mathcal{X}$. Further

$$\mathbf{J}(\mathbf{x})=\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}},\ \mathbf{J}_s(\mathbf{x})=\frac{1}{2}\left[\mathbf{J}(\mathbf{x})+\mathbf{J}^T(\mathbf{x})\right]\text{ and }\mathbf{P}(\mathbf{x}_1,\mathbf{x}_2)=\int_0^1\mathbf{J}\left[(\mathbf{x}_2-\mathbf{x}_1)\cdot s+\mathbf{x}_1\right]\cdot ds$$

$\underline{\sigma}_P$: minimum singular value of $\mathbf{P}(\mathbf{x}_1,\mathbf{x}_2)$

$\lambda_i[\mathbf{J}_s(\mathbf{x})]$: i^{th} eigenvalue of $\mathbf{J}_s(\mathbf{x})$

Then $\mathbf{f}$ is a C^k diffeomorphism of $\mathcal{X}$ onto $\mathcal{Y}$,

if $\det[\mathbf{J}(\mathbf{x})] \neq 0$ and either of the conditions is fulfilled for some $k > 0$

1. $\underline{\sigma}_P \geq k$
2. $\mathbf{J}_s(\mathbf{x})$ is positive or negative definite, such that
   either $\lambda_i[\mathbf{J}_s(\mathbf{x})] \geq k$ or $\lambda_i[\mathbf{J}_s(\mathbf{x})] \leq -k$ for $i = 1, \ldots, n$

Local Solvability of Systems of Nonlinear Equations

From section 4.1.2 it is known that control affine systems that have a relative degree can be solved for the controls globally, i.e. one can always find a control $\mathbf{u}$ such that

$$\mathbf{a}(\zeta, \boldsymbol{\eta}) + \mathbf{B}(\zeta, \boldsymbol{\eta})\mathbf{u} = \mathbf{v}$$

for any $\mathbf{v} \in \mathbb{R}^m$. Global feedback linearizability of nonaffine dynamic systems with a relative degree as introduced in section 4.1.7 is not given a priori but only local solvability can be assured. In the latter case, the equation to be solved for the controls is given by

$$\mathbf{a}(\zeta, \boldsymbol{\eta}, \mathbf{u}_N) = \mathbf{v}\,. \tag{B.23}$$

For the remainder, we will consider the state $\zeta, \boldsymbol{\eta}$ as some fixed parameter, while $\mathbf{u}_N$ is the variable to be solved for. As one can expect, the equation cannot be solved for any $\mathbf{v}$ in general. However, if one can find some $\mathbf{u}_N^0, \mathbf{v}^0$ that solve (B.23), it can be shown, under certain conditions, that it has a solution in a vicinity of $\mathbf{v}^0$, too. Towards this end recall the implicit function theorem, which is well-known in literature and a proof is e.g. presented in [Oli13].

Theorem B.12 Implicit Function

Let $\mathcal{X} \subset \mathbb{R}^m, \mathcal{Y} \subset \mathbb{R}^n$ a C^k map: $\mathbf{f} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^m$

Assume, for some $(\mathbf{x}_0, \mathbf{y}_0) \in \mathcal{X} \times \mathcal{Y}$, $\mathbf{f}(\mathbf{x}_0, \mathbf{y}_0) = \mathbf{0}$

If $\dfrac{\partial \mathbf{f}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}}$ is regular at $(\mathbf{x}_0, \mathbf{y}_0)$

Then, there are vicinities $U \subset \mathcal{X}$ of $\mathbf{x}_0$ and $V \subset \mathcal{Y}$ of $\mathbf{y}_0$
and a C^k map $\mathbf{g} : U \to V$ such that

$\mathbf{f}(\mathbf{x}, \mathbf{g}(\mathbf{x})) = \mathbf{0}$ for all $\mathbf{x} \in U$

Using the Implicit Function Theorem, the following fact can be shown.

Corollary B.2 Local Feedback Linearizability of Nonaffine Systems

Let $\mathcal{D}_\zeta \subset \mathbb{R}^r, \mathcal{D}_\eta \subset \mathbb{R}^{n-r}, \mathcal{U} \subset \mathbb{R}^{p_N}$

and a continuously differentiable map $\mathbf{a}: \mathcal{D}_\zeta \times \mathcal{D}_\eta \times \mathcal{U} \to \mathbb{R}^m$, $m \le r \le n$, $p_N \ge m$

Suppose that there are $\boldsymbol{\zeta}^0 \in \mathcal{D}_\zeta, \boldsymbol{\eta}^0 \in \mathcal{D}_\eta, \mathbf{u}_N^0 \in \mathcal{U}$ and $\mathbf{v}^0 \in \mathbb{R}^m$

such that $\mathbf{a}(\boldsymbol{\zeta}^0, \boldsymbol{\eta}^0, \mathbf{u}_N^0) = \mathbf{v}^0$.

Then, there are vicinities $\mathcal{V}$ of $\mathbf{v}^0$, $\mathcal{Z}$ of $\boldsymbol{\zeta}^0$ $\mathcal{Y}$ of $\boldsymbol{\eta}^0$ and U of $\mathbf{u}_N^0$

such that, for any $\zeta \in \mathcal{Z}, \zeta \in \mathcal{Y}, \mathbf{v} \in \mathcal{V}$, $\mathbf{a}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N) = \mathbf{v}$ is solved for some $\mathbf{u}_N \in U$

if $\dfrac{\partial \mathbf{a}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N)}{\partial \mathbf{u}_N}$ has full row rank at $(\boldsymbol{\zeta}^0, \boldsymbol{\eta}^0, \mathbf{u}_N^0)$

Proof

At first, denote

$$\mathbf{J}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N) := \left[\mathbf{j}_1(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N) \quad \cdots \quad \mathbf{j}_{p_N}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N)\right] := \frac{\partial \mathbf{a}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N)}{\partial \mathbf{u}_N}$$

where $\mathbf{j}_i(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N) \in \mathbb{R}^m$. The full rank condition implies that $\mathbf{J}(\boldsymbol{\zeta}^0, \boldsymbol{\eta}^0, \mathbf{u}_N^0)$ has m linearly independent columns (Theorem B.14). Without loss of generality, assume that the first m columns are linearly independent. If it is not the case, resort the elements of $\mathbf{u}_N$. Then, split the control vector

$$\mathbf{u}_N = \begin{pmatrix} \mathbf{u}_{N1} \\ \mathbf{u}_{N2} \end{pmatrix}, \quad \mathbf{u}_N^0 = \begin{pmatrix} \mathbf{u}_{N1}^0 \\ \mathbf{u}_{N2}^0 \end{pmatrix}$$

where $\mathbf{u}_{N1}$ contains the first m elements, and $\mathbf{u}_{N2}$ the remaining elements of $\mathbf{u}_N$. Further, define:

- The regular quadratic matrix $\bar{\mathbf{J}}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N) = \dfrac{\partial \mathbf{a}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N)}{\partial \mathbf{u}_{N1}} = \left[\mathbf{j}_1(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N) \quad \cdots \quad \mathbf{j}_m(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N)\right]$
- vectors $\mathbf{x} = \mathbf{u}_{N1}$, $\mathbf{y}^T = \begin{pmatrix} \mathbf{u}_{N2}^T & \boldsymbol{\zeta}^T & \boldsymbol{\eta}^T & \mathbf{v}^T \end{pmatrix}$
- the map $\mathbf{f}(\mathbf{x}, \mathbf{y}) = \mathbf{a}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N) - \mathbf{v}$

Then

$$\frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{y})}{\partial \mathbf{x}} = \bar{\mathbf{J}}(\boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{u}_N)$$

is regular at $\mathbf{x}_0 = \mathbf{u}_{N1}^0$, $(\mathbf{y}^0)^T = \begin{pmatrix} \mathbf{u}_{N2}^0 & \boldsymbol{\zeta}^0 & \boldsymbol{\eta}^0 & \mathbf{v}^0 \end{pmatrix}$ and hence by Theorem B.12 there are vicinities U_1 of $\mathbf{u}_{N1}^0$ and $U_2 \times \mathcal{Z} \times \mathcal{Y} \times \mathcal{V}$ of $\begin{pmatrix} \mathbf{u}_{N2}^0 & \boldsymbol{\zeta}^0 & \boldsymbol{\eta}^0 & \mathbf{v}^0 \end{pmatrix}$ and a C^1 map

$$\mathbf{g}: U_1 \to U_2 \times \mathcal{Z} \times \mathcal{Y} \times \mathcal{V}$$

such that $\mathbf{f}(\mathbf{x}, \mathbf{g}(\mathbf{x})) = \mathbf{0}$. Transferring back to original notation, and splitting up $\mathbf{g}$

$$\begin{pmatrix} \mathbf{u}_{N2} \\ \boldsymbol{\zeta} \\ \boldsymbol{\eta} \\ \boldsymbol{\nu} \end{pmatrix} = \begin{pmatrix} \mathbf{g}_u(\mathbf{u}_{N1}) \\ \mathbf{g}_\zeta(\mathbf{u}_{N1}) \\ \mathbf{g}_\eta(\mathbf{u}_{N1}) \\ \mathbf{g}_\nu(\mathbf{u}_{N1}) \end{pmatrix} = \mathbf{g}(\mathbf{u}_{N1})$$

we arrive at

$$\mathbf{a}\left[\mathbf{u}_N, \mathbf{g}_\zeta(\mathbf{u}_{N1}), \mathbf{g}_\eta(\mathbf{u}_{N1})\right] - \mathbf{g}_\nu(\mathbf{u}_{N1}) = \mathbf{0}$$

for $\mathbf{u}_{N1} \in U_1$ and

$$\mathbf{u}_N = \begin{pmatrix} \mathbf{u}_{N1} \\ \mathbf{g}_u(\mathbf{u}_{N1}) \end{pmatrix}.$$

Hence, the equation of the corollary is also solved in a vicinity $U = U_1 \times U_2$ of $\mathbf{u}_N^0$,.

□

B.5 Matrix Properties

Subsequently some properties, pertaining to matrices that appear to be useful throughout the thesis are presented. The proofs are quite easy and most of them can be found in any appropriate mathematical textbook, however, for better understanding they are included here.

Theorem B.13 Hermitian Matrix - Inverse

For a regular matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$, the following equation is valid.

$$\left(\mathbf{A}^{-1}\right)^H = \left(\mathbf{A}^H\right)^{-1}$$

Proof

$$\left(\mathbf{A}\mathbf{A}^{-1}\right)^H = \mathbf{I} \quad \Rightarrow \quad \left(\mathbf{A}^{-1}\right)^H \mathbf{A}^H = \mathbf{I} \Rightarrow \quad \left(\mathbf{A}^{-1}\right)^H = \left(\mathbf{A}^H\right)^{-1}$$

□

Theorem B.14 Row Rank – Column Rank

Let a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$. Let further p and q be the number of linearly independent rows and columns in $\mathbf{A}$ respectively.

Then $p=q$ and $p=q$ is denoted as $\operatorname{rank}(\mathbf{A})$

Proof

Let a scalar expansion of $\mathbf{A}$ be given by

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix},$$

the i^{th} row is denoted as $\mathbf{a}_i^T = \begin{pmatrix} a_{i1} & \cdots & a_{im} \end{pmatrix}$.

Then, since there are p linearly independent rows, they span a subspace of dimension p. Let a basis of the subspace be given by

$$\mathbf{b}_1^T = \begin{pmatrix} b_{11} & \cdots & b_{1m} \end{pmatrix} \quad \ldots \; \mathbf{b}_p^T = \begin{pmatrix} b_{p1} & \cdots & b_{pm} \end{pmatrix},$$

and the i^{th} row of $\mathbf{A}$ can be composed as a linear combination of the basis vectors, i.e.

$$\mathbf{a}_i^T = k_{i1}\mathbf{b}_1^T + \ldots + k_{ip}\mathbf{b}_p^T.$$

and the whole matrix, in terms of the basis vectors reads as

$$\mathbf{A} = \begin{bmatrix} \mathbf{k}_1 & \cdots & \mathbf{k}_p \end{bmatrix} \begin{bmatrix} \mathbf{b}_1^T \\ \vdots \\ \mathbf{b}_p^T \end{bmatrix}$$

where $\mathbf{k}_k^T = \begin{pmatrix} k_{1k} & \cdots & k_{nk} \end{pmatrix}$, $k = 1, \ldots, p$. Obviously, each column of $\mathbf{A}$ is a linear combination of the column vectors $\mathbf{k}_1 \ldots \mathbf{k}_p$ which implies that there are, at maximum, p linearly independent columns and hence $q \leq p$. An analog consideration, starting with the q linearly independent columns, results in $p \leq q$ which, finally, implies $p=q$.

□

Corollary B.3 Rank of A and A^T

Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$. Then $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^T)$

Proof

The proof follows directly from Theorem B.14.

□

Theorem B.15 Rank and Kernel

Let a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ with $\dim(\ker(\mathbf{A})) = k$.

Then $\operatorname{rank}(\mathbf{A}) = r$ such that $r + k = m$

Proof

Let the columns of $\mathbf{V}=[\mathbf{v}_1 \ \cdots \ \mathbf{v}_k]$, $\mathbf{v}_i \in \mathbb{R}^m$ be a basis for $\ker(\mathbf{A})$. Let further $\mathbf{U}=[\mathbf{u}_1 \ \cdots \ \mathbf{u}_{m-k}]$ complete a basis in $\mathbb{R}^m$. Since $\operatorname{rank}(\mathbf{A})=\dim(\operatorname{image}(\mathbf{A}))$, it needs to be shown that $\dim(\operatorname{image}(\mathbf{A}))=m-k$. To this end notice that a vector $\mathbf{x}$ produces a nonzero output by application of $\mathbf{A}$, if and only if it is composed from a nontrivial linear combination of the columns of $\mathbf{U}$. In other words, if

$$\mathbf{x} = \mathbf{U} \cdot \mathbf{c}$$

for some $\mathbf{c} \in \mathbb{R}^{m-k}$, then

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{A} \cdot (\mathbf{U}\mathbf{c}) = \mathbf{0} \text{ if and only if } \mathbf{c} = \mathbf{0}.$$

This implies that columns of the matrix

$$\mathbf{A}_U = \mathbf{A}\mathbf{U} \in \mathbb{R}^{n\times(m-k)}$$

are linearly independent. Hence, the space that is spanned by $\mathbf{A}_U\mathbf{c}$ has dimension m-k and is equal to the space that is spanned by $\mathbf{A}\mathbf{x}$. Thus $\dim(\operatorname{image}(\mathbf{A}))=m-k$.

□

Theorem B.16 Kernel A and A^T

Let a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$ with $\operatorname{rank}(\mathbf{A})=r$, $\dim(\ker(\mathbf{A}))=k_1$ and $\dim(\ker(\mathbf{A}^T))=k_2$.
Then $k_1=m-r$ and $k_2=n-r$

Proof

The proof follows from the fact that column and row rank of a matrix are equal (Theorem B.14) and by application of Theorem B.15 to $\mathbf{A}$ and $\mathbf{A}^T$.

□

Theorem B.17 Symmetric Positive Definite Matrices, Eigenvalues, Singular Values, Inverse

Let $\mathbf{\Gamma} \in \mathbb{R}^{n\times n}$ be symmetric positive semi-definite. Then, for $i=1,\ldots,n$,

- Eigenvalues equal singular values $\lambda_{\Gamma,i}=\sigma_{\Gamma,i}(\mathbf{\Gamma})$
- It exists, the eigenvalues of $\mathbf{\Gamma}^{-1}$ equal the inverse of the eigenvalues of $\mathbf{\Gamma}$: $\lambda_i(\mathbf{\Gamma}^{-1})=\lambda_i^{-1}(\mathbf{\Gamma})$

Proof

The singular values of $\mathbf{\Gamma}$ are computed from the square-roots of the eigenvalues of $\mathbf{\Gamma}^T\mathbf{\Gamma}$. Therefore, let

$$\mathbf{\Gamma} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \tag{B.24}$$

be an eigenvalue decomposition of $\mathbf{\Gamma}$, where $\mathbf{Q}$ is an orthonormal matrix (i.e. $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$) with eigenvectors in its columns and

$$\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_{\Gamma,1}, \ldots, \lambda_{\Gamma,n}\right)$$

a diagonal matrix containing the nonnegative eigenvalues. Then

$$\mathbf{\Gamma}^T\mathbf{\Gamma} = \mathbf{Q}\mathbf{\Lambda}^2\mathbf{Q}^T \ , \ \mathbf{\Lambda}^2 = \mathrm{diag}\left(\lambda_{\Gamma,1}^2, \ldots, \lambda_{\Gamma,n}^2\right)$$

and λ_i^2 are the eigenvalues of $\mathbf{\Gamma}^T\mathbf{\Gamma}$. Since λ_i are nonnegative and singular values of $\mathbf{\Gamma}$ equal the square-roots of the eigenvalues of $\mathbf{\Gamma}^T\mathbf{\Gamma}$, the first part is proved. For the second part, notice that

$$\mathbf{\Gamma}^{-1} = \mathbf{Q}\mathbf{\Lambda}^{-1}\mathbf{Q}^T \tag{B.25}$$

which is obtained from (B.24), and computation of $\mathbf{\Gamma}^{-1}\mathbf{\Gamma}$. Thereby $\mathbf{\Lambda}^{-1}$ is a diagonal matrix with eigenvalues of $\mathbf{\Gamma}^{-1}$ on its diagonal. It evaluates to

$$\mathbf{\Lambda}^{-1} = \mathrm{diag}\left(\lambda_{\Gamma,1}^{-1}, \ldots, \lambda_{\Gamma,n}^{-1}\right)$$

which proves the second part.

□

Theorem B.18 Singular Value Bounds for the 2-Norm of Linear Maps

For a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, $n \geq m$, and any $\mathbf{x} \in \mathbb{R}^m$

$$\underline{\sigma}_A \cdot \left\|\mathbf{x}\right\|_2 \leq \left\|\mathbf{A}\cdot\mathbf{x}\right\|_2 \leq \overline{\sigma}_A \cdot \left\|\mathbf{x}\right\|_2$$

if $n=m$ and $\mathbf{A}$ regular

$$\frac{\left\|\mathbf{x}\right\|_2}{\overline{\sigma}_A} \leq \left\|\mathbf{A}^{-1}\cdot\mathbf{x}\right\|_2 \leq \frac{\left\|\mathbf{x}\right\|_2}{\underline{\sigma}_A}$$

Proof

The singular value decomposition (SVD) of $\mathbf{A}$ is given by $\mathbf{A} = \mathbf{V}\cdot\mathbf{\Sigma}\cdot\mathbf{U}^T$ where $\mathbf{V} \in \mathbb{R}^{n\times n}, \mathbf{U} \in \mathbb{R}^{m\times m}$ are orthonormal matrices and $\mathbf{\Sigma} \in \mathbb{R}^{n\times m}$ is a diagonal type matrix containing the singular values. Then, with $\mathbf{A}^T\mathbf{A} = \mathbf{U}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{U}^T$, where

$$\mathbf{\Sigma}^T\mathbf{\Sigma} = \mathrm{diag}\left[\overline{\sigma}_A^2 \quad \cdots \quad \underline{\sigma}_A^2\right]$$

And hence

$$\left\|\mathbf{A}\mathbf{x}\right\|_2^2 = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} = \left(\mathbf{U}^T\mathbf{x}\right)^T\mathbf{\Sigma}^T\mathbf{\Sigma}\left(\mathbf{U}^T\mathbf{x}\right)$$

Since orthonormal matrices preserve the 2-norm of vectors ([Lüt96]), we also have:

$$\|\mathbf{x}\|_2 = \|\mathbf{U}^T\mathbf{x}\|_2$$

With definition $\mathbf{y} = \mathbf{U}^T\mathbf{x}$, we get

$$\|\mathbf{A}\mathbf{x}\|_2^2 = \mathbf{y}^T\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\mathbf{y} = \overline{\sigma}_A^2 y_1^2 + \ldots + \underline{\sigma}_A^2 y_m^2$$

which implies that

$$\underline{\sigma}_A\|\mathbf{x}\|_2 = \underline{\sigma}_A\|\mathbf{y}\|_2 \leq \|\mathbf{A}\mathbf{x}\|_2 = \sqrt{\mathbf{y}^T\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\mathbf{y}} \leq \overline{\sigma}_A\|\mathbf{y}\|_2 = \overline{\sigma}_A\|\mathbf{x}\|_2$$

and the first claim is proved. Further, if $\mathbf{A}$ is regular, it can be easily verified (by computation of $\mathbf{A}^{-1}\mathbf{A}$) that the SVD of the matrix inverse is

$$\mathbf{A}^{-1} = \mathbf{U} \cdot \boldsymbol{\Sigma}^{-1} \cdot \mathbf{V}^T$$

where $\boldsymbol{\Sigma}^{-1}$ is of diagonal form and the diagonal elements are the inverses of the singular values of $\mathbf{A}$. This implies

$$\overline{\sigma}_{A^{-1}} = \frac{1}{\underline{\sigma}_A}\ ,\ \underline{\sigma}_{A^{-1}} = \frac{1}{\overline{\sigma}_A}.$$

which proves the second claim.

□

Theorem B.19 Quadratic Forms for Non-Symmetric Matrices

For a quadratic matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$, and any $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = 0 \text{ and } \mathbf{x}^T\mathbf{B}\mathbf{x} = \mathbf{x}^T\mathbf{S}\mathbf{x}$$

where $\mathbf{S} = \frac{1}{2}\left(\mathbf{B}^T + \mathbf{B}\right)$ is the symmetric part of $\mathbf{B}$

and $\mathbf{A} = \frac{1}{2}\left(\mathbf{B} - \mathbf{B}^T\right)$ is the skew-symmetric part of $\mathbf{B}$

Proof

Every quadratic matrix can be separated into a symmetric and skew-symmetric part, such that

$$\mathbf{B} = \mathbf{A} + \mathbf{S}$$

where $\mathbf{S} = \mathbf{S}^T$ and $\mathbf{A} = -\mathbf{A}^T$. For the skew-symmetric part, consider the quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$. Since it is a scalar, it is equal to its transpose and hence

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{A}^T\mathbf{x} = -\mathbf{x}^T\mathbf{A}\mathbf{x}$$

which requires that $\mathbf{x}^T\mathbf{A}\mathbf{x} = 0$. This proves the first part. Further

$$\mathbf{x}^T\mathbf{B}\mathbf{x} = \mathbf{x}^T\left(\mathbf{S} + \mathbf{A}\right)\mathbf{x} = \mathbf{x}^T\mathbf{S}\mathbf{x} + \underbrace{\mathbf{x}^T\mathbf{A}\mathbf{x}}_{=0}$$

proves the second part.

□

Theorem B.20 Bounds on Quadratic Forms

Let $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$ and $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times n}$ symmetric and positive definite.

Then $\left|\mathbf{y}^T \boldsymbol{\Gamma} \mathbf{x}\right| \leq \overline{\lambda}_\Gamma \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$ and $\underline{\lambda}_\Gamma \|\mathbf{x}\|_2^2 \leq \mathbf{x}^T \boldsymbol{\Gamma} \mathbf{x} \leq \overline{\lambda}_\Gamma \|\mathbf{x}\|_2^2$

where $\underline{\lambda}_\Gamma$, $\overline{\lambda}_\Gamma$ denote the minimum and maximum eigenvalues of $\boldsymbol{\Gamma}$.

Proof

Since $\boldsymbol{\Gamma}$ is symmetric and positive definite, it has a full system of orthogonal eigenvectors, such that

$$\boldsymbol{\Gamma} = \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^T$$

where $\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix}$ contains the orthogonal eigenvectors and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \overline{\lambda}_\Gamma & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \underline{\lambda}_\Gamma \end{bmatrix}$$

contains the eigenvalues on its main diagonal in descending order. Further, it is known that multiplication of a vector by an orthonormal matrix preserves its Euclidean norm ([Lüt96]). Hence, we define $\mathbf{x}_q = \mathbf{Q}^T \mathbf{x}$, $\mathbf{y}_q = \mathbf{Q}^T \mathbf{y}$ and obtain:

first inequality:

$$\begin{aligned} \mathbf{x}^T \boldsymbol{\Gamma} \mathbf{y} &= \mathbf{x}_q^T \boldsymbol{\Lambda} \mathbf{y}_q = \overline{\lambda}_\Gamma x_{q1} y_{q1} + \ldots + \underline{\lambda}_\Gamma x_{qn} y_{qn} \leq \overline{\lambda}_\Gamma (x) \left(x_{q1} y_{q1} + \ldots + x_{qn} y_{qn} \right) \\ &= \overline{\lambda}_\Gamma \mathbf{x}_q^T \mathbf{y}_q \leq \overline{\lambda}_\Gamma \|\mathbf{x}_q\|_2 \|\mathbf{y}_q\|_2 = \overline{\lambda}_\Gamma \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \end{aligned}$$

second inequality:

$$\begin{aligned} \mathbf{x}^T \boldsymbol{\Gamma} \mathbf{x} &= \mathbf{x}_q^T \boldsymbol{\Lambda} \mathbf{x}_q = \overline{\lambda}_\Gamma x_{q1}^2 + \ldots + \underline{\lambda}_\Gamma x_{q2}^2 \leq \overline{\lambda}_\Gamma \left(x_{q1}^2 + \ldots + x_{qn}^2 \right) \\ &= \overline{\lambda}_\Gamma \mathbf{x}_q^T \mathbf{x}_q = \overline{\lambda}_\Gamma \|\mathbf{x}_q\|_2^2 = \overline{\lambda}_\Gamma \|\mathbf{x}\|_2^2 \end{aligned}$$

$$\begin{aligned} \mathbf{x}^T \boldsymbol{\Gamma} \mathbf{x} &= \mathbf{x}_q^T \boldsymbol{\Lambda} \mathbf{x}_q = \overline{\lambda}_\Gamma x_{q1}^2 + \ldots + \underline{\lambda}_\Gamma x_{q2}^2 \geq \underline{\lambda}_\Gamma \left(x_{q1}^2 + \ldots + x_{qn}^2 \right) \\ &= \underline{\lambda}_\Gamma \mathbf{x}_q^T \mathbf{x}_q = \underline{\lambda}_\Gamma \|\mathbf{x}_q\|_2^2 = \underline{\lambda}_\Gamma \|\mathbf{x}\|_2^2 \end{aligned}$$

□

Theorem B.21 Bound on Symmetric Trace Expression

Let $\boldsymbol{\Theta} \in \mathbb{R}^{n\times m}$, $\boldsymbol{\Gamma} \in \mathbb{R}^{n\times n}$ symmetric and positive definite.

Then $\underline{\lambda}_\Gamma \left\|\boldsymbol{\Theta}\right\|_F^2 \leq \mathrm{tr}\left(\boldsymbol{\Theta}^T \boldsymbol{\Gamma} \boldsymbol{\Theta}\right) \leq \overline{\lambda}_\Gamma \left\|\boldsymbol{\Theta}\right\|_F^2$

where $\underline{\lambda}_\Gamma, \overline{\lambda}_\Gamma$: minimum and maximum eigenvalues of $\boldsymbol{\Gamma}$

Proof

Since $\boldsymbol{\Gamma}$ is symmetric and positive definite, it has a full system of orthogonal eigenvectors, such that

$$\boldsymbol{\Gamma} = \mathbf{Q} \boldsymbol{\Lambda} \mathbf{Q}^T$$

where $\mathbf{Q} = \begin{bmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_n \end{bmatrix}$ contains the orthogonal eigenvectors and

$$\boldsymbol{\Lambda} = \begin{bmatrix} \overline{\lambda}_\Gamma & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \underline{\lambda}_\Gamma \end{bmatrix}$$

contains the eigenvalues on its main diagonal in descending order. Let the columns of $\boldsymbol{\Theta}$ be denoted with $\boldsymbol{\theta}_i$, $i = 1, \ldots, m$ and define

$$\hat{\boldsymbol{\Theta}} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_1 & \cdots & \hat{\boldsymbol{\theta}}_m \end{bmatrix} = \begin{bmatrix} \mathbf{Q}^T \boldsymbol{\theta}_1 & \cdots & \mathbf{Q}^T \boldsymbol{\theta}_m \end{bmatrix}.$$

Then

$$\mathrm{tr}\left(\boldsymbol{\Theta}^T \boldsymbol{\Gamma} \boldsymbol{\Theta}\right) = \mathrm{tr}\left(\hat{\boldsymbol{\Theta}}^T \boldsymbol{\Lambda} \hat{\boldsymbol{\Theta}}\right) = \overline{\lambda}_\Gamma \hat{\boldsymbol{\theta}}_1^T \hat{\boldsymbol{\theta}}_1 + \ldots + \underline{\lambda}_\Gamma \hat{\boldsymbol{\theta}}_m^T \hat{\boldsymbol{\theta}}_m$$

and an upper and lower for this expression is given by

$$\underline{\lambda}_\Gamma \underbrace{\left(\hat{\boldsymbol{\theta}}_1^T \hat{\boldsymbol{\theta}}_1 + \ldots + \hat{\boldsymbol{\theta}}_m^T \hat{\boldsymbol{\theta}}_m\right)}_{*} \leq \mathrm{tr}\left(\boldsymbol{\Theta}^T \boldsymbol{\Gamma} \boldsymbol{\Theta}\right) \leq \overline{\lambda}_\Gamma \underbrace{\left(\hat{\boldsymbol{\theta}}_1^T \hat{\boldsymbol{\theta}}_1 + \ldots + \hat{\boldsymbol{\theta}}_m^T \hat{\boldsymbol{\theta}}_m\right)}_{*}. \tag{B.26}$$

The term in brackets (*) is the sum of squares of all elements of $\hat{\boldsymbol{\Theta}}$ and hence equal to the square of the Frobenius norm.

$$\left(\hat{\boldsymbol{\theta}}_1^T \hat{\boldsymbol{\theta}}_1 + \ldots + \hat{\boldsymbol{\theta}}_m^T \hat{\boldsymbol{\theta}}_m\right) = \left\|\hat{\boldsymbol{\Theta}}\right\|_F^2 = \mathrm{tr}\left(\hat{\boldsymbol{\Theta}}^T \hat{\boldsymbol{\Theta}}\right)$$

Back substitution and the fact that $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}$ yields

$$\mathrm{tr}\left(\hat{\boldsymbol{\Theta}}^T \hat{\boldsymbol{\Theta}}\right) = \mathrm{tr}\left(\boldsymbol{\Theta}^T \mathbf{Q}\mathbf{Q}^T \boldsymbol{\Theta}\right) = \mathrm{tr}\left(\boldsymbol{\Theta}^T \boldsymbol{\Theta}\right) = \left\|\boldsymbol{\Theta}\right\|_F^2$$

and using this result in (B.26) finally proves the claim.

□

Theorem B.22 Regularity of Sum of Matrices

Let some matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$, and $\mathbf{B} \in \mathbb{R}^{n,xn}$ and

$\overline{\sigma}_A$: maximum singular value of $\mathbf{A}$

$\underline{\sigma}_B$ minimum singular values of $\mathbf{B}$.

If $\overline{\sigma}_A < \underline{\sigma}_B$, then

$\mathbf{A}+\mathbf{B}$ is regular

Proof

If $\mathbf{C} = \mathbf{A} + \mathbf{B}$ is singular, then it has a non-trivial null space, i.e. there is some $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{Cx} = \mathbf{0}$ and its quadratic form $\mathbf{x}^T\mathbf{Cx} = 0$. If it can be shown, that there is no such $\mathbf{x}$, then matrix is regular. FromTheorem B.18 we have

$$\|\mathbf{Ax}\|_2 \leq \overline{\sigma}_A \|\mathbf{x}\|_2 \text{ and } \|\mathbf{Bx}\|_2 \geq \underline{\sigma}_B \|\mathbf{x}\|_2$$

which implies

$$\mathbf{x}^T\mathbf{Cx} = \mathbf{x}^T\mathbf{Ax} + \mathbf{x}^T\mathbf{Bx} \geq \underline{\sigma}_B \|\mathbf{x}\|_2 - \overline{\sigma}_A \|\mathbf{x}\|_2 = (\underline{\sigma}_B - \overline{\sigma}_A)\|\mathbf{x}\|_2 > 0$$

and the proof is complete.

□

B.6 Introduction to Quaternions

The drawback of a description of the aircraft attitude by means of 3 Euler angles is the singularity for pitch angle $\Theta = \pm\pi/2$. A method to circumvent this is the utilization of quaternions, which were first introduced by W.R. Hamilton in the 19th century([Ste03]). A comprehensive introduction to quaternions is given in [Ste03].

B.6.1 Mathematical Point of View

Quaternions are a generalization of the complex numbers where the imaginary part is extended to 3 dimensions. So a quaternion is typically of the form

$$\underline{\mathbf{q}} = q_0 + q_1 \cdot \mathbf{i} + q_2 \cdot \mathbf{j} + q_3 \cdot \mathbf{k}$$

where $\mathbf{i},\mathbf{j},\mathbf{k}$ are unit vectors for the 3 dimensional imaginary part and

$$\begin{aligned} \mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{i}\cdot\mathbf{j}\cdot\mathbf{k} = -1 \quad , \\ \mathbf{i}\cdot\mathbf{j} = -\mathbf{j}\cdot\mathbf{i} = \mathbf{k} \quad , \quad \mathbf{j}\cdot\mathbf{k} = -\mathbf{k}\cdot\mathbf{j} = \mathbf{i} \quad , \quad \mathbf{k}\cdot\mathbf{i} = -\mathbf{i}\cdot\mathbf{k} = \mathbf{j} \end{aligned} \tag{B.27}$$

q_0 is the scalar part and $\mathbf{q}^T = (q_1 \quad q_2 \quad q_3)$ is referred to as vector part of the quaternion. A quaternion is also compactly written as vector in $\mathbb{R}^4$.

$$\underline{\mathbf{q}} = \begin{pmatrix} q_0 \\ \mathbf{q} \end{pmatrix}$$

As complex numbers, also a quaternion has a conjugate element:

$$\underline{\mathbf{q}}^* = \begin{pmatrix} q_0 \\ -\mathbf{q} \end{pmatrix}$$

While quaternion summation is equal to addition of vectors, the quaternion multiplication obeys some special principles.

Norm of a Quaternion

The norm of a quaternion is defined as

$$\left\| \underline{\mathbf{q}} \right\| = \underline{\mathbf{q}}^* * \underline{\mathbf{q}} = q_0^2 + q_1^2 + q_2^2 + q_3^2 \tag{B.28}$$

where the quaternion norm is not to be mixed up with the 2-norm of an ordinary vector in $\mathbb{R}^4$. It is in fact the square of the 2-norm.

Quaternion Multiplication

Multiplication of quaternions is carried out by accounting for properties (B.27). Let $\mathbf{p}$, $\mathbf{q}$ be quaternions, then

$$\underline{\mathbf{p}} * \underline{\mathbf{q}} = \begin{pmatrix} p_0 q_0 - \mathbf{p}^T \cdot \mathbf{q} \\ p_0 \mathbf{q} + q_0 \mathbf{p} + \mathbf{p} \times \mathbf{q} \end{pmatrix} \tag{B.29}$$

where "$*$" denotes quaternion multiplication and "$\times$" denotes the vector (cross) product in $\mathbb{R}^3$.

Neutral Element

Considering (B.29), the neutral element w.r.t. quaternion multiplication is

$$\underline{\mathbf{I}} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix} \tag{B.30}$$

Inverse Element

In light of (B.29) and (B.30) inverse element $\mathbf{q}^{-1}$ is obtained by

$$\underline{\mathbf{q}}^{-1} = \frac{\underline{\mathbf{q}}^*}{\left\| \underline{\mathbf{q}} \right\|} = \frac{1}{q_0^1 + q_1^2 + q_2^2 + q_3^2} \cdot \begin{pmatrix} q_0 \\ -\mathbf{q} \end{pmatrix} \tag{B.31}$$

Quaternion Properties

Let

$$\underline{\mathbf{p}} = \begin{pmatrix} p_0 \\ \mathbf{p} \end{pmatrix}, \ \underline{\mathbf{q}} = \begin{pmatrix} q_0 \\ \mathbf{q} \end{pmatrix}, \ \underline{\mathbf{r}} = \begin{pmatrix} r_0 \\ \mathbf{r} \end{pmatrix}$$

be quaternions. Then the following properties, concerning the quaternion product, hold.

Non Commutativity

$$\underline{\mathbf{p}} * \underline{\mathbf{q}} - \underline{\mathbf{q}} * \underline{\mathbf{p}} = \begin{pmatrix} 0 \\ 2 \cdot \mathbf{p} \times \mathbf{q} \end{pmatrix}$$

Associativity

$$\underline{\mathbf{p}} * (\underline{\mathbf{q}} * \underline{\mathbf{r}}) = (\underline{\mathbf{p}} * \underline{\mathbf{q}}) * \underline{\mathbf{r}}$$

Norm

$$\left\| \underline{\mathbf{p}} * \underline{\mathbf{q}} \right\| = \left\| \underline{\mathbf{p}} \right\| \cdot \left\| \underline{\mathbf{q}} \right\|$$

Inverse

$$(\underline{\mathbf{p}} * \underline{\mathbf{q}})^{-1} = \underline{\mathbf{q}}^{-1} * \underline{\mathbf{p}}^{-1}$$

B.6.2 Some Useful Syntax

Some notations were introduced [Jac92], which allow a compact and well-arranged presentation of operations involving quaternions. In a slight modified version, notations that appeared to be useful are stated here.

To a quaternion

$$\underline{\mathbf{q}} = \begin{pmatrix} q_0 \\ \mathbf{q} \end{pmatrix} \quad , \quad \mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}$$

we define

$$\overset{+}{\mathbf{Q}} = \begin{bmatrix} q_0 & -\mathbf{q}^T \\ \mathbf{q} & q_0 \mathbf{I} + \tilde{\mathbf{q}} \end{bmatrix} \quad , \quad \overset{-}{\mathbf{Q}} = \begin{bmatrix} q_0 & -\mathbf{q}^T \\ \mathbf{q} & q_0 \mathbf{I} - \tilde{\mathbf{q}} \end{bmatrix} \tag{B.32}$$

where

$$\tilde{\mathbf{q}} = \begin{bmatrix} 0 & -q_3 & q_2 \\ q_3 & 0 & -q_1 \\ -q_2 & q_1 & 0 \end{bmatrix}$$

is the skew symmetric cross product matrix associated with the vector part and $\mathbf{I}$ denotes the identity matrix in $\mathbb{R}^3$. The respective matrices to the conjugate quaternion is defined accordingly

$$\overset{+}{\mathbf{Q}}^{*} = \begin{bmatrix} q_0 & \mathbf{q}^T \\ -\mathbf{q} & q_0\mathbf{I} - \tilde{\mathbf{q}} \end{bmatrix} \quad , \quad \overset{-}{\mathbf{Q}}^{*} = \begin{bmatrix} q_0 & \mathbf{q}^T \\ -\mathbf{q} & q_0\mathbf{I} + \tilde{\mathbf{q}} \end{bmatrix} \tag{B.33}$$

and it follows that

$$\overset{+}{\mathbf{Q}}^{*} = \overset{+}{\mathbf{Q}}^{T} \quad , \quad \overset{-}{\mathbf{Q}}^{*} = \overset{-}{\mathbf{Q}}^{T} \tag{B.34}$$

It also reveals that $\overset{+}{\mathbf{Q}}$ and $\overset{-}{\mathbf{Q}}$ as well as their conjugates are orthonormal matrices if $\underline{\mathbf{q}}$ is a unit quaternion. This can be verified by computation of

$$\overset{+}{\mathbf{Q}}^{T} \cdot \overset{+}{\mathbf{Q}} = \begin{bmatrix} q_0^2 + \mathbf{q}^T \cdot \mathbf{q} & -q_0 \cdot \mathbf{q} + \mathbf{q}^T (q_0\mathbf{I} + \tilde{\mathbf{q}}) \\ -q_0\mathbf{q} + (q_0\mathbf{I} - \tilde{\mathbf{q}})\mathbf{q} & \mathbf{q}\mathbf{q}^T + (q_0\mathbf{I} - \tilde{\mathbf{q}})(q_0\mathbf{I} + \tilde{\mathbf{q}}) \end{bmatrix} = \overbrace{\|\underline{\mathbf{q}}\|}^{=1} \cdot \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \tag{B.35}$$

With these definitions, the quaternion product can be written as matrix vector product. Let $\underline{\mathbf{p}}$ be another quaternion and $\overset{+}{\mathbf{P}}$, $\overset{-}{\mathbf{P}}$ according to (B.32), then

$$\underline{\mathbf{p}} * \underline{\mathbf{q}} = \overset{+}{\mathbf{P}} \cdot \underline{\mathbf{q}} \quad , \quad \underline{\mathbf{q}} * \underline{\mathbf{p}} = \overset{-}{\mathbf{P}} \cdot \underline{\mathbf{q}} \tag{B.36}$$

and the non-commutativity of quaternion product is simply resolved by changing the sign of $\tilde{\mathbf{p}}$. We further define matrices

$$\overset{+}{\mathbf{E}}_q = \begin{bmatrix} -\mathbf{q} & q_0\mathbf{I} + \tilde{\mathbf{q}} \end{bmatrix} \quad , \quad \overset{-}{\mathbf{E}}_q = \begin{bmatrix} -\mathbf{q} & q_0\mathbf{I} - \tilde{\mathbf{q}} \end{bmatrix} \tag{B.37}$$

and their conjugate associates

$$\overset{+}{\mathbf{E}}_{q^*} = \begin{bmatrix} \mathbf{q} & q_0\mathbf{I} - \tilde{\mathbf{q}} \end{bmatrix} \quad , \quad \overset{-}{\mathbf{E}}_{q^*} = \begin{bmatrix} \mathbf{q} & q_0\mathbf{I} + \tilde{\mathbf{q}} \end{bmatrix} \tag{B.38}$$

Then (B.32) also reads as

$$\overset{+}{\mathbf{Q}} = \begin{bmatrix} \underline{\mathbf{q}} & \overset{-}{\mathbf{E}}_q^T \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{q}}^{*T} \\ \overset{-}{\mathbf{E}}_{q^*} \end{bmatrix} \quad , \quad \overset{-}{\mathbf{Q}} = \begin{bmatrix} \underline{\mathbf{q}} & \overset{+}{\mathbf{E}}_q^T \end{bmatrix} = \begin{bmatrix} \underline{\mathbf{q}}^{*T} \\ \overset{+}{\mathbf{E}}_{q^*} \end{bmatrix} \tag{B.39}$$

Note that orthonormal property of $\overset{+}{\mathbf{Q}}_q$, $\overset{-}{\mathbf{Q}}_q$, $\overset{+}{\mathbf{Q}}_{q^*}$, $\overset{-}{\mathbf{Q}}_{q^*}$ and (B.39) imply that in each case the rows of $\overset{+}{\mathbf{E}}_q$, $\overset{-}{\mathbf{E}}_q$, $\overset{+}{\mathbf{E}}_{q^*}$ and $\overset{-}{\mathbf{E}}_{q^*}$ are orthogonal to each other. Additionally $\underline{\mathbf{q}}$ is orthogonal to each row of $\overset{+}{\mathbf{E}}_q$ and $\overset{-}{\mathbf{E}}_q$ and $\underline{\mathbf{q}}^*$ is orthogonal to each row of $\overset{+}{\mathbf{E}}_{q^*}$ and $\overset{-}{\mathbf{E}}_{q^*}$. Particularly, the following equations hold:

$$\overset{+}{\mathbf{E}}_q \cdot \overset{+}{\mathbf{E}}_q^T = \mathbf{I} \quad , \quad \overset{-}{\mathbf{E}}_q \cdot \overset{-}{\mathbf{E}}_q^T = \mathbf{I} \tag{B.40}$$

$$\overset{+}{\mathbf{E}}_{q^*} \cdot \overset{+}{\mathbf{E}}_{q^*}^T = \mathbf{I} \quad , \quad \overset{-}{\mathbf{E}}_{q^*} \cdot \overset{-}{\mathbf{E}}_{q^*}^T = \mathbf{I} \tag{B.41}$$

$$\overset{+}{\mathbf{E}}_q \cdot \underline{\mathbf{q}} = \mathbf{0} \quad , \quad \overset{-}{\mathbf{E}}_q \cdot \underline{\mathbf{q}} = \mathbf{0} \tag{B.42}$$

$$\overset{+}{\mathbf{E}}_{q^*} \cdot \underline{\mathbf{q}}^* = \mathbf{0} \quad , \quad \overset{-}{\mathbf{E}}_{q^*} \cdot \underline{\mathbf{q}}^* = \mathbf{0} \tag{B.43}$$

B.6.3 Quaternions and Rotations in Euclidean Space

In the field of engineering, quaternions are often utilized for representation of vector rotations in Euclidean space, where the quaternion vector part represents the direction of the rotation axis while scalar part represents the angle of rotation.

Starting with Goldstein's formula for vector rotation in Euclidean space ([Gol01]), the resulting expression is compared with a certain type of quaternion product and it will be seen, that both are equal.

Rotation of a Vector

Figure B.4 sketches a left hand rotation of the vector $\vec{\mathbf{u}}$ onto the vector $\vec{\mathbf{v}}$ by some angle μ. Thereby $\vec{\mathbf{n}}$ is a unit length vector pointing into the direction of the rotation axis and the disc E indicates the plane of rotation, perpendicular to $\vec{\mathbf{n}}$. Then, from Figure B.4, we derive the Goldstein's rotation Formula

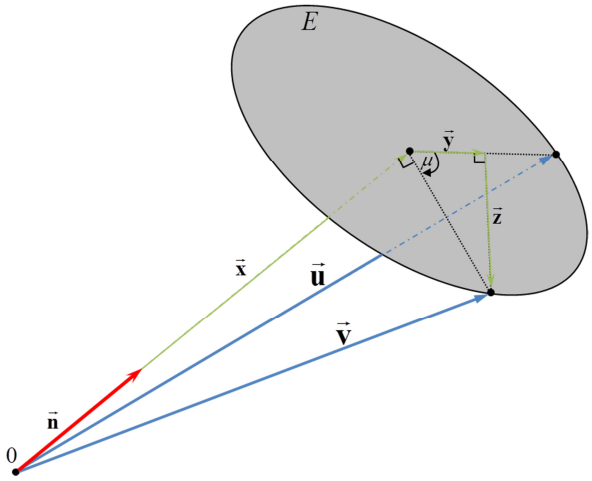


Figure B.4 Left Hand Vector Rotation

$$\begin{aligned} \vec{\mathbf{v}} &= \vec{\mathbf{x}} + \vec{\mathbf{y}} + \mathbf{z} \\ &= \vec{\mathbf{n}} \cdot \vec{\mathbf{n}}^T \cdot \vec{\mathbf{u}} + \left(\vec{\mathbf{u}} - \vec{\mathbf{n}} \cdot \vec{\mathbf{n}}^T \cdot \vec{\mathbf{u}}\right)\cos\mu - \left(\vec{\mathbf{n}} \times \vec{\mathbf{u}}\right)\sin\mu \\ \vec{\mathbf{v}} &= \left(1 - \cos\mu\right) \cdot \vec{\mathbf{n}} \cdot \vec{\mathbf{n}}^T \cdot \vec{\mathbf{u}} + \vec{\mathbf{u}} \cdot \cos\mu - \left(\vec{\mathbf{n}} \times \vec{\mathbf{u}}\right) \cdot \sin\mu \end{aligned} \tag{B.44}$$

With knowledge of rotation axis and rotation angle, some arbitrary vector $\vec{\mathbf{u}}$ is rotated in a left hand sense by (B.44).

Quaternion Rotation

Vector rotations can also be described by means of *unit quaternion*

$$\underline{\mathbf{q}} = \begin{pmatrix} q_0 \\ \mathbf{q} \end{pmatrix}$$

which is a quaternion of unity length, i.e. $\left\| \underline{\mathbf{q}} \right\| = 1$.

In order to show, how a vector $\vec{\mathbf{u}}$ is rotate, utilizing quaternion products, we at first have to embed $\vec{\mathbf{u}}$ into quaternion context. This is done by defining

$$\underline{\mathbf{u}} = \begin{pmatrix} 0 \\ \vec{\mathbf{u}} \end{pmatrix} \tag{B.45}$$

i.e. the associated quaternion is obtained by inserting $\vec{\mathbf{u}}$ into the vector part of the quaternion and setting the scalar quaternion part to 0. The following quaternion product

$$\underline{\mathbf{q}}^{-1} * \underline{\mathbf{u}} * \underline{\mathbf{q}} = \begin{pmatrix} -q_0 \cdot \vec{\mathbf{u}}^T \cdot \mathbf{q} + \mathbf{q}^T \cdot (q_0 \cdot \vec{\mathbf{u}} + \vec{\mathbf{u}} \times \mathbf{q}) \\ \mathbf{q} \cdot \vec{\mathbf{u}}^T \cdot \mathbf{q} + q_0 \cdot (q_0 \cdot \vec{\mathbf{u}} + \vec{\mathbf{u}} \times \mathbf{q}) - \mathbf{q} \times (q_0 \cdot \vec{\mathbf{u}} + \vec{\mathbf{u}} \times \mathbf{q}) \end{pmatrix}$$

reduces to

$$\underline{\mathbf{q}}^{-1} * \underline{\mathbf{u}} * \underline{\mathbf{q}} = \begin{pmatrix} 0 \\ 2\mathbf{q} \cdot \mathbf{q}^T \cdot \vec{\mathbf{u}} + \mathbf{u} \cdot \left(q_0^2 - \mathbf{q}^T \cdot \mathbf{q} \right) - 2q_0 \cdot \mathbf{q} \times \vec{\mathbf{u}} \end{pmatrix} \tag{B.46}$$

where Grassmann's identity has been used for the double cross-product term. Comparing (B.44) and (B.46), it can be concluded that

$$\mathbf{q} = \vec{\mathbf{n}} \cdot \sin\left(\mu/2\right) \text{ and } q_0 = \cos\left(\mu/2\right) \tag{B.47}$$

make both expressions equal. This becomes clear by recalling the double angle formulas

$$\sin\left(\mu/2\right) = \sqrt{\frac{1}{2}\left(1 - \cos\mu\right)} \text{ and } \cos\left(\mu/2\right) = \sqrt{\frac{1}{2}\left(1 + \cos\mu\right)}$$

and inserting the latter into the vector part of (B.46):

$$2\left(\sqrt{\frac{1}{2}(1-\cos\mu)}\right)^2\cdot\vec{\mathbf{n}}\cdot\vec{\mathbf{n}}^T\cdot\mathbf{u}+\vec{\mathbf{u}}\cdot\left(\frac{1}{2}(1+\cos\mu)-\frac{1}{2}(1-\cos\mu)\right)\vec{\mathbf{n}}^T\cdot\vec{\mathbf{n}}-\sqrt{\frac{1}{2}(1-\cos\mu)}\cdot\sqrt{\frac{1}{2}(1+\cos\mu)}\cdot\vec{\mathbf{n}}\times\vec{\mathbf{u}}$$

$$=(1-\cos\mu)\cdot\vec{\mathbf{n}}\cdot\vec{\mathbf{n}}^T\cdot\vec{\mathbf{u}}+\vec{\mathbf{u}}\cdot\cos\mu-\sqrt{1-\cos^2\mu}\cdot\vec{\mathbf{n}}\times\vec{\mathbf{u}}$$

$$=(1-\cos\mu)\cdot\vec{\mathbf{n}}\cdot\vec{\mathbf{n}}^T\cdot\vec{\mathbf{u}}+\vec{\mathbf{u}}\cdot\cos\mu-\sin\mu\cdot\vec{\mathbf{n}}\times\vec{\mathbf{u}}$$

and obviously, the resulting expression equals (B.44). Summing up a quaternion, describing a left hand rotation around axis $\vec{\mathbf{n}}$ and angle μ,is of the form

$$\underline{\mathbf{q}}=\begin{pmatrix}\cos(\mu/2)\\ \vec{\mathbf{n}}\cdot\sin(\mu/2)\end{pmatrix} \tag{B.48}$$

where $\vec{\mathbf{n}}$ has unity length and hence also $\|\underline{\mathbf{q}}\|=\cos^2(\mu/2)+\sin^2(\mu/2)=1$. It is worth to be noticed that $\underline{\mathbf{q}}$ and $-\underline{\mathbf{q}}$ describe the same rotation, which is clear from (B.46).

Vector Rotation and Coordinate Transformation

As we have seen, Goldstein's rotation formula (B.44) as well as the quaternion product (B.46) represent a left-handed rotation of $\vec{\mathbf{u}}$ onto $\vec{\mathbf{v}}$. This also can be interpreted equivalently as a right-handed rotation of the coordinate frame. Let the original system be denoted with index 0, the rotated coordinate frame with index B and a rotation matrix from 0 to B frame $\mathbf{M}_{BO}$. Then it is well-known that a coordinate transformation $0\rightarrow B$ is obtained by

$$(\vec{\mathbf{u}})_B=\mathbf{M}_{BO}\cdot(\vec{\mathbf{u}})_O \tag{B.49}$$

By the results, derived so far, there is an associated quaternion that represents the same coordinate transformation, which will henceforth be denoted as $\underline{\mathbf{q}}_{BO}$.

$$\underline{\mathbf{u}}_B=\underline{\mathbf{q}}_{BO}^{-1}*\underline{\mathbf{u}}_O*\underline{\mathbf{q}}_{BO} \tag{B.50}$$

where $\underline{\mathbf{u}}_O$, $\underline{\mathbf{u}}_B$ are the quaternions associated with the vectors $(\vec{\mathbf{u}})_O$, $(\vec{\mathbf{u}})_B$ according to (B.45). This operation is of course also invertible by solving (B.50) for $\underline{\mathbf{u}}_O$.

$$\begin{aligned}\underline{\mathbf{u}}_O&=\underline{\mathbf{q}}_{BO}*\underline{\mathbf{u}}_B*\underline{\mathbf{q}}_{BO}^{-1}\\ &=\underline{\mathbf{q}}_{OB}^{-1}*\underline{\mathbf{u}}_B*\underline{\mathbf{q}}_{OB}\end{aligned}$$

and the quaternion, associated with the inverse coordinate transformation, is

$$\underline{\mathbf{q}}_{OB}=\underline{\mathbf{q}}_{BO}^{-1}. \tag{B.51}$$

If $\underline{\mathbf{u}}_B$ is further transformed to a frame C with a quaternion $\underline{\mathbf{q}}_{CB}$, the associated operation is

$$\underline{\mathbf{u}}_C = \underline{\mathbf{q}}_{CB}^{-1} * \underline{\mathbf{u}}_B * \underline{\mathbf{q}}_{CB} = \underline{\mathbf{q}}_{CB}^{-1} * \underline{\mathbf{q}}_{BO}^{-1} * \underline{\mathbf{u}}_O * \underline{\mathbf{q}}_{BO} * \underline{\mathbf{q}}_{CB}$$

and consequently, the quaternion for an iterated transformation $O \to B \to C$ is

$$\underline{\mathbf{q}}_{CO} = \underline{\mathbf{q}}_{BO} * \underline{\mathbf{q}}_{CB} \tag{B.52}$$

Using definitions of section B.6.2,the quaternion rotation operation (B.50) can also be written in a compact matrix vector notation. With (B.34), (B.36) and (B.39), we obtain

$$\begin{pmatrix} 0 \\ \mathbf{u}_B \end{pmatrix} = \underline{\mathbf{q}}_{BO}^{-1} * \underline{\mathbf{u}}_A * \underline{\mathbf{q}}_{BO} = \overset{+}{\mathbf{Q}}{}^T_{q_{BO}} \cdot \overset{-}{\mathbf{Q}}_{q_{BO}} \cdot \begin{pmatrix} 0 \\ \mathbf{u}_O \end{pmatrix}$$

$$= \begin{bmatrix} \underline{\mathbf{q}}_{BO}^T \\ \overset{-}{\mathbf{E}}_{q_{BO}} \end{bmatrix} \cdot \begin{bmatrix} \underline{\mathbf{q}}_{BO} & \overset{+}{\mathbf{E}}{}^T_{q_{BO}} \end{bmatrix} \cdot \begin{pmatrix} 0 \\ \mathbf{u}_O \end{pmatrix} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \overset{-}{\mathbf{E}}_{q_{BO}} \overset{+}{\mathbf{E}}{}^T_{q_{BO}} \end{bmatrix} \cdot \begin{pmatrix} 0 \\ \mathbf{u}_O \end{pmatrix}$$

The scalar part automatically results to zero while we obtain for the vector part:

$$\mathbf{u}_B = \overset{-}{\mathbf{E}}_{q_{BO}} \cdot \overset{+}{\mathbf{E}}{}^T_{q_{BO}} \cdot \mathbf{u}_A \tag{B.53}$$

Euler's Rotation Theorem

It is intuitively clear and affirmed by Euler's rotation theorem ([Ste03]), if a rigid body is arbitrarily rotated about some fixed point, the change of attitude can also be obtained by a rotation about a single axis. This statement is compliant with the fact that rotation matrices in $\mathbb{R}^3$ have exactly one invariant direction - the axis of rotation. This is clear by the well-known fact that nontrivial rotation matrices in $\mathbb{R}^3$ have exactly one eigenvalue that equals to *1* (representing the rotation axis) and pair of conjugate complex eigenvalues (representing the rotation)([Ste03]). This implies that the rotation axis vector $\vec{\mathbf{n}}$ of a quaternion $\underline{\mathbf{q}}_{BO}$, representing a coordinate transformation $0 \to B$ is equal in both frames.

Another consequence of Euler's theorem is, that a change in attitude, obtained by consecutive rotations about coordinate frame axes (as is e.g. the case using Euler angles) is equivalently obtained by a rotation about a single rotation axis (which is in fact the quaternion vector part or the eigenvector to eigenvalue *1* of the associated transformation matrix).

Quaternion and Euler Angles

The aircraft attitude relative to the earth surface (NED-frame) is described by 3 consecutive rotations of the Euler angles: z-axis: Ψ, y-axis: Θ, x-axis Φ (refer to Appendix A). Each elementary rotation is associated with a quaternion

$$\underline{\mathbf{q}}_{\Psi} = \begin{pmatrix} \cos(\Psi/2) \\ 0 \\ 0 \\ \sin(\Psi/2) \end{pmatrix} \quad , \quad \underline{\mathbf{q}}_{\Theta} = \begin{pmatrix} \cos(\Theta/2) \\ 0 \\ \sin(\Theta/2) \\ 0 \end{pmatrix} \quad , \quad \underline{\mathbf{q}}_{\Phi} = \begin{pmatrix} \cos(\Phi/2) \\ \sin(\Phi/2) \\ 0 \\ 0 \end{pmatrix}$$

and the whole rotation obtained by

$$\underline{\mathbf{q}}_{E} = \underline{\mathbf{q}}_{\Psi} * \underline{\mathbf{q}}_{\Theta} * \underline{\mathbf{q}}_{\Phi}$$

which leads the following relation between Euler angles and associated quaternions:

$$\begin{aligned} q_{E,0} &= \pm[\cos(\Phi/2)\cos(\Theta/2)\cos(\Psi/2) + \sin(\Phi/2)\sin(\Theta/2)\sin(\Psi/2)] \\ q_{E,1} &= \pm[\sin(\Phi/2)\cos(\Theta/2)\cos(\Psi/2) - \cos(\Phi/2)\sin(\Theta/2)\sin(\Psi/2)] \\ q_{E,2} &= \pm[\cos(\Phi/2)\sin(\Theta/2)\cos(\Psi/2) + \sin(\Phi/2)\cos(\Theta/2)\sin(\Psi/2)] \\ q_{E,3} &= \pm[\cos(\Phi/2)\cos(\Theta/2)\sin(\Psi/2) - \sin(\Phi/2)\sin(\Theta/2)\cos(\Psi/2)] \end{aligned} \tag{B.54}$$

The $\pm$ sign is justified since $\underline{\mathbf{q}}$ and $-\underline{\mathbf{q}}$ describe the same rotation.

Quaternion and Transformation Matrix

A coordinate transformation between frames O and B is equivalently represented by a unit quaternion $\mathbf{q}_{BO}$ and a transformation matrix $\mathbf{M}_{BO}$. In order to derive the relationship between the two equivalent representations, consider the quaternion rotation operation in matrix vector notation (B.53), which reads, when expanded

$$\mathbf{M}_{BO} = \bar{\mathbf{E}}^{T}_{q_{BO}} \cdot \mathbf{E}^{+}_{q_{BO}} = [-\mathbf{q} \quad q_0\mathbf{I} - \tilde{\mathbf{q}}] \cdot \begin{bmatrix} -\mathbf{q}^T \\ q_0\mathbf{I} - \tilde{\mathbf{q}} \end{bmatrix} = \mathbf{q} \cdot \mathbf{q}^T + (q_0\mathbf{I} - \tilde{\mathbf{q}}) \cdot (q_0\mathbf{I} - \tilde{\mathbf{q}})$$

and we obtain the following relationship by further scalar expansion of the quaternion.

$$\mathbf{M}_{BO} = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 + q_0q_3) & 2(q_1q_3 - q_0q_2) \\ 2(q_1q_2 - q_0q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2q_3 + q_0q_1) \\ 2(q_1q_3 + q_0q_2) & 2(q_2q_3 - q_0q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \tag{B.55}$$

Let

$$\mathbf{M}_{BO} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}$$

be a scalar expansion of the rotation matrix. Then a transformation from rotation matrix to unit quaternion is obtained, by solving the 4 equations

$$q_0^2 + q_1^2 - q_2^2 - q_3^2 = m_{11}$$
$$q_0^2 - q_1^2 + q_2^2 - q_3^2 = m_{22}$$
$$q_0^2 - q_1^2 - q_2^2 + q_3^2 = m_{33}$$
$$q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$$

which are obtained from the 3 diagonal entries of the rotation matrix and the unit quaternion condition, for the quaternion.

$$\begin{aligned} q_0 &= \pm\frac{1}{4}\sqrt{1 + m_{11} + m_{22} + m_{33}} \\ q_1 &= \pm\frac{1}{4}\sqrt{1 + m_{11} - m_{22} - m_{33}} \\ q_2 &= \pm\frac{1}{4}\sqrt{1 - m_{11} + m_{22} - m_{33}} \\ q_3 &= \pm\frac{1}{4}\sqrt{1 - m_{11} - m_{22} + m_{33}} \end{aligned} \qquad \text{(B.56)}$$

Quaternion Kinematic

In aircraft kinematic, the body-fixed frame (index B) is rotated w.r.t. the NED frame (index O) and the relative rotation rate is prescribed by the angular rate $\left(\vec{\boldsymbol{\omega}}^{OB}\right)_B$ whose components are given in body-fixed coordinates. The associated attitude is described by some quaternion $\mathbf{q}_{BO}(t)$ that depends on time t. It is therefore desired to have a differential equation that describes that change of the quaternion with time, dependent on the prescribed angular rate. Therefore, the angular rate is written in terms of a unity length direction vector $(\vec{\mathbf{n}})_B(t)$ and an absolute value ω:

$$\left(\vec{\boldsymbol{\omega}}^{OB}\right)_B(t) = \left[\omega \cdot (\vec{\mathbf{n}})_B\right](t)$$

It follows that, within some infinitesimal time interval dt the B-frame is rotating around the rotation axis $(\vec{\mathbf{n}})_B$ about a small angle, which amounts, in a first order approximation, to $d\mu = \omega \cdot dt$ and the quaternion describing this rotation is given by

$$d\underline{\mathbf{q}} = \begin{pmatrix} 1 \\ (\vec{\mathbf{n}})_B \, \omega/2 \cdot dt \end{pmatrix}$$

where (B.48) and the small angle approximation has been used. Hence the quaternion, describing the transformation $0 \to B$ at $t + dt$ is, in light of (B.52)

$$\underline{\mathbf{q}}_{BO}(t + dt) = \underline{\mathbf{q}}_{BO}(t) * d\underline{\mathbf{q}}$$

and the time derivative evaluates to

$$\dot{\underline{\mathbf{q}}}(t) = \lim_{dt \to 0} \left[\frac{\underline{\mathbf{q}}_{BO}(t) * \left(d\underline{\mathbf{q}} - \underline{\mathbf{I}} \right)}{dt} \right] = \frac{1}{2} \cdot \underline{\mathbf{q}}_{BO}(t) * \begin{pmatrix} 0 \\ \left(\vec{\boldsymbol{\omega}}^{OB} \right)_B (t) \end{pmatrix} \tag{B.57}$$

Using the definitions of section B.6.2, the quaternion kinematic equation (B.57) is compactly written in matrix vector notation:

$$\dot{\underline{\mathbf{q}}} = \frac{1}{2} \cdot \bar{\mathbf{E}}_q^T \cdot \left(\vec{\boldsymbol{\omega}}^{OB} \right)_B \tag{B.58}$$

It can be concluded from properties (B.41) and (B.42), that the quaternion differential equation automatically preserves a unit quaternion as time evolves. This becomes reasonable by computing the time derivative of the quaternion norm and utilizing (B.36) and (B.39)

$$\begin{aligned} \frac{d}{dt}\left(\left\| \underline{\mathbf{q}} \right\| \right) &= \underline{\mathbf{q}}^* * \dot{\underline{\mathbf{q}}} + \dot{\underline{\mathbf{q}}}^* * \underline{\mathbf{q}} = \overset{+}{\mathbf{Q}}^* \cdot \dot{\underline{\mathbf{q}}} + \bar{\mathbf{Q}} \cdot \dot{\underline{\mathbf{q}}}^* \\ &= \frac{1}{2} \left(\begin{bmatrix} \underline{\mathbf{q}}^T \\ \bar{\mathbf{E}}_q \end{bmatrix} \cdot \bar{\mathbf{E}}_q^T + \begin{bmatrix} \underline{\mathbf{q}}^{*T} \\ \overset{+}{\mathbf{E}}_{q^*} \end{bmatrix} \cdot \begin{pmatrix} - \overset{+}{\mathbf{E}}_{q^*}^T \end{pmatrix} \right) \cdot \left(\vec{\boldsymbol{\omega}}^{OB} \right)_B \end{aligned}$$

We have used

$$\dot{\underline{\mathbf{q}}}^* = \begin{pmatrix} \dot{q}_0 \\ -\dot{\mathbf{q}} \end{pmatrix} = \frac{1}{2} \begin{bmatrix} -\mathbf{q}^T \\ -q_0 \mathbf{I} - \tilde{\mathbf{q}} \end{bmatrix} \cdot \left(\vec{\boldsymbol{\omega}}^{OB} \right)_B = -\frac{1}{2} \overset{+}{\mathbf{E}}_{q^*}^T \cdot \left(\vec{\boldsymbol{\omega}}^{OB} \right)_B$$

and with properties (B.40) - (B.43) we finally get

$$\frac{d}{dt}\left(\left\| \underline{\mathbf{q}} \right\| \right) = \mathbf{0}$$

and hence, the length of a quaternion, subjected to (B.58) is preserved.

Appendix C

Fundamentals on Lyapunov Stability

In classical linear control, stability of dynamic systems is analyzed by consideration of eigenvalues. If all eigenvalues lie in the open left half complex plane, the system is asymptotically stable, if at least one eigenvalue lies on the imaginary axis, the system is stable and if at least one eigenvalue lies in the open right half complex plane, the system is unstable.

In adaptive control systems, even if it is the imaginable simplest one, the closed loop dynamic is per se nonlinear. Since, due to adaptation, controller gains are dynamic states of the system and the control law generally is a product of plant state and controller parameter – thus a multiplication of states – the system is rendered nonlinear. For this reason, there is need for stability concepts suitable for nonlinear systems of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}, t) \;, \quad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{C.1}$$

with state vector $\mathbf{x}(t) \in \mathcal{D}$ and a mapping $\mathbf{f} : \mathcal{D} \times \mathbb{R}_0^+ \to \mathbb{R}^n$ locally Lipschitz in $\mathbf{x}$ and piecewise continuous in t, where $\mathcal{D} \subset \mathbb{R}^n$ is an open connected set containing the origin. Due to the explicit dependence on t, the system is called *non-autonomous*. Further, in some cases it is required that $\mathbf{x} = \mathbf{0}$ is an equilibrium, i.e. $\mathbf{f}(\mathbf{0}, t) = \mathbf{0}$ for all $t \geq t_0$. The question on stability of the equilibrium that arises is the following. How does a trajectory, starting near $\mathbf{x}=\mathbf{0}$, behave relative to the equilibrium.

In the 19th century, the Russian mathematician Aleksandr Mikhailovich Lyapunov developed a stability theory for systems of nonlinear ordinary differential equations without explicit knowledge of its solution. The basic idea thereby is the fact, that mechanical systems, commonly described by differential equations, have a total energy, consisting of kinetic and potential energy, which depends on the system states. If the system is dissipative, meaning that energy gets lost as time progresses, the system intuitively approaches some equilibrium of minimum total energy.

This insight was generalized by Lyapunov by introducing a state dependent generalized energy function, which is also denoted as *Lyapunov function candidate*. The system is dissipative if the time derivative of the Lyapunov function candidate is smaller than zero, in this case, it is called a *Lyapunov function*. Hence it can be concluded, that the system is stable, i.e. the states will not grow unbounded, since this meant that the generalized energy grew unbounded.

The Lyapunov stability framework is presented in a rigorous mathematical manner, focus is put on a detailed derivation. However, in order not to overstretch the chapter, some basics facts about existence and uniqueness of solutions to the aforementioned system, which are tied to the Lipschitz condition for $\mathbf{f}(\mathbf{x},t)$, are assumed to be known. These facts are though comprehensively presented in [Kha02]. At first, we start with various definitions of stability, used in the remainder of the chapter.

C.1 Stability Definitions

Stability

Definition C.1 Stability

The equilibrium of (C.1) is said to be *stable*,
if for each $\varepsilon > 0$ and $t_0 \geq 0$ one can find a $\delta(\varepsilon,t_0) > 0$
such that $\|\mathbf{x}_0\| < \delta(\varepsilon,t_0)$ implies that $\|\mathbf{x}(t)\| < \varepsilon$ for all $t \geq t_0$

Remarks

A state trajectory is stable, if it is possible to force a trajectory to reside in an arbitrarily small ball around the equilibrium for all $t \geq t_0$, i.e. $\|\mathbf{x}(t)\| < \varepsilon$, if the initial state is sufficiently close to the equilibrium, i.e. $\|\mathbf{x}_0\| < \delta$. To a given ε and t_0, one can find infinitely many $\delta(\varepsilon,t_0)$ because, if some δ_1 is found to some ε and t_0 then also every $\delta_2 \in (0,\delta_1)$ is also a valid one. The supremum of all $\delta(\varepsilon,t_0)$ will be denoted as $\bar{\delta}(\varepsilon,t_0)$, i.e.

$$\bar{\delta}(\varepsilon,t_0) = \sup(\delta(\varepsilon,t_0)) \tag{C.2}$$

But even as $\varepsilon \to \infty$, $\bar{\delta}(\varepsilon,t_0)$ does not necessarily grow unbounded but, for some $\bar{\delta}_\infty(t_0) > 0$:

$$\lim_{\varepsilon\to\infty}\left(\bar{\delta}(\varepsilon,t_0)\right) = \bar{\delta}_\infty(t_0) < \infty \tag{C.3}$$

The following specifications can be attributed to stability

1. Stability is attributed to be *uniform*, if δ can be chosen independent of t_0.
2. Stability is attributed to be *global*, if δ can be chosen to be arbitrarily large for a sufficiently large ε, i.e. $\lim_{\varepsilon\to\infty}(\delta(\varepsilon,t_0)) = \infty$, $\forall t_0 \geq 0$.

Asymptotic Stability

Definition C.2 Asymptotic Stability

The equilibrium of (C.1) is said to be *asymptotically stable*, if it is stable and additionally there is a $\rho(t_0)$ such that for any $\eta > 0$ there is a $T(\eta, t_0, \mathbf{x}_0)$ such that $\|\mathbf{x}_0\| < \rho(t_0)$ implies that $\|\mathbf{x}(t)\| < \eta$ for all $t \geq t_0 + T$.

Remarks

Asymptotic stability additionally requires that the state trajectory approaches the equilibrium as time goes to infinity. The rate of convergence thereby depends on initial time and initial condition.

Again, the following specifications can be attributed to asymptotic stability

1. Asymptotic stability is attributed to be *uniform*, if the equilibrium is uniformly stable and T is independent of t_0, $\mathbf{x}_0$ and ρ is independent of t_0. T merely depends on some $\delta \in (0, \rho]$ i.e. $T(\eta, \delta)$ such that $\|\mathbf{x}(t)\| < \eta$ for all $t \geq t_0 + T(\eta, \delta)$ and $\|\mathbf{x}_0\| < \delta$.
2. Asymptotic stability is attributed to be *global*, if the equilibrium is globally stable and additionally $\rho(t_0) = \infty$ for all t_0.

Exponential Stability

The weak point of asymptotic stability is, that it does not provide any statement about the rate of convergence. It is merely assures that the trajectory approaches the equilibrium, but the time that is needed therefore could be arbitrarily large. A remedy is the definition of exponential stability.

Definition C.3 Exponential Stability

The equilibrium of (C.1) is said to be *exponentially stable*, if there are constants $c, k, \lambda > 0$ such that $\|\mathbf{x}\| \leq k\|\mathbf{x}_0\| e^{-\lambda(t-t_0)}$ if $\|\mathbf{x}_0\| < c$

If $x = \infty$ the equilibrium is *globally exponentially stable*.

λ is denoted as *convergence rate.*

Boundedness and Ultimate Boundedness

The various definitions of stability require that the system has an isolated equilibrium. This, however, is not true in any case. Particularly when analyzing systems, subjected to time variant disturbances, there is potentially no equilibrium at all. This essentially

means that $\mathbf{f}(\mathbf{0},t)=\mathbf{0}$ is not fulfilled for all $t \geq t_0$. In such cases the best, one can hope is that, a state trajectory is bounded.

Definition C.4 Boundedness and Ultimate Boundedness

Consider a solution of (C.1) $\mathbf{x}(t)$ for $t \geq t_0$ and $\mathbf{x}(t_0)=\boldsymbol{x}_0$ where $\mathbf{x}=0$ is not necessarily an equilibrium. Then $\mathbf{x}(t)$ is said to be

1. *uniformly bounded*
   if, for every $\delta \in [0, \rho]$ and some $\rho > 0$ there is a $\varepsilon(\delta)$ such that
   $\|\mathbf{x}_0\| \leq \delta$ implies $\|\mathbf{x}(t)\| \leq \varepsilon(\delta)$ for all $t \geq t_0$.
2. *globally uniformly bounded*
   if an $\varepsilon(\delta)$ is found for any $\delta > 0$.
3. *uniformly ultimately bounded with ultimate bound b*
   if, for every $\delta \in [0, \rho]$ and some $\rho > 0$,
   there is a $T(\delta, b)$ for some $b > 0$ such that
   $\|\mathbf{x}_0\| \leq \delta$ implies $\|\mathbf{x}(t)\| \leq b$ for all $t \geq t_0 + T(\delta, b)$
4. *globally uniformly ultimately bounded with ultimate bound b*
   if $T(\delta, b)$ exists for every $\delta > 0$.

Remark

The boundedness is uniform since the $\varepsilon(\delta)$ to be found is independent of the initial time t_0. Further uniform boundedness seems to be similar to the definition of stability. In fact stability requires to find a δ for a given ε while uniform boundedness is defined the other way round, i.e.to find an ε for a given δ. In other words, stability assures that the trajectory remains in an arbitrarily small ball around the equilibrium if only the initial value is sufficiently close the equilibrium whereas boundedness start with a bound on the initial condition and requires finding a bound, the trajectory does not leave for $t \geq t_0$.

C.2 Positive Definite Functions

Lyapunov functions have to fulfill the property, that they accept a minimum in the equilibrium, which is clear from the fact, that a stable equilibrium is the point with lowest energy. This property is precisely defined next.

Definition C.5 Positive (Semi-) Definite Function

Let $\mathcal{D} \subset \mathbb{R}^n$ be an open connected set containing the origin,
let further $V : \mathcal{D} \to \mathbb{R}_0^+$ be a continuously differentiable function.
Then V is said to be positive (semi-) definite if
$V(\mathbf{0}) = 0$ and $V(\mathbf{x}) > 0$ $(V(\mathbf{x}) \geq 0)$ for all $\mathbf{x} \in \mathcal{D} \setminus \mathbf{0}$

Remark

A positive (semi-) definite function V is said to be *radially unbounded*, if $\lim_{\|\mathbf{x}\|\to\infty}(V(\mathbf{x}))=\infty$.

C.3 Comparison Principles

In the non-autonomous case, Lyapunov functions generally depend explicitly on time. In order to establish uniform stability by the methods of Lyapunov, it is necessary, that the time variant Lyapunov function is enclosed between two comparison functions that do not depend explicitly on time. A comprehensive introduction to comparison functions is given in [Kha02].

Class $\mathcal{K}$ function

Definition C.6 Class $\mathcal{K}$ Function

A continuous function $\alpha:[0,a)\to\mathbb{R}_0^+$ is said to belong to class $\mathcal{K}$, if it is strictly increasing and $\alpha(0)=0$.

It is said to belong to class $\mathcal{K}_\infty$, if $a=\infty$ and $\lim_{x\to\infty}(\alpha(x))=\infty$

Class $\mathcal{KL}$ function

Definition C.7 Class $\mathcal{KL}$ Function

A continuous function $\beta:[0,a)\times[0,\infty)\to\mathbb{R}_0^+$ is said to belong to class $\mathcal{KL}$ if, for a fixed t_0, the function $\beta(x,t_0)$ is a class $\mathcal{K}$ function with respect to x and, for each fixed x_0, the function $\beta(x_0,t)$ is nonincreasing with respect to t and

$$\lim_{t\to\infty}\beta(x_0,t)=0$$

Remarks

Let

- $\alpha_1(x), \alpha_2(x)$ be some class $\mathcal{K}$ functions on $[0,a)$,
- $\alpha_3(x), \alpha_4(x)$ class $\mathcal{K}_\infty$ functions,
- $\beta(x,t)$ a class $\mathcal{KL}$ function on $[0,a)$
- $f:(0,a)\to\mathbb{R}^+$ be a positive nondecreasing function, not necessarily continuous

- $g:[0,a)\to\mathbb{R}_0^+$ be a continuous, nondecreasing function where $g(0)=0$ *and* $g(x)>0$ for $0<x<a$.

Then the following properties can be easily verified:

1. The inverse α_1^{-1} belongs to $\mathcal{K}$.
2. The inverse α_3^{-1} belongs to $\mathcal{K}_\infty$.
3. $\alpha_1(\alpha_2(x))$ belongs to $\mathcal{K}$.
4. $\alpha_3(\alpha_4(x))$ belongs to $\mathcal{K}_\infty$.
5. $\sigma(x,t)=\alpha_1(\beta(\alpha_2(x),t))$ belongs to $\mathcal{KL}$.
6. There is a class $\mathcal{K}$ function $\alpha:[0,a)\to\mathbb{R}_0^+$ such that $\alpha(x)<f(x)$ on $x\in(0,a)$
7. There are class $\mathcal{K}$ functions $\beta_1(x)$ and $\beta_2(x)$ on $x\in[0,a)$ such that $\beta_1(x)<g(x)<\beta_2(x)$ for $x\in(0,a)$

It is sometimes convenient to express stability in the sense of Lyapunov in terms of class $\mathcal{K}$ and $\mathcal{KL}$ functions, according to the following lemma.

Lemma C.1 Stability and Class $\mathcal{K}$, $\mathcal{KL}$ Functions

The equilibrium of (C.1) is

1. *uniformly stable*
   if and only if there exists a class $\mathcal{K}$ function $\alpha(x)$ and a positive constant ρ independent of t_0, such that
   $\|\mathbf{x}(t)\|\le\alpha(\|\mathbf{x}_0\|)$ for all $t\ge t_0$ and $\|\mathbf{x}_0\|<\rho$
2. *uniformly asymptotically stable*
   if and only if there exists a class $\mathcal{KL}$ function $\beta(x,t)$ and a positive constant ρ, such that
   $\|\mathbf{x}(t)\|\le\beta(\|\mathbf{x}_0\|,t-t_0)$ for all $t\ge t_0$ and $\|\mathbf{x}_0\|<\rho$
3. *globally uniformly asymptotically stable*
   if and only the inequality for uniform asymptotical stability
   $\|\mathbf{x}(t)\|\le\beta(\|\mathbf{x}_0\|,t-t_0)$ holds for all $t\ge t_0$ and all $\mathbf{x}_0\in\mathbb{R}^n$

Proof Part 1

NECESSARY PART:

"*If there is a class $\mathcal{K}$ function such that $\|\mathbf{x}(t)\|\le\alpha(\|\mathbf{x}_0\|)$ for all $t\ge t_0$ and $\|\mathbf{x}_0\|<\rho$, then the equilibrium is uniformly stable*."

For a given ε, choose

$$\delta(\varepsilon)=\begin{cases}\alpha^{-1}(\varepsilon) & \text{if } \varepsilon<\overline{\alpha}\\ \rho & \text{if } \varepsilon\geq\overline{\alpha}\end{cases} \quad \text{where } \overline{\alpha}=\lim_{x\to\rho^-}\alpha(x)$$

Hence $\delta(\varepsilon)$ is independent of t_0 which implies uniform stability.

SUFFICIENT PART

"*If the equilibrium is uniformly stable, then there exists a class $\mathcal{K}$ function $\alpha(x)$ such that $\|\mathbf{x}(t)\|\leq\alpha(\|\mathbf{x}_0\|)$ for all $t\geq t_0$ and $\|\mathbf{x}_0\|<\rho$.*"

Uniform stability means that, for every ε, there is a $\delta(\varepsilon)$ such that $\|\mathbf{x}_0\|<\delta$ implies $\|\mathbf{x}(t)\|<\varepsilon$ for all $t\geq t_0$ and $\delta(\varepsilon)$ can be chosen to be nondecreasing. In order to accept this fact, note that, for some ε_1 and the associated $\delta(\varepsilon_1)$, we have $\|\mathbf{x}(t)\|<\varepsilon_1$ if $\|\mathbf{x}_0\|<\delta(\varepsilon_1)$. If we require $\|\mathbf{x}(t)\|<\varepsilon_2$ for some $\varepsilon_2>\varepsilon_1$, it is surely fulfilled if $\|\mathbf{x}_0\|<\delta(\varepsilon_1)$. Hence we can, at least, choose $\delta(\varepsilon_2)=\delta(\varepsilon_1)$, possibly also $\delta(\varepsilon_2)>\delta(\varepsilon_1)$. Hence $\delta(\varepsilon)$ is a positive, nondecreasing function and, by property 6, there exists some class $\mathcal{K}$ function $\beta(\varepsilon)<\delta(\varepsilon)$ for $\varepsilon>0$. Further, define

$$\lim_{\varepsilon\to\infty}(\beta(\varepsilon))=\rho\,.$$

It follows that, if $\|\mathbf{x}_0\|<\beta(\varepsilon)$, then $\|\mathbf{x}(t)\|<\varepsilon$. Then

$$\alpha(x)=\beta^{-1}(x) \tag{C.4}$$

is a class $\mathcal{K}$ function on $[0,\rho)$ and we have $\|\mathbf{x}(t)\|\leq\alpha(\|\mathbf{x}_0\|)$ if $\|\mathbf{x}_0\|<\rho$.

□

Proof Part 2

NECESSARY PART

"*If there is a class $\mathcal{KL}$ function $\beta(x,t)$ such that $\|\mathbf{x}(t)\|\leq\beta(\|\mathbf{x}_0\|,t-t_0)$ for $t\geq t_0$ and $\|\mathbf{x}_0\|<\rho$, then the equilibrium is uniformly asymptotically stable.*"

For a given ε choose

$$\delta=\begin{cases}\beta^{-1}(\varepsilon,0) & \text{if } \varepsilon<\overline{\beta}\\ \rho & \text{if } \varepsilon\geq\overline{\beta}\end{cases} \quad \text{where } \overline{\beta}=\lim_{x\to\rho^-}\beta(x,0) \tag{C.5}$$

where $\beta^{-1}(\cdot,\cdot)$ denotes the inverse w.r.t. to the first argument. Therefore $\|\mathbf{x}(t)\|<\varepsilon$, if $\|\mathbf{x}_0\|<\delta$ which proves uniform stability, as δ is independent of t_0. Further, let

$$\overline{\beta}(x,t-t_0)=\begin{cases}\beta(x,t-t_0) & \text{for } 0\leq x<\rho\\ \lim_{x\to\rho^-}\beta(x,t-t_0) & \text{for } x=\rho\end{cases}$$

(If the limit does not exist, it is ∞ such that every real number is smaller.) Then, for every $\eta>0$ and $\delta\in(0,\rho]$ there is some $T(\eta,\delta)$ such that $\bar{\beta}(\delta,T)<\eta$ and hence $\|\mathbf{x}(t)\|<\eta$ for all $t\geq t_0+T(\eta,\delta)$, if $\|\mathbf{x}_0\|<\delta$, which shows uniform asymptotic stability.

SUFFICIENT PART

"*If the equilibrium is uniformly asymptotically stable, then there exists some class $\mathcal{KL}$ function such that* $\|\mathbf{x}(t)\|\leq\beta(\|\mathbf{x}_0\|,t-t_0)$ *for all* $t\geq t_0$ *and* $\|\mathbf{x}_0\|<\rho$."

Since uniform asymptotic stability also includes uniform stability, from the first part, there exists some class $\mathcal{K}$ function $\alpha(x)$ such that

$$\|\mathbf{x}(t)\|\leq\alpha(\|\mathbf{x}_0\|) \tag{C.6}$$

for $t\geq t_0$ and $\|\mathbf{x}_0\|<\rho_\alpha$. Further, let

$$\bar{\alpha}(x)=\begin{cases}\alpha(x) & \text{for } 0<x<\rho_\alpha\\ \lim_{x\to\infty}(\alpha(x)) & \text{for } x=\rho_\alpha.\end{cases} \tag{C.7}$$

(If the limit does not exist, it is ∞ such that every real number is smaller.) By uniform asymptotic stability, there exist $T(\eta,\delta)$ for any $\eta>0$ and $\delta\in(0,\rho_\beta]$, $\rho_\beta>0$, such that, $\|\mathbf{x}(t)\|<\eta$ if $t\geq t_0+T(\eta,\delta)$ and $\|\mathbf{x}_0\|<\delta$. In the following, we choose $\rho=\min(\rho_\alpha,\rho_\beta)$. It will be shown next that $T(\eta,\delta)$ can be chosen such that, for $\eta>0$ and $\delta\in(0,\rho]$,

- $T(\eta,\delta)=0$ for $\eta\geq\bar{\alpha}(\delta)$
- $T(\eta,\delta)$ is nonnegative,
- $T(\eta,\delta)$ is nonincreasing in η and
- $T(\eta,\delta)$ is nondecreasing in δ

as will be shown in the following.

The first property directly follows from (C.6) and (C.7). Non-negativity of T is clear from the fact that we merely consider times $t\geq t_0$. Moreover, consider some $\eta_1>0$, $\eta_2>\eta_1$ and its associated $T(\eta_1,\delta)$, $T(\eta_2,\delta)$. By definition, $\|\mathbf{x}(t)\|$ falls below η_1 at least at $T(\eta_1,\delta)$. Since $\eta_2>\eta_1$ it is also below η_2 at the same time, or even earlier. Hence $T(\eta_2,\delta)\leq T(\eta_1,\delta)$.

Analogously, consider some $\delta_1>0$, $\delta_2>\delta_1$ and its associated $T(\eta,\delta_1)$, $T(\eta,\delta_2)$. If $\|\mathbf{x}_0\|<\delta_2$ then $\|\mathbf{x}(t)\|<\eta$ at $t_0+T(\eta,\delta_2)$. Furthermore, if $\|\mathbf{x}_0\|<\delta_1$ then $\|\mathbf{x}_0\|<\delta_2$, too and it is assured that $\|\mathbf{x}(t)\|<\eta$ at the same time, or even earlier. Hence $T(\eta,\delta_1)\leq T(\eta,\delta_2)$.

Next we define, for $(\delta,\eta)\in(0,\rho]\times(0,2\bar{\alpha}(\delta)]$

$$\hat{T}_\delta(\eta)=\frac{2}{\eta}\int_{\eta/2}^{\eta}T(\xi,\delta)d\xi+\frac{1}{\eta}-\frac{1}{2\bar{\alpha}(\delta)} \tag{C.8}$$

which has the following properties.

1. $\hat{T}_\delta(\eta)$ is nonnegative
2. $\hat{T}_\delta(\eta)$ is continuous w.r.t. η
3. $\hat{T}_\delta(\eta)$ is strictly decreasing w.r.t. η
4. $\hat{T}_\delta(\eta)$ has the following behavior at the boundaries of the definition space:

$$T_\delta(2\overline{\alpha}(\delta))=0 \quad , \quad \lim_{\eta\to 0^+}(T_\delta(\eta))=\infty$$

5. $\hat{T}_\delta(\eta)$ is strictly increasing w.r.t.δ
6. $\hat{T}_\delta(\eta)\geq T(\eta,\delta)$

Remarks

1: $\hat{T}_\delta(\eta)$ is nonnegative since $T(\eta,\delta)$ is nonnegative and $1/\eta-1/(2\overline{\alpha}(\delta))\geq 0$ in the definition space.

2: Consider, at fixed η_0, $\left|\hat{T}_\delta(\eta_0+h)-\hat{T}_\delta(\eta_0)\right|$ for a sufficiently small h such that $\left[\eta_0/2+h,\eta_0+h\right]\subset(0,2\overline{\alpha}(\delta)]$. Note that $T(\eta,\delta)\leq M$ on the interval $\left[\eta_0/2+h,\eta_0+h\right]$ for some $M>0$. Then

$$\begin{aligned}\left|\hat{T}_\delta(\eta_0+h)-\hat{T}_\delta(\eta_0)\right| &= \left|\frac{2}{\eta_0}\int_{\frac{\eta_0}{2}+h}^{\eta_0+h}T(\xi,\delta)d\xi+\frac{1}{\eta_0+h}-\frac{1}{2\overline{\alpha}(\delta)}-\frac{2}{\eta_0}\int_{\frac{\eta_0}{2}}^{\eta_0}T(\xi,\delta)d\xi-\frac{1}{\eta_0}+\frac{1}{2\overline{\alpha}(\delta)}\right| \\ &\leq \left|\frac{2}{\eta_0}\int_{\frac{\eta_0}{2}}^{\frac{\eta_0}{2}+h}T(\xi,\delta)d\xi\right|+\left|\frac{2}{\eta_0}\int_{\eta_0}^{\eta_0+h}T(\xi,\delta)d\xi\right|+\frac{|h|}{\eta_0(\eta_0-|h|)} \\ &\leq \left(\frac{4M}{\eta_0}+\frac{1}{\eta_0(\eta_0-|h|)}\right)|h|\end{aligned}$$

and for every $\varepsilon>0$, one can find a $\delta>0$ such that $\left|\hat{T}_\delta(\eta_0+h)-\hat{T}_\delta(\eta_0)\right|<\varepsilon$ if $|h|<\delta$ which shows continuity.

3:Let $\eta_2>\eta_1>0$, and $\eta_2=k\cdot\eta_1$, $k>1$. Using a change of variables $\xi=k\overline{\xi}$, we get

$$\begin{aligned}\hat{T}_\delta(\eta_2) &= \frac{2}{\eta_2}\int_{\frac{\eta_2}{2}}^{\eta_2}T(\xi,\delta)d\xi+\frac{1}{\eta_2}-\frac{1}{2\overline{\alpha}(\delta)}=\frac{2}{k\eta_1}\int_{\frac{k\eta_1}{2}}^{k\eta_1}T(\xi,\delta)d\xi+\frac{1}{k\eta_1}-\frac{1}{2\overline{\alpha}(\delta)} \\ &= \frac{2k}{k\eta_1}\int_{\frac{\eta_1}{2}}^{\eta_1}T\left(k\overline{\xi},\delta\right)d\overline{\xi}+\frac{1}{k\eta_1}-\frac{1}{2\overline{\alpha}(\delta)}=\frac{2}{\eta_1}\int_{\frac{\eta_1}{2}}^{\eta_1}T\left(k\overline{\xi},\delta\right)d\overline{\xi}+\frac{1}{k\eta_1}-\frac{1}{2\overline{\alpha}(\delta)}.\end{aligned}$$

Since $T(\xi,\delta)$ is nonincreasing with respect to ξ, we have $T(k\xi,\delta)\leq T(\xi,\delta)$ and therefore

$$\hat{T}_\delta(\eta_2)=\frac{2}{\eta_1}\int_{\frac{\eta_1}{2}}^{\eta_1}T\left(k\overline{\xi},\delta\right)d\overline{\xi}+\frac{1}{k\eta_1}-\frac{1}{2\overline{\alpha}(\delta)}<\frac{2}{\eta_1}\int_{\frac{\eta_1}{2}}^{\eta_1}T\left(\overline{\xi},\delta\right)d\overline{\xi}+\frac{1}{\eta_1}-\frac{1}{2\overline{\alpha}(\delta)}=\hat{T}_\delta(\eta_1)$$

4: The asymptotic behavior at the boundary of the definition set is due to the additional term $1/\eta-1/(2\overline{\alpha}(\delta))$ and the fact that the integral vanishes for $\eta\geq 2\overline{\alpha}(\delta)$, where $\overline{\alpha}(\delta)\leq\overline{\alpha}(\rho)$.

5: Let $\delta_1>0$, $\delta_2 > \delta_1>0$ and some fixed $\eta_0>0$. The fact that $T(\xi,\delta)$ is nondecreasing with respect to δ, yields

$$\hat{T}_{\delta_2}(\eta_0)=\frac{2}{\eta_0}\int_{\frac{\eta_0}{2}}^{\eta_0}T(\xi,\delta_2)d\xi+\frac{1}{\eta_0}-\frac{1}{2\overline{\alpha}(\delta_2)}>\frac{2}{\eta_0}\int_{\frac{\eta_0}{2}}^{\eta_0}T(\xi,\delta_1)d\xi+\frac{1}{\eta_0}-\frac{1}{2\overline{\alpha}(\delta_1)}=\hat{T}_{\delta_1}(\eta_0)$$

6: Since $T(\xi,\delta)$ is nonincreasing w.r.t. ξ, we have

$$\frac{2}{\eta}\int_{\frac{\eta}{2}}^{\eta}T(\xi,\delta)d\xi\geq\frac{2}{\eta}\cdot T(\eta,\delta)\cdot\int_{\frac{\eta}{2}}^{\eta}d\xi=T(\eta,\delta)$$

Further, $1/\eta-1/(2\overline{\alpha}(\delta))\geq 0$ in the definition space yields $\hat{T}_\delta(\eta)\geq T(\eta,\delta)$

Since $\hat{T}_\delta(\eta)$ is strictly decreasing w.r.t. η, the inverse

$$\hat{\eta}_\delta(\tau)=\hat{T}_\delta^{-1}(\tau) \qquad \text{(C.9)}$$

exists for $\tau\in[0,\infty)$ for each allowed δ and $\hat{\eta}_\delta(\tau)$ has the following properties

1. $\hat{\eta}_\delta(\tau)$ is continuous, positive and strictly decreasing in τ (from properties 1-3 of $\hat{T}_\delta(\eta)$)
2. $\hat{\eta}_\delta(\tau)$ is positive and strictly increasing in δ (from property 5 of $\hat{T}_\delta(\eta)$)
3. $\hat{\eta}_\delta(0)=2\overline{\alpha}(\delta)$ (from property 4 of $\hat{T}_\delta(\eta)$)
4. $\lim_{\tau\to\infty}(\hat{\eta}_\delta(\tau))=0$ (from property 4 of $\hat{T}_\delta(\eta)$)

Then – from property 6 – for a fixed $T_0>0$, we have

$$T(\hat{\eta}_\delta(T_0),\delta)<\hat{T}_\delta(\hat{\eta}_\delta(T_0))=T_0.$$

Since $\|\mathbf{x}(t)\|<\hat{\eta}_\delta(T_0)$ is fulfilled for all $t\geq t_0+T(\hat{\eta}_\delta(T_0),\delta)$, it is also fulfilled for $t\geq t_0+T_0$. Consequently

$$\|\mathbf{x}(t)\|<\hat{\eta}_\delta(t-t_0) \qquad \text{(C.10)}$$

for all $t-t_0\geq 0$ and $\|\mathbf{x}_0\|<\delta$. By (C.6) and (C.10), we have $\|\mathbf{x}(t)\|^2\leq\alpha(\|\mathbf{x}_0\|)\cdot\hat{\eta}_\delta(t-t_0)$ and finally

$$\|\mathbf{x}(t)\|\leq\beta(\|\mathbf{x}_0\|,(t-t_0)) \text{ for all } t\geq t_0 \text{ and } \|\mathbf{x}_0\|<\rho \qquad \text{(C.11)}$$

where $\beta(\|\mathbf{x}_0\|,t-t_0)=\sqrt{\alpha(\|\mathbf{x}_0\|)\cdot\hat{\eta}_\rho(t-t_0)}$ is a class $\mathcal{KL}$ function.

□

Proof Part 3

NECESSARY PART

"*If there exists a class $\mathcal{KL}$ function $\beta(x,t)$ such that $\|\mathbf{x}(t)\|\leq\beta(\|\mathbf{x}_0\|,t-t_0)$ for all $t\geq t_0$ and $\mathbf{x}_0\in\mathbb{R}^n$, then the equilibrium is globally asymptotically stable.*"

Since $\beta(x,t)$ is a class $\mathcal{K}$ function, defined on $[0,\infty)$ for every fixed t, there is also a class $\mathcal{KL}$ function $\beta_\infty(x,t)\geq\beta(x,t)$, that belongs to class $\mathcal{K}_\infty$ for fixed t. Hence, we also have

$$\|\mathbf{x}(t)\| \le \beta_\infty(\|\mathbf{x}_0\|, t-t_0)$$

for all $t \ge t_0$ and $\mathbf{x}_0 \in \mathbb{R}^n$. Analogously to (C.5), we choose

$$\delta(\varepsilon) = \beta_\infty^{-1}(\varepsilon, 0)$$

and we have $\delta \to \infty$ as $\varepsilon \to \infty$, which proves global uniform stability. Moreover, for every $\eta, \delta > 0$, there exists a $T(\eta, \delta)$ such that $\beta(\delta, T) < \eta$ and hence $\|\mathbf{x}(t)\| < \eta$ for all $t \ge t_0 + T(\eta, \delta)$ if $\|\mathbf{x}_0\| < \delta$ which shows global uniform asymptotic stability.

SUFFICIENT PART

"*If the equilibrium is globally asymptotically stable, then there exists a class $\mathcal{KL}$ function $\beta(x,t)$ such that $\|\mathbf{x}(t)\| \le \beta(\|\mathbf{x}_0\|, t-t_0)$ for all $t \ge t_0$ and $\mathbf{x}_0 \in \mathbb{R}^n$.*"

Since global uniform asymptotical stability implies global uniform stability $\delta(\varepsilon)$ can be chosen such that

$$\lim_{\varepsilon \to \infty} \delta(\varepsilon) = \infty$$

Therefore, $\alpha(x)$ in (C.4) can be chosen to belong to class $\mathcal{K}_\infty$ and (C.4) holds for all $t \ge t_0$ and $\mathbf{x}_0 \in \mathbb{R}^n$. Moreover, also $\hat{T}_\delta$, $\hat{\eta}_\delta$ exist and (C.10) holds for all for all $\delta > 0$ and

$$\|\mathbf{x}(t)\| \le \Psi(\|\mathbf{x}_0\|, t-t_0) \qquad \text{(C.12)}$$

where

$$\Psi(x,\tau) = \min(\alpha(x), \hat{\eta}_{x+1}(\tau))$$

Unfortunately $\Psi(x,t)$ does not necessarily belong to class $\mathcal{KL}$, since $\hat{\eta}_{x+1}(\tau)$,is not necessarily continuous in x. Let us define

$$Y(x,\tau) = \int_x^{x+1} \Psi(\xi,\tau)\, d\xi + \frac{x}{(x+1)(\tau+1)}$$

which has the following properties

1. $Y(x,t)$ is continuous in both arguments
2. $Y(x,t)$ is strictly increasing w.r.t. x
3. $Y(x,t)$ is strictly decreasing w.r.t. τ and $\lim_{\tau \to \infty}(Y(x,\tau)) = 0$
4. $Y(x,t) \ge \Psi(x,t)$

1: Continuity in x follows from similar considerations as done for property 2 of $\hat{T}_r(\eta)$. Continuity in τ follows from continuity of the second term and continuity of $\Psi(x,t)$ in τ.

2: $Y(x,t)$ is strictly increasing in x because, on the one hand, $\Psi(x,t)$ is strictly increasing in x, which, in turn, follows from the fact that α (x), and $\eta_{x+1}(\tau)$ are strictly increasing in x. On the other hand, the second term in Y is strictly increasing in x. Following similar arguments as used for property 5 of $\hat{T}_r(\eta)$ the claim is proved.

3: $Y(x,t)$ is strictly decreasing in τ because, on the one hand, $\Psi(x,t)$ is strictly decreasing in τ which, in turn follows from the fact that $\eta_{x+1}(\tau)$ strictly decreasing in τ. On the other hand, the second term is strictly decreasing in τ. Following similar arguments as used for property 3 of $\hat{T}_r(\eta)$ the claim is proved.

$\lim_{\tau\to\infty}(Y(x,\tau))=0$ follows from

$$\lim_{\tau\to\infty}\hat{\eta}_\delta(\tau)=0 \text{ (property 4) which implies } \lim_{\tau\to\infty}(\Psi(x,\tau))=0$$

and from the fact that the second term in Y goes to 0 as $\tau\to\infty$.

4: $Y(x,t)\geq\Psi(x,t)$ follows from similar arguments as used for property 6 of $\hat{T}_r(\eta)$.

With (C.12) and property 4, we get

$$\|\mathbf{x}(t)\|\leq Y(\|\mathbf{x}_0\|,t-t_0) \tag{C.13}$$

for all $t\geq t_0$ and $\mathbf{x}_0\in\mathbb{R}^n$. Moreover, properties 1-3 assure that

$$\beta(x,t)=\sqrt{\alpha(x)\cdot Y(x,\tau)}$$

belongs to class $\mathcal{KL}$ and, analogously to (C.11), we obtain

$$\|\mathbf{x}(t)\|\leq\beta(\|\mathbf{x}_0\|,t-t_0)$$

for all $\mathbf{x}_0\in\mathbb{R}^n$ and $t\geq t_0$.

□

Positive definition functions can be enclosed by class $\mathcal{K}$ functions, which appear to be useful in Lyapunov stability theorems.

Lemma C.2 Positive Definite Function and Class $\mathcal{K}$ Functions [Kha02]

Let $W:\mathcal{D}\to\mathbb{R}_0^+$ a positive definite function, $\mathcal{D}\in\mathbb{R}^n$ an open connected set containing the origin. Let, for some $r>0$
$\mathcal{B}_r=\{\mathbf{x}\in\mathbb{R}^n\,|\,\|\mathbf{x}\|<r\}$ and r sufficiently small such that $\mathcal{B}_r\subset\mathcal{D}$.
Then there exist class $\mathcal{K}$ functions $\alpha_1(x)$, $\alpha_2(x)$ on $[0,r)$ such that
$\alpha_1(\|\mathbf{x}\|)\leq W(\mathbf{x})\leq\alpha_2(\|\mathbf{x}\|)$ for all $\mathbf{x}\in\mathcal{B}_r$
If $\mathcal{D}=\mathbb{R}^n$ and $W(\mathbf{x})$ is radially unbounded, r can be chosen to be ∞ and $\alpha_1(x)$, $\alpha_2(x)$ can be chosen to belong to class $\mathcal{K}_\infty$

Proof

Define functions on $[0,r)$:

$$\psi(x) = \inf_{x \le \|\mathbf{x}\| < r} W(\mathbf{x}) \qquad \phi(x) = \sup_{\|\mathbf{x}\| \le x} W(\mathbf{x})$$

Both functions are

- continuous, since W is continuous
- nondecreasing
- positive for $x>0$ since $W(\mathbf{x})>0$ for $\mathbf{x}\neq\mathbf{0}$
- 0 for $\mathbf{x}=\mathbf{0}$, since $W(\mathbf{0})=0$

and satisfy the following inequality

$$\psi(\|\mathbf{x}\|) \le W(\mathbf{x}) \le \phi(\|\mathbf{x}\|)$$

Due to the properties of $\psi(x), \phi(x)$, there are class $\mathcal{K}$ functions $\alpha_1(x)$ and $\alpha_2(x)$ such that

$$\alpha_1(x) \le \psi(x) \text{ and } \phi(x) \le \alpha_2(x) \qquad \text{(C.14)}$$

(property 7 of class $\mathcal{K}$ functions) and hence

$$\alpha_1(\|\mathbf{x}\|) \le W(\mathbf{x}) \le \alpha_2(\|\mathbf{x}\|)$$

If $\mathcal{D}=\mathbb{R}^n$ and if $W(\mathbf{x})$ is radially unbounded, we choose $r=\infty$ and obtain

$$\lim_{x\to\infty}(\psi(x)) = \infty$$

and therefore $\alpha_1(x)$ in (C.14) can be chosen such that

$$\lim_{x\to\infty}(\alpha_1(x)) = \infty$$

$\alpha_2(x)$ can be chosen to go to infinity anyway, as x goes to infinity, since $\alpha_2(x) \ge \phi(x)$ and hence $\alpha_1(x)$, $\alpha_2(x)$ can be chosen to belong to class $\mathcal{K}_\infty$.

□

The next lemma plays an important role in the proof of exponential stability.

Lemma C.3 Comparison Lemma [Kha02]

Consider the scalar differential equation $\dot{u} = f(u,t)$, $u(t_0) = u_0$,

where $f(u,t)$ is locally Lipschitz in u, uniformly in t in some domain $\mathcal{D} \subset \mathbb{R}$. Let $[t_0,T)$ be the maximal interval of existence for the solution $u(t)$ and $u(t) \in \mathcal{D}$ for all $t \in [t_0,T)$.

Let $v(t)$ be a function whose upper right Dini derivative satisfies

$$D^+ v(t) \le f(v(t),t),\ v(t_0) \le u_0 \text{ and } v(t) \in \mathcal{D} \text{ for all } t \in [t_0,T),$$

then $v(t) \le u(t)$ for all $t \in [t_0,T)$

Proof

For a definition of Dini's derivatives, refer to Definition B.30.

PART 1

It will be at first shown that some auxiliary function $z(t,\lambda)$ that satisfies the differential equation

$$\dot{z}(t) = f(z(t),t) + \lambda,\ z(t_0) = u_0 \tag{C.15}$$

with some small positive constant λ lies arbitrarily close to $u(t)$ in some interval $[t_0, t_1]$, i.e.

$$|z(t,\lambda) - u(t)| < \varepsilon$$

for some $\varepsilon > 0$, if $\lambda \le \delta$ for some $\delta > 0$. First note that the solutions of $z(t,\lambda)$ and $u(t)$ satisfy

$$u(t) = u_0 + \int_{t_0}^{t} f(u(\tau),\tau) d\tau$$

$$z(t,\lambda) = u_0 + \int_{t_0}^{t} f(z(\tau),\tau) d\tau + \lambda(t - t_0).$$

Further, since $[t_0, t_1]$ is compact and $u(t)$ is a continuous function of time, it follows that $u(t)$ is bounded on $[t_0, t_1]$. Define some tube around the nominal solution

$$U_\varepsilon = \{(z(t),t) \in \mathcal{D} \times [t_0,t_1] \,|\, |z(t) - u(t)| \le \varepsilon\}.$$

This tube is clearly compact, as it is closed and bounded and therefore $f(z,t)$ is Lipschitz on U_ε with the same constant L that holds for all elements in U_ε. Subtracting

the two solutions under the assumption that $z(t,\lambda)\in U_\varepsilon$ and accounting for the Lipschitz conditions yields an upper bound for its absolute value:

$$\begin{aligned}|u(t)-z(t,\lambda)| &\le \int_{t_0}^{t} |f(u(\tau),\tau)-f(z(\tau,\lambda),\tau)|d\tau+\lambda(t-t_0)\\ &\le L\cdot\int_{t_0}^{t}|u(\tau)-z(\tau,\lambda)|d\tau+\lambda(t-t_0)\end{aligned}$$

Application of Bellman-Gronwall Lemma (see e.g. [Tao03]) yields

$$|u(t)-z(t,\lambda)|\le\lambda(t-t_0)+L\lambda\int_{t_0}^{t}(\tau-t_0)\cdot e^{L(t-\tau)}$$

Integrating the right hand side by parts further results in

$$\begin{aligned}|u(t)-z(t,\lambda)| &\le \lambda(t-t_0)-L\lambda t_0 e^{Lt}\int_{t_0}^{t}e^{-L\tau}d\tau+L\lambda e^{Lt}\int_{t_0}^{t}\tau e^{-L\tau}d\tau\\ &= \lambda(t-t_0)-L\lambda t_0 e^{Lt}\frac{e^{-Lt}-e^{-Lt_0}}{-L}+L\lambda e^{Lt}\frac{te^{-Lt}-t_0e^{-Lt_0}}{-L}-L\lambda e^{Lt}\frac{\int_{t_0}^{t}e^{-L\tau}d\tau}{-L}\\ &= \lambda(t-t_0)+\lambda t_0\left(1-e^{L(t-t_0)}\right)-\lambda\left(t-t_0e^{L(t-t_0)}\right)-L\lambda e^{Lt}\frac{e^{-Lt}-e^{-Lt_0}}{L^2}\\ &= \lambda(t-t_0)+\lambda t_0\left(1-e^{L(t-t_0)}\right)-\lambda\left(t-t_0e^{L(t-t_0)}\right)-\frac{\lambda}{L}\left(1-e^{L(t-t_0)}\right)\\ &= \lambda(t-t_0)-\lambda(t-t_0)-\frac{\lambda}{L}\left(1-e^{L(t-t_0)}\right)\\ &= \frac{\lambda}{L}\left(e^{L(t-t_0)}-1\right)\\ &< \frac{\lambda}{L}e^{L(t-t_0)}\end{aligned}$$

If we choose $\lambda<\varepsilon L e^{-L(t_1-t_0)}=:\delta$, we have

$$|u(t)-z(t,\lambda)|<\varepsilon\cdot e^{-L(t_1-t)}$$

which shows that $z(t,\lambda)$ does not leave U_ε within the interval $[t_0,t_1]$. Therefore, the assumption of $f(z,t)$ being Lipschitz with constant L is fulfilled.

PART 2

Next, it will be shown that $v(t)\le z(t,\lambda)$ in $[t_0,t_1]$ for $\lambda>0$.

To this end, assume that, in some interval $t\in(a,b]$, where $a,b\in[t_0,t_1]$, $v(t)>z(t,\lambda)$ and $v(a)=z(a,\lambda)$. This in turn implies that

$$\frac{v(a+h)-v(a)}{h}>\frac{z(a+h,\lambda)-z(a,\lambda)}{h}$$

for all $h \in (0, b-a]$. This particularly means that for every sequence $\{h_k\}$ – $k \in \mathbb{N}$ and $h_k \in (0, b-a]$ – such that $h_k > 0$ and $\lim_{k\to\infty}(h_k) = 0$, we have

$$a_k := \frac{v(a+h_k) - v(a)}{h_k} > \frac{z(a+h_k, \lambda) - z(a, \lambda)}{h_k} =: b_k$$

Since $a_k > b_k$ for all $k \in \mathbb{N}$ it can be easily concluded that

$$\limsup_{k\to\infty}(a_k) \geq \limsup_{k\to\infty}(b_k)$$

if the respective limits exist. The left limit exists from the assumption of the theorem, since $D^+v(a) = \limsup_{k\to\infty}(a_k)$. Further since the time derivative $\dot{z}(a, \lambda)$ exists (since in obeys the differential equation), we also have

$$\limsup_{k\to\infty}(b_k) = \lim_{k\to\infty}(b_k) = \dot{z}(a, \lambda) = f(z(a), a) + \lambda$$

which implies that

$$D^+v(a) = \limsup_{k\to\infty}(a_k) > f(z(a), a)$$

which contradicts the assumption $D^+v(a) \leq f(z(a), a)$.

PART 3

Using the results of parts 1 and 2 it can be shown that $v(t) \leq u(t)$ for all $t \in [t_0, t_1]$.

Therefore, assume that at some $a \in [t_0, t_1]$, we have

$$v(a) = \frac{\varepsilon}{2} + u(a)$$

for some $\varepsilon > 0$. Then for some sufficiently small δ and $\lambda \leq \delta$, we have

$$|u(a) - z(a, \lambda)| \leq \varepsilon$$

from part 1. Then taking the difference yields

$$\begin{aligned} v(a) - z(a, \lambda) &= (v(a) - u(a)) + (u(a) - z(a, \lambda)) \\ &\geq (v(a) - u(a)) - \varepsilon \\ &= 2\varepsilon - \varepsilon = \varepsilon \end{aligned}$$

which contradicts part 2 of the proof, and hence $v(t) \leq u(t)$ for all $t \in [t_0, t_1]$.

□

C.4 Lyapunov's Direct Method

Lyapunov's direct method is considered as the main theorem in Lyapunov stability analysis and is intensively employed in the field of adaptive control. The theorem including its proof is presented in the following.

Theorem C.1 Lyapunov's Direct Method

Consider system (C.1).

Let $V:\mathcal{D}\times\mathbb{R}_0^+\to\mathbb{R}_0^+$ be a continuously differentiable function in all arguments such that

1. $W_1(\mathbf{x})\le V(\mathbf{x},t)\le W_2(\mathbf{x})$
2. $\frac{dV}{dt}=\frac{\partial V}{\partial t}+\frac{\partial V}{\partial \mathbf{x}}\cdot\mathbf{f}(\mathbf{x},t)\le -W_3(\mathbf{x})$

where are $W_1(\mathbf{x})$, $W_2(\mathbf{x})$ are positive definite functions on $\mathcal{D}$.

The equilibrium of (C.1) is

1. *uniformly stable* if $W_3(\mathbf{x})$ is positive semi-definite on $\mathcal{D}$.
2. *uniformly asymptotically stable* if $W_3(\mathbf{x})$ is positive definite on $\mathcal{D}$.

If $\mathcal{D}=\mathbb{R}^n$ and $\lim_{\|\mathbf{x}\|\to\infty}(W_1(\mathbf{x}))=\infty$, the above statements hold globally.

Remark

It is actually sufficient that V is the continuously differentiable on $\mathcal{D}\backslash 0$. As will be shown in the proof it is merely required that V is continuous at $\mathbf{x}=0$. It is therefore also allowed that

$$\frac{\partial V(\mathbf{x},t)}{\partial \mathbf{x}}\quad,\quad\frac{\partial V(\mathbf{x},t)}{\partial t}$$

do not exist for $\mathbf{x}=\mathbf{0}$ at all and the second condition is weakened to

$$\frac{\partial V(\mathbf{x},t)}{\partial t}+\frac{\partial V(\mathbf{x},t)}{\partial \mathbf{x}}\mathbf{f}(\mathbf{x},t)\le -W_3(\mathbf{x})\ \text{ on }\ \mathbf{x}\in\mathcal{D}\setminus\mathbf{0}$$

Proof

Uniform Stability

For the case that $\mathbf{x}(t_1)$ equals $\mathbf{0}$ for some $t_1\ge t_0$, $\mathbf{x}(t)$ will remain $\mathbf{0}$ for all $t\ge t_1$ since $\mathbf{x}=\mathbf{0}$ is an equilibrium. Hence in the remainder we consider the case, where $\mathbf{x}\neq\mathbf{0}$ for all $t\ge t_0$.

Choose some $\bar{r}_0>0$, sufficiently small such that the ball $\mathcal{B}_{\bar{r}_0}(\mathbf{0})$ (see. Definition B.6) completely lies in $\mathcal{D}$. For readability, the argument in $\mathcal{B}_{\bar{r}_0}(\mathbf{0})$ is dropped in the following.

By Lemma C.2 there are class $\mathcal{K}$ functions $\alpha_1(x)$, $\alpha_2(x)$ on $[0,\bar{r}_0)$ such that $\alpha_1(\|\mathbf{x}\|)\leq W_1(\|\mathbf{x}\|)$ and $W_2(\|\mathbf{x}\|)\leq \alpha_2(\|\mathbf{x}\|)$. Then

$$\alpha_1\big(\|\mathbf{x}\|\big)\leq V(\mathbf{x},t)\leq \alpha_2\big(\|\mathbf{x}\|\big). \tag{C.16}$$

For $r\in(0,r_0]$, $r_0 < \bar{r}_0$, choose some constant

$$\beta=\alpha_1(r) \tag{C.17}$$

and the following set

$$\Omega_{\beta,t}=\left\{\mathbf{x}\in\overline{\mathcal{B}}_r \,\middle|\, V(\mathbf{x},t)<\beta\right\}$$

resides in the interior of $\overline{\mathcal{B}}_r$ (does not hit the boundary) and contains the origin. This is clear since, on the boundary, we have $V(\mathbf{x},t)\geq\alpha_1(r)=\beta$ but on $\Omega_{\beta,t}$, we have $V(\mathbf{x},t)<\beta$. Further $V(\mathbf{0},t)=0<\beta$ and thus $\mathbf{0}\in\Omega_{\beta,t}$. Hence, $\Omega_{\beta,t}$ can be thought of as a time varying region, which is residing inside $\overline{\mathcal{B}}_r$. The situation is depicted in Figure C.1. The region

$$\Theta_\beta=\left\{\mathbf{x}\in\overline{\mathcal{B}}_r \,\middle|\, \alpha_2\big(\|\mathbf{x}\|\big)<\beta\right\} \tag{C.18}$$

is not time varying and part of $\Omega_{\beta,t}$ at any time instant since, if $\alpha_2(\|\mathbf{x}\|)<\beta$ then also $V(\mathbf{x},t)<\beta$ from inequality (C.16). Θ_β is in fact an open ball around the origin of radius,

$$\delta(r)=\alpha_2^{-1}(\beta)=\alpha_2^{-1}(\alpha_1(r)) \tag{C.19}$$

i.e. $\Theta_\beta=\mathcal{B}_\delta$. Further $\Omega_{\beta,t}$ is an invariant set meaning that, if $\mathbf{x}_0$ is in $\Omega_{\beta,t}$ then $\mathbf{x}(t)$ will stay in $\Omega_{\beta,t}$ for all times. Since, if $\mathbf{x}_0\in\Omega_{\beta,t}$, then $V(\mathbf{x}_0,t_0)<\beta$, and $\dot{V}(\mathbf{x},t)\leq 0$ on $\Omega_{\beta,t}\backslash\mathbf{0}$ implies that the trajectory can never hit the boundary of $\Omega_{\beta,t}$ since then $V(\mathbf{x},t)=\beta$. (Note that case $\mathbf{x}=0$ has been treated separately).

Moreover, if $\|\mathbf{x}_0\|<\delta$ then the trajectory starts within the time invariant set Θ_β which in turn lies within the time varying set $\Omega_{\beta,t}$. From the fact that $\Omega_{\beta,t}$ lies in the interior of $\overline{\mathcal{B}}_r$, it follows that $\|\mathbf{x}(t)\|<\varepsilon$ for all $t\geq t_0$ and for any $\varepsilon\geq r$, which proves uniform stability.

So far, we found a δ for a fixed r or ε. Of practical interest is, how far we can start away from the equilibrium such that the system states remain stable. In other words, what is the maximum δ such that the trajectories remain bounded. We have found δ for $r\leq r_0$, however, the stability definition requires a δ for every $\varepsilon > 0$. In order to achieve this, we define

$$r(\varepsilon)=\min(\varepsilon,r_0) \tag{C.20}$$

In view of equation (C.19), we obtain $\delta(\varepsilon)=\alpha_1^{-1}(\beta(r(\varepsilon)))$, and

$$\delta_\infty := \lim_{\varepsilon\to\infty}(\delta(\varepsilon)) = \alpha_2^{-1}(\alpha_1(r_0)) = \delta(r_0) \tag{C.21}$$

Spoken descriptive, as long as $||\mathbf{x}_0||<\delta_\infty$ it is guaranteed that $||\mathbf{x}(t)||<r_0$ for all $t\geq t_0$. δ_∞ is in fact the stability limit that is obtained from these considerations and $\delta(\varepsilon)=\delta_\infty$ for all $\varepsilon\geq r_0$. Note also that necessarily $\delta_\infty\leq r_0$ since otherwise it might occur that $\mathbf{x}_0\notin\mathcal{D}$ which undoes all previous arguments.

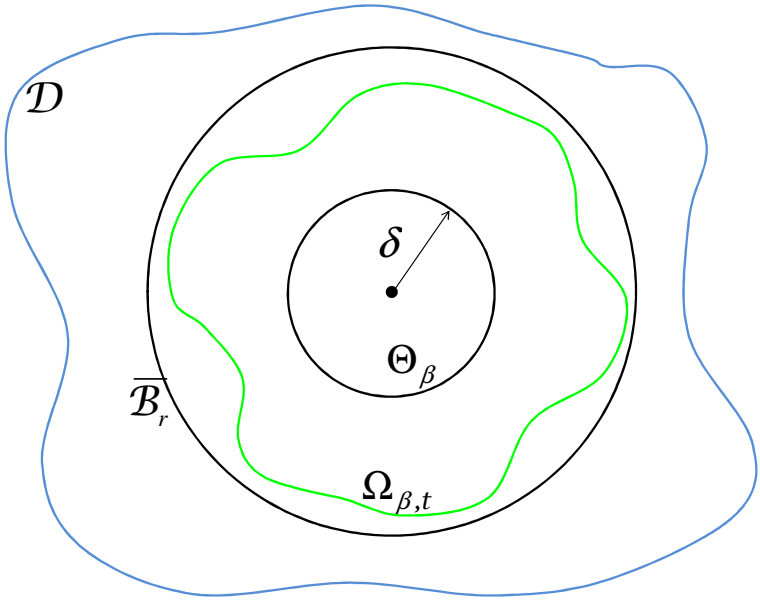


Figure C.1 Domains used in Lyapunov Stability Proof

Uniform Asymptotic Stability

For the case that $\mathbf{x}(t_1)$ equals $\mathbf{0}$ for some $t_1\geq t_0$, $\mathbf{x}(t)$ will remain $\mathbf{0}$ for all $t\geq t_1$ since $\mathbf{x}=\mathbf{0}$ is an equilibrium. Hence, in the remainder we consider the case, where $\mathbf{x}(t)\neq\mathbf{0}$ for all $t\geq t_0$.

For $\|\mathbf{x}_0\|<\delta_\infty$, $\mathbf{x}(t)$ will stay in $\overline{\mathcal{B}}_{r0}\subset\mathcal{D}$ and hence $\dot{V}(\mathbf{x}(t),t)\leq W_3(x)\leq 0$ for all $t\geq t_0$. Consequently $V(\mathbf{x}(t),t)$ is decreasing with t, and is bounded from below by 0. This implies that V has a limit $c\geq 0$ as t goes to infinity as will be shown next.

For $k\in\mathbb{N}$, define a sequence of time instances $t_k=t_0+k$ and a sequence $x_k=V(\mathbf{x}(t_k),t_k)$. This sequence is bounded from below by 0 and decreasing. Therefore, it converges to some $c\geq 0$ by Theorem B.7. This implies that for every $\varepsilon>0$ there is some N such that

$$V(\mathbf{x}(t_N),t_N) = x_N < c+\varepsilon$$

Since $V(\mathbf{x}(t),t)$ is nonincreasing it follows that also

$$V(\mathbf{x}(t),t) < c+\varepsilon$$

for all $t\geq t_N$. On the other hand, also

$$V(\mathbf{x}(t),t) \geq c$$

for all $t \geq t_0$ since, assume that there is a τ such that $V(\mathbf{x}(t),t)<c$ then, as $V(\mathbf{x}(t),t)$ is nonincreasing, there is some M and t_M of the sequence defined recently with $t_M \geq \tau$ such that $x_M=V(\mathbf{x}(t_M),t_M)<c$ which is a contradiction, since all $x_k \geq c$. Hence for every $\varepsilon>0$, there exists an N such that

$$\left|V(\mathbf{x}(t),t)-c\right|<\varepsilon$$

for all $t \geq t_N$ and hence

$$\lim_{t\to\infty}(V(\mathbf{x}(t),t))=c\geq 0$$

Next it is shown by contradiction that $c=0$. Therefore assume $c>0$. Then $V(\mathbf{x},t)\geq c$, in light of (C.16), implies that

$$\|\mathbf{x}(t)\| \geq \alpha_2^{-1}(c) \tag{C.22}$$

Employing Lemma C.2 there is a class $\mathcal{K}$ function $\alpha_3(x)$ such that

$$\alpha_3(\|\mathbf{x}\|)\leq W_3(\mathbf{x})$$

For $||\mathbf{x}_0||\in[0,r_0]$ and , by(C.22) we arrive at

$$W_3(\mathbf{x})\geq \alpha_3(\alpha_2^{-1}(c))=:\gamma(c) \tag{C.23}$$

where γ is also a class $\mathcal{K}$ function and hence $\gamma(c)>0$ if $c>0$. Then, (C.23) implies

$$\dot{V}(\mathbf{x}(t),t)\leq -\gamma(c).$$

If $V(\mathbf{x},t)\geq c$, integration of $\dot{V}$ yields

$$V(\mathbf{x}(t),t)-V(\mathbf{x}_0,t_0)=\int_{t_0}^{t}\dot{V}(\mathbf{x}(\tau),\tau)d\tau\leq -\gamma(c)(t-t_0)$$

Furthermore by (C.16), $V(\mathbf{x}(t_0),t_0)\leq\alpha_2(\|\mathbf{x}_0\|)$ and hence

$$V(\mathbf{x}(t),t)\leq \alpha_2(\|\mathbf{x}(t_0)\|)-\gamma(c)(t-t_0) \tag{C.24}$$

and, for some $t_1\geq t_0+\gamma^{-1}\alpha_2(\|\mathbf{x}_0\|)$ we have $V(\mathbf{x}(t_1),t_1)\leq 0$ which contradicts the assumption $c>0$ and hence $c=0$. In order to find a $T(\eta,\delta)$, introduced in the definition of asymptotic stability (Definition C.2), we first restrict the initial condition

$$\|\mathbf{x}_0\|<\delta(r) \tag{C.25}$$

where δ is defined in equation (C.19) for $r\in(0,r_0]$ and $\rho=\delta(r_0)=\delta_\infty$. Figure C.2 depicts the situation.

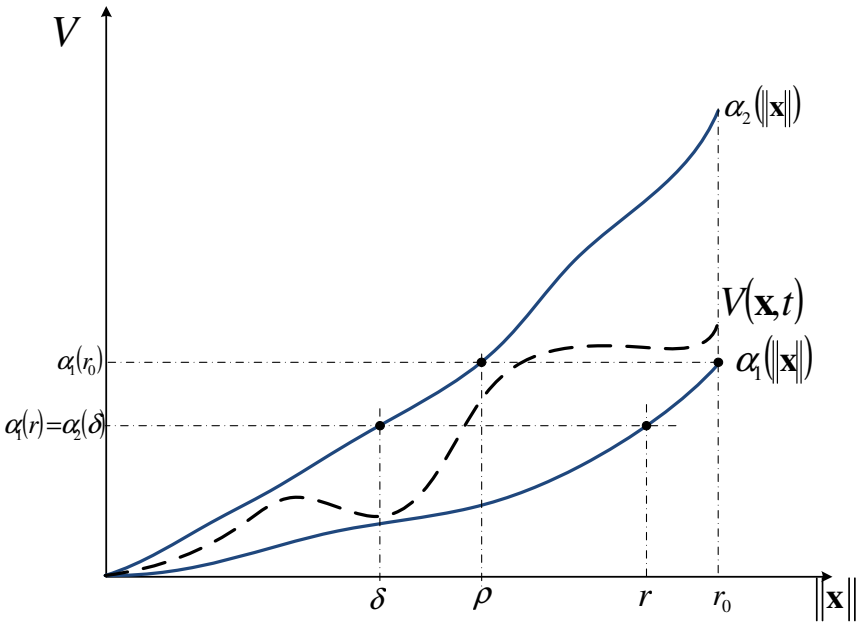


Figure C.2 Lyapunov Asymptotic Stability

It is clear from first part of this proof, that $\|\mathbf{x}(t)\|<r$ for all $t\geq t_0$, if $\|\mathbf{x}_0\|<\delta(r)$. Next, it will be shown that, for any $d>0$ and $\delta(r)$, there is a $\hat{T}(d,\delta)$ such that $V(\mathbf{x}(t),t)<d$ for all $t\geq t_0+\hat{T}(d,\delta)$ and $\|\mathbf{x}_0\|<\delta$. Therefore note that $\hat{T}(d,\delta)=0$ for $d\geq\alpha_2(\delta)$ $(=\alpha_1(r))$ since, by constraint (C.25) and inequality (C.16), it is clear that $V(\mathbf{x}_0,t_0)<\alpha_2(\delta)$ and $\dot{V}(\mathbf{x}(t),t)\leq 0$ implies that $V(\mathbf{x},t)<\alpha_2(\delta)$ for all $t\geq t_0$. It is left to investigate the case $d<\alpha_2(d)$. In this case, for some $t_1\geq t_0$, there is an interval $[t_0,t_1)$ such that $V(\mathbf{x}(t),t)\geq d$. By (C.24) and condition (C.25), we obtain

$$V(\mathbf{x}(t),t)<\alpha_2(\delta)-\gamma(d)(t_1-t_0)\overset{!}{\geq}d$$

one obtains that for

$$t_1=t_0+\frac{\alpha_2(\delta)-d}{\gamma(d)}$$

we have

$$V(\mathbf{x}(t),t)<d$$

for all $t\geq t_1$ and hence we find

$$\hat{T}(d,\delta)=\frac{\alpha_2(\delta)-d}{\gamma(d)}=\frac{\alpha_2(\delta)-d}{\alpha_3(\alpha_2^{-1}(d))} \tag{C.26}$$

If $V(\mathbf{x}(t),t)<d$ then (C.16) implies that $\|\mathbf{x}(t)\|<\eta(d)$,where

$$\eta(d):=\alpha_1^{-1}(d)\text{ for }d\in(0,\alpha_1(r)]$$

is a class $\mathcal{K}$ function and hence $d(\eta):=\alpha_1(\eta)$, which is defined on $(0,r]$. Then by (C.26) there is a

$$T(\eta,\delta) := \hat{T}(d(\eta),\delta) = \frac{\alpha_2(\delta)-\alpha_1(\eta)}{\alpha_3\left(\alpha_2^{-1}(\alpha_1(\eta))\right)} \tag{C.27}$$

for any $\eta \in (0, r]$ such that $||\mathbf{x}(t)|| < \eta$ for all $t \geq t_0 + T(\eta,\delta)$. For $\eta = r$ and $\alpha_1(r) = \alpha_2(\delta)$ we obtain

$$T(r,\delta) = \frac{\alpha_1(r)-\alpha_1(r)}{\alpha_3\left(\alpha_2^{-1}(\alpha_1(r))\right)} = 0$$

Which is clear from the first part of the proof, since $\delta(r)$ is chosen such that $||\mathbf{x}(t)|| < r$ for all $t \geq t_0$ and thus for $\eta > r$ we choose $T(\eta,\delta) = 0$. Summing up

- for $0 < \eta \leq r$: $T(\eta,\delta)$ according to (C.27)
- for $\eta > r$: $T(\eta,\delta) = 0$

for every $\delta \in (0, \rho]$ and uniform asymptotic stability is proved.

Global Uniform Asymptotic Stability

If $\mathcal{D} = \mathbb{R}^n$ then $r_0 = \infty$ and $r \in (0,\infty)$. Consequently, from (C.20) $r = \varepsilon$ for all $r > 0$ and

$$\lim_{\varepsilon \to \infty}(r(\varepsilon)) = \infty \tag{C.28}$$

If further $W_1(x)$ is radially unbounded then $\alpha_1(\cdot)$ can be chosen to belong to class $\mathcal{K}_\infty$, $\beta(r)$ as defined in (C.17) becomes arbitrarily large for a sufficiently large r, in other words:

$$\lim_{r \to \infty}(\beta(r)) = \infty \tag{C.29}$$

Also $W_2(x)$ is radially unbounded from condition 1 of the theorem. For (C.16) and condition 2 of the theorem, $\alpha_2(\cdot)$ belongs to class $\mathcal{K}_\infty$, too which implies that also their inverses $\alpha_1^{-1}(\cdot), \alpha_2^{-1}(\cdot)$ belong to class $\mathcal{K}_\infty$. Then with δ as defined in (C.19), we have

$$\delta_\infty = \lim_{\beta \to \infty}(\delta(\beta)) = \lim_{\varepsilon \to \infty}\left(\alpha_2^{-1}(\alpha_1(\varepsilon))\right) = \infty \,. \tag{C.30}$$

which proves global uniform stability.

□

Lyapunov's direct method, as presented so far, only guarantees stability and asymptotic stability respectively. But there is no statement on how fast the equilibrium will be approached. It is an accepted opinion that Lyapunov asymptotic stability only guarantees that the trajectory will converge to the equilibrium, but not how fast this will occur. However, in the proof we have computed time dependent bounds for the

trajectory by introducing class $\mathcal{K}$ functions that bound the Lyapunov function as well as its time derivative. This result will be summarized in the subsequent corollary.

Global uniform asymptotic stability follows from the fact that $\rho=\delta_\infty=\infty$ and the derivations for uniform asymptotic stability are valid for al $\delta\in(0,\infty)$.

□

Theorem C.1 does not specify explicit values on the rate of convergence, i.e. explicit values for $T(\eta,\delta)$. However, the recent proof reveals that, explicit bounds can be computed, if the Lyapunov function and its time derivative are enclosed by class $\mathcal{K}$ functions, which is summarized in the following.

Corollary C.1 Lyapunov's Direct Method with Explicit Bounds

Consider system (C.1). Let

- $V:\mathcal{D}\times\mathbb{R}_0^+\to\mathbb{R}_0^+$ be a continuously differentiable function in all arguments.
- some $r_0>0$, and the closed ball $\overline{\mathcal{B}_{r_o}}=\left\{\mathbf{x}\in\mathbb{R}^n\,\middle|\,\|\mathbf{x}\|\le r_0\right\}$ be part of $\mathcal{D}$.
- class $\mathcal{K}$ functions $\alpha_1(\cdot)$, $\alpha_2(\cdot)$, $\alpha_3(\cdot)$ and for all $\|\mathbf{x}\|\in[0,r_0]$ such that

1. $\alpha_1(\|\mathbf{x}\|)\le V(\mathbf{x},t)\le\alpha_2(\|\mathbf{x}\|)$
2. $\dfrac{dV}{dt}=\dfrac{\partial V}{\partial t}+\dfrac{\partial V}{\partial\mathbf{x}}\mathbf{f}(\mathbf{x},t)\le-\alpha_3(\|\mathbf{x}\|)$

then the equilibrium of (C.1) is uniformly asymptotically stable

such that $\|\mathbf{x}(t)\|<\eta$ for all $t\ge t_0+T(\eta,\rho)$ if $\|\mathbf{x}(t_0)\|<\delta$, $\delta\in(0,\rho]$

for all $r\in(0,r_0]$, $\rho=\alpha_2^{-1}(\alpha_1(r_0))$, $\eta>0$ where $T(\eta,\delta)=\max\left[0,\dfrac{\alpha_2(\delta)-\alpha_1(\eta)}{\alpha_3(\alpha_2^{-1}(\alpha_1(\eta)))}\right]$

If the class $\mathcal{K}$ functions are monomials, even exponential stability can be shown.

Theorem C.2 Lyapunov Exponential Stability [Kha02]

Consider system (C.1). Let

- $V:\mathcal{D}\times\mathbb{R}_0^+\to\mathbb{R}_0^+$ be a continuously differentiable function in all arguments.
- some $r_0>0$, and the closed ball $\overline{\mathcal{B}_{r_0}}=\left\{\mathbf{x}\in\mathbb{R}^n\,\middle|\,\|\mathbf{x}\|\le r_0\right\}$ be part of $\mathcal{D}$.
- positive constant k_1, k_2, k_3 and some $\alpha>1$ such that, for $\|\mathbf{x}\|\in[0,r_0]$

1. $k_1\cdot\|\mathbf{x}\|^\alpha\le V(\mathbf{x},t)\le k_2\cdot\|\mathbf{x}\|^\alpha$
2. $\dfrac{dV}{dt}=\dfrac{\partial V}{\partial t}+\dfrac{\partial V}{\partial\mathbf{x}}\cdot\mathbf{f}(\mathbf{x},t)\le-k_3\cdot\|\mathbf{x}\|^\alpha$

Then the equilibrium of (C.1) is exponentially stable with convergence rate

$$\lambda = \frac{k_3}{\alpha \cdot k_2}$$

and $\mathbf{x}(t) \le \|\mathbf{x}_0\| \cdot \left(\frac{k_2}{k_1}\right)^{\frac{1}{\alpha}} \cdot e^{-\lambda(t-t_0)}$ if $\|\mathbf{x}_0\| < \left(\frac{k_1}{k_2}\right)^{\frac{1}{\alpha}} \cdot r_0$

If $\mathcal{D}=\mathbb{R}^n$, then the equilibrium is globally exponentially stable.

Proof

At first, note that

$$\|\mathbf{x}(t_0)\| < \rho = \left(\frac{k_1}{k_2}\right)^{\frac{1}{\alpha}} r_0$$

implies $V(\mathbf{x}(t),t) \le V(\mathbf{x}_0,t_0) < k_1 r_0^{\alpha}$ and consequently $\|\mathbf{x}(t)\| < r_0$ for all $t \ge t_0$. Hence with this initial condition constraint, the trajectories will not leave the definition set. It can be further conclude that

$$-\|\mathbf{x}\|^{\alpha} \le -\frac{V(\mathbf{x},t)}{k_2} \text{ and } \dot{V}(\mathbf{x},t) \le -k_3\|\mathbf{x}\|^{\alpha}$$

which results in

$$\frac{dV(\mathbf{x},t)}{dt} \le -k_3\|\mathbf{x}\|^{\alpha} \le -\frac{k_3}{k_2}V(\mathbf{x},t)$$

and Lemma C.3 (Comparison Lemma) implies that

$$V(\mathbf{x},t) \le V(\mathbf{x}_0,t_0) \cdot e^{-\frac{k_3}{k_2}(t-t_0)}$$

It can be concluded from inequality 1 of the theorem that

$$\|\mathbf{x}(t)\| \le \left(\frac{V(\mathbf{x}_0,t_0) \cdot e^{-\frac{k_3}{k_2}(t-t_0)}}{k_1}\right)^{\frac{1}{\alpha}} \le \left(\frac{k_2 \cdot \|\mathbf{x}_0\|^{\alpha} \cdot e^{-\frac{k_3}{k_2}(t-t_0)}}{k_1}\right)^{\frac{1}{\alpha}} = \left(\frac{k_2}{k_1}\right)^{\frac{1}{\alpha}} \cdot \|\mathbf{x}_0\| \cdot e^{-\frac{k_3}{\alpha \cdot k_2}(t-t_0)}$$

which establishes exponential convergence with convergence rate $\lambda = \frac{k_3}{\alpha \cdot k_2}$.

□

C.5 Lyapunov's Indirect Method

It is common to analyze the stability of a setpoint of a nonlinear system by investigation of a substitute system that is linearized at the setpoint. We restrict ourselves the autonomous case of the form.

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}) \ , \quad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{C.31}$$

Where $\mathbf{x}(t) \in \mathcal{D} \subset \mathbb{R}^n$, $\mathbf{f}: \mathcal{D} \rightarrow \mathbb{R}^n$ is a continuously differentiable function, and $\mathbf{x}=\mathbf{0}$ is an equilibrium, i.e. $\mathbf{f}(\mathbf{0})=\mathbf{0}$. The substituted linearized system is

$$\dot{\mathbf{x}}(t) = \mathbf{A} \cdot \mathbf{x}(t) \ , \quad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{C.32}$$

where

$$\mathbf{A} = \left. \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{0}}$$

The reason why it is justified to conclude local stability of the nonlinear system from asymptotic stability of the substituted linearized system is known as Lyapunov's indirect method. Before establishing this, a preliminary lemma is necessary.

Lemma C.4 Existence of a Solution to the Lyapunov Equation [Kha02]

Consider system (C.32).

The matrix $\mathbf{A}$ is Hurwitz, i.e. for all eigenvalues λ_i of $\mathbf{A}$, $\mathrm{Re}(\lambda_i)<0$ if and only if

for any symmetric positive definite matrix $\mathbf{Q} \in \mathbb{R}^{n \times n}$ the Lyapunov equation

$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q}$ has a symmetric positive definite solution.

Moreover $\mathbf{P}$ is the unique solution and

$$\|\mathbf{x}(t)\|_2 \leq \|\mathbf{x}_0\|_2 \sqrt{\frac{\bar{\lambda}_P}{\underline{\lambda}_P}} \exp\left(-\frac{\underline{\lambda}_Q}{2\bar{\lambda}_P} \cdot (t - t_0) \right)$$

Proof

SUFFICIENT PART:
"*If the Lyapunov equation has a positive definite solution, then* $\mathbf{A}$ *is Hurwitz*".

We define the positive definite Lyapunov function candidate

$$V(\mathbf{x}) = \mathbf{x}^T\mathbf{P}\mathbf{x}$$

and take the time derivative along the system trajectories.

$$\begin{aligned} \dot{V} &= \dot{\mathbf{x}}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\dot{\mathbf{x}} \\ &= \mathbf{x}^T\left(\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A}\right)\mathbf{x} \\ &= -\mathbf{x}^T\mathbf{Q}\mathbf{x} < 0 \end{aligned}$$

Further, note that

$$\underline{\lambda}_P \|\mathbf{x}\|_2^2 \le V(\mathbf{x}) \le \overline{\lambda}_P \|\mathbf{x}\|_2^2$$

and

$$\dot{V}(\mathbf{x}) \le \underline{\lambda}_Q \|\mathbf{x}\|_2^2 .$$

which follows from Theorem B.20. By Theorem C.2, system (C.32) is exponentially stable

$$\|\mathbf{x}(t)\|_2 \le \|\mathbf{x}_0\|_2 \sqrt{\frac{\overline{\lambda}_P}{\underline{\lambda}_P}} \exp\left(-\frac{\underline{\lambda}_Q}{2\overline{\lambda}_P} \cdot (t - t_0) \right) \ , \ \lambda = \frac{\underline{\lambda}_Q}{2\overline{\lambda}_P}$$

This however is possible if and only if for all eigenvalues λ_i of $\mathbf{A}$, $\mathrm{Re}(\lambda_i) < 0$ which proves the claim.

NECESSARY PART

"*If* $\mathbf{A}$ *is Hurwitz, the Lyapunov equation has a positive definite solution.*"

Consider therefore the following integral

$$\mathbf{P} = \int_0^\infty e^{\mathbf{A}^T t} \mathbf{Q} \mathbf{e}^{\mathbf{A} t} dt \qquad \text{(C.33)}$$

Clearly, $\mathbf{P}$, if it exists, is symmetric. It has to be shown that $\mathbf{P}$ exists and is positive definite. Therefore, consider the Jordan decomposition of $\mathbf{A}$

$$\mathbf{A} = \mathbf{V} \mathbf{J} \mathbf{V}^{-1}$$

where the columns of $\mathbf{V}$ contain the eigenvectors and generalized eigenvectors of $\mathbf{A}$ and $\mathbf{J}$ is a block diagonal matrix, whose diagonal elements are Jordan blocks.

$$\mathbf{J}_i = \begin{bmatrix} \lambda_i & 1 & & \mathbf{0} \\ 0 & \lambda_i & & \\ \vdots & \ddots & & 1 \\ 0 & \cdots & 0 & \lambda_i \end{bmatrix}$$

By recalling the definition of the matrix exponential and inserting the Jordan decomposition one obtains

$$\begin{aligned} e^{\mathbf{A}t} &= \mathbf{V} \left\{ \mathbf{I} + \mathbf{J}t + \mathbf{J}^2 \frac{t^2}{2} + \ldots + \mathbf{J}^k \frac{t^k}{k} + \ldots \right\} \mathbf{V}^{-1} \\ &= \mathbf{V} \cdot e^{\mathbf{J}t} \cdot \mathbf{V}^{-1} \end{aligned}$$

where $e^{\mathbf{J}t}$ has block diagonal structure with diagonal elements

$$e^{\mathbf{J}_i t} = \begin{bmatrix} e^{\lambda_i t} & te^{\lambda_i t} & t^2 e^{\lambda_i t} & \cdots \\ 0 & e^{\lambda_i t} & te^{\lambda_i t} & \cdots \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & 0 & e^{\lambda_i t} \end{bmatrix}$$

Hence the integral definition of $\mathbf{P}$ is

$$\mathbf{P} = \mathbf{V}^{-T} \cdot \left(\int_0^\infty e^{\mathbf{J}t} \mathbf{V}^T \mathbf{Q} \mathbf{V} e^{\mathbf{J}t} dt \right) \cdot \mathbf{V}^{-1}. \tag{C.34}$$

Since $\mathrm{Re}(\lambda_i)<0$, one can conclude that the integral is finite and hence $\mathbf{P}$ exists. Further, define

$$\mathbf{y}(t) = \mathbf{V} e^{\mathbf{J}t} \mathbf{V}^{-1} \mathbf{x}$$

for some $\mathbf{x} \in \mathbb{R}^n$ and note that $\mathbf{y}(t) \neq 0$ if $\mathbf{x} \neq 0$ since $\mathbf{V}$ (since its columns contain the linearly independent eigenvectors) and $e^{\mathbf{J}t}$ (due to its diagonal structure) are regular for all t. Then (C.34) becomes.

$$\mathbf{x}^T \mathbf{P} \mathbf{x} = \left(\int_0^\infty \mathbf{y}^T(t) \cdot \mathbf{Q} \cdot \mathbf{y}(t) dt \right)$$

Now, since $\mathbf{Q}$ is positive definite, the integrand is positive if $\mathbf{x} \neq 0$ which renders $\mathbf{P}$ positive definite. It is only left to show that $\mathbf{P}$ satisfies the Lyapunov equation. Therefore, consider the expression

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} = \int_0^\infty \left(\mathbf{A}^T e^{\mathbf{A}^T t} \mathbf{Q} e^{\mathbf{A}t} + e^{\mathbf{A}^T t} \mathbf{Q} e^{\mathbf{A}t} \mathbf{A} \right) dt$$

and recall the fundamental property of the matrix exponential

$$\frac{d\left(e^{\mathbf{A}t}\right)}{dt} = \mathbf{A} \cdot e^{\mathbf{A}t} = e^{\mathbf{A}t} \cdot \mathbf{A}$$

which results in

$$\mathbf{A}^T \mathbf{P} + \mathbf{P} \mathbf{A} = \int_0^\infty \frac{d}{dt} \left(e^{\mathbf{A}^T t} \mathbf{Q} e^{\mathbf{A}t} \right) dt = \left[e^{\mathbf{A}^T t} \mathbf{Q} e^{\mathbf{A}t} \right]_0^\infty = -\mathbf{Q}.$$

Uniqueness of $\mathbf{P}$ is shown by assuming that $\tilde{\mathbf{P}}$ also satisfies the Lyapunov equation. Then

$$\mathbf{A}^T \left(\mathbf{P} - \tilde{\mathbf{P}} \right) + \left(\mathbf{P} - \tilde{\mathbf{P}} \right) \mathbf{A} = \mathbf{0} \tag{C.35}$$

and after pre and post multiplying (C.35) with $e^{\mathbf{A}^T t}$ and $e^{\mathbf{A}t}$ respectively one obtains

$$\mathbf{0} = e^{\mathbf{A}^T t} \mathbf{A}^T \left(\mathbf{P} - \tilde{\mathbf{P}} \right) e^{\mathbf{A}t} + e^{\mathbf{A}^T t} \left(\mathbf{P} - \tilde{\mathbf{P}} \right) \mathbf{A} e^{\mathbf{A}t} = \frac{d}{dt} \left(e^{\mathbf{A}^T t} \left(\mathbf{P} - \tilde{\mathbf{P}} \right) \mathbf{A} e^{\mathbf{A}t} \right).$$

Obviously

$$e^{\mathbf{A}^T t}\left(\mathbf{P}-\tilde{\mathbf{P}}\right)e^{\mathbf{A}t}=\text{const.}$$

for all t. Inserting $t=0$ and $t=\infty$ finally yields $\mathbf{P}-\tilde{\mathbf{P}}=\mathbf{0}$ which proves uniqueness of $\mathbf{P}$.

□

Now preparation for Lyapunov's indirect method has been done.

Theorem C.3 Lyapunov's Indirect Method [Kha02]

The equilibrium of (C.31) is locally asymptotically stable, if the matrix $\mathbf{A}$ of the linearized substitute (C.32) is Hurwitz.

Proof

A Lyapunov function candidate is given by

$$V(\mathbf{x})=\mathbf{x}^T\mathbf{P}\mathbf{x}$$

where $\mathbf{P}$ satisfies $\mathbf{A}^T\mathbf{P}+\mathbf{P}\mathbf{A}=-\mathbf{Q}$ and is symmetric positive definite by Lemma C.4 and hence

$$\underline{\lambda}_P\|\mathbf{x}\|_2^2\le V(\mathbf{x})\le\overline{\lambda}_P\|\mathbf{x}\|_2^2$$

The time derivative along the trajectories of the nonlinear system (C.31) is

$$\dot{V}(\mathbf{x})=\mathbf{f}^T(\mathbf{x})\mathbf{P}\mathbf{x}+\mathbf{x}^T\mathbf{P}\mathbf{f}(\mathbf{x}).$$

By the mean value theorem

$$\mathbf{f}(\mathbf{x})=\mathbf{f}(\mathbf{0})+\left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}}\right|_{\mathbf{x}=\xi}\cdot\mathbf{x}$$

Where ξ is some point on the line connecting $\mathbf{0}$ and $\mathbf{x}$, i.e. $\xi=k\cdot\mathbf{x}$ for $k\in[0,1]$. Note that $\mathbf{f}(\mathbf{0})=\mathbf{0}$ and

$$\left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}}\right|_{\mathbf{x}=\xi}=\mathbf{A}+\left(\left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}}\right|_{\mathbf{x}=\xi}-\left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}}\right|_{\mathbf{x}=0}\right).$$

Therefore $\dot{V}$ becomes

$$\begin{aligned}\dot{V}(\mathbf{x})&=\mathbf{x}^T\left(\mathbf{A}^T\mathbf{P}+\mathbf{P}\mathbf{A}\right)\mathbf{x}+2\mathbf{x}^T\mathbf{G}(\xi)\mathbf{P}\mathbf{x}\\&=-\mathbf{x}^T\mathbf{Q}\mathbf{x}+2\mathbf{x}^T\mathbf{G}(\xi)\mathbf{P}\mathbf{x}\end{aligned}$$

Where we abbreviated

$$\mathbf{G}(\xi)=\left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}}\right|_{\mathbf{x}=\xi}-\left.\frac{\partial\mathbf{f}(\mathbf{x})}{\partial\mathbf{x}}\right|_{\mathbf{x}=0}$$

An upper bound for the second term of $\dot{V}$ is given by

$$\mathbf{x}^T\mathbf{G}(\xi)\mathbf{x}\le\|\mathbf{G}(\xi)\mathbf{P}\|_2\cdot\|\mathbf{x}\|_2^2$$

where $\|\mathbf{G}(\mathbf{x})\mathbf{P}\|_2$ is the induced matrix 2-norm. An upper bound for $\dot{V}$ is given by

$$V(\mathbf{x}) \leq \left(\underline{\lambda}_Q - 2\|\mathbf{G}(\xi)\mathbf{P}\|_2 \right) \cdot \|\mathbf{x}\|_2^2$$

Since $\mathbf{f}(\mathbf{x})$ is assumed to be continuously differentiable, we have

$$2\|\mathbf{G}(\xi)\mathbf{P}\|_2 < \underline{\lambda}_Q \quad \text{if} \quad \|\mathbf{x}\|_2 < r$$

for a sufficiently small $r>0$. Therefore, $\dot{V}(\mathbf{x})$ is bounded from above by a negative definite function on some set

$$\mathcal{D} = \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, \|\mathbf{x}\| < r \right\}$$

and Theorem C.1 assures local asymptotic stability of (C.31).

□

C.6 Converse Theorem

Lyapunov's direct method guarantees (uniform asymptotic) stability from the existence of a Lyapunov function. In nonlinear control, it is sometimes known, that parts of the whole system are asymptotically stable. In order to proof stability of the overall system, it is necessary to know about the existence of a Lyapunov function for the partial system. Such theorems are known as converse theorems [Kha02], as it assures the existence of a Lyapunov function from the knowledge of asymptotic stability of a dynamic system. The proof is done by construction of such Lyapunov functions, but in the general, it is not possible to specify the Lyapunov function explicitly.
There are some lemmas required in the proof of the theorem, nonetheless the converse theorem is stated first and the required preliminaries are presented afterwards. However, it is recommendable to read Lemma C.5 – Lemma C.7 at first in order to understand Theorem C.4.

Theorem C.4 Converse Lyapunov [Kha02]

Assume that $\mathbf{f}(\mathbf{x},t)$ in system (C.1) is continuously differentiable on $\overline{\mathcal{B}_r} = \left\{ \mathbf{x} \in \mathbb{R}^n \,\middle|\, \|\mathbf{x}\| \leq r \right\}$, where $r>0$ and sufficiently small such that $\overline{\mathcal{B}_r} \subset \mathcal{D}$.

Let further the Jacobian

$$\frac{\partial \mathbf{f}(\mathbf{x},t)}{\partial \mathbf{x}}$$

be bounded on $\overline{\mathcal{B}}_r$ uniformly in t.

If the system is uniformly asymptotically stable, then there exists a function $V : \mathcal{B}_\rho \times [0,\infty) \to \mathbb{R}_0^+$ continuously differentiable in all arguments which satisfies the following inequalities

1. $\alpha_1(\|\mathbf{x}\|) \leq V(\mathbf{x},t) \leq \alpha_2(\|\mathbf{x}\|)$
2. $\dot{V}(\mathbf{x},t) = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x},t) \leq -\alpha_3(\|\mathbf{x}\|)$

3. $\left\|\frac{\partial V(\mathbf{x},t)}{\partial \mathbf{x}}\right\| \le \alpha_4(\|\mathbf{x}\|)$

where $\mathcal{B}_\rho = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\| < \rho\}$ for some $\rho>0$

and ρ sufficiently small such that $\|\mathbf{x}_0\| < \rho$ implies $\|\mathbf{x}(t)\| < r$ for all $t \ge t_0$

and $\alpha_1(x), \alpha_2(x), \alpha_3(x), \alpha_4$ belong to class $\mathcal{K}$ on $x \in [0, \rho)$

If the system is autonomous, V can be chosen to be independent of t and inequality 2 simplifies to

$$\dot{V}(\mathbf{x}) = \frac{\partial V(\mathbf{x})}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x}) \le -\alpha_3(\mathbf{x})$$

Proof

First we change notation, by defining $\boldsymbol{\varphi}(\tau, \mathbf{x}, t)$ as the solution to (C.1) with initial value $\mathbf{x}$ and initial time t observed at time τ. By Lemma C.1 there is a class $\mathcal{KL}$ function $\beta(x,t)$ such that

$$\|\boldsymbol{\varphi}(\tau, \mathbf{x}, t)\|_2 \le \beta(\|\mathbf{x}\|_2, \tau - t) \qquad \text{(C.36)}$$

for all $\tau \ge t$ and $||\mathbf{x}|| < \rho$ where ρ is chosen such that $\beta(x,0) \le r$ for $x < \rho$. In the remainder of the proof, a Lyapunov function is constructed that satisfies the inequalities. The following Lyapunov function is suggested

$$V(\mathbf{x}, t) = \int_t^\infty f(\|\boldsymbol{\varphi}(\tau, \mathbf{x}, t)\|_2) d\tau \qquad \text{(C.37)}$$

where $f(\cdot)$ is some class $\mathcal{K}$ function to be determined.

Inequality 1

Since the Jacobian of $\mathbf{f}(\mathbf{x},t)$ is bounded on $\mathcal{B}_r$, uniformly in t, let us say

$$\left\|\frac{\partial \mathbf{f}(\mathbf{x},t)}{\partial \mathbf{x}}\right\|_2 \le L$$

the mean value theorem implies that

$$\|\mathbf{f}(\mathbf{x},t)\|_2 \le L\|\mathbf{x}\|_2$$

on $\mathcal{B}_r$ (recall that $\mathbf{f}(\mathbf{0},t) = \mathbf{0}$) for all $t \ge 0$). For admissible initial conditions, $||\mathbf{x}|| < \rho$, Lemma C.6 assures that

$$\|\boldsymbol{\varphi}(\tau, \mathbf{x}, t)\|_2 \ge \|\mathbf{x}\|_2 \cdot e^{-L(\tau - t)}. \qquad \text{(C.38)}$$

Inserting inequality (C.38) into the suggested Lyapunov function yields

$$V(\mathbf{x},t) \ge \int_t^\infty f(\|\mathbf{x}\|_2 e^{-L(\tau-t)}) d\tau = \int_0^\infty f(\|\mathbf{x}\|_2 e^{-L\bar{\tau}}) d\bar{\tau} \ge \int_0^1 f(\|\mathbf{x}\|_2 e^{-L}) d\bar{\tau} = f(\|\mathbf{x}\|_2 e^{-L}) =: \alpha_1(\|\mathbf{x}\|_2)$$

$\alpha_1(\cdot)$ belongs to class $\mathcal{K}$ since $f(\cdot)$ belongs to class $\mathcal{K}$ which proves the left part of inequality 1. In order to proof the right part, we use (C.36) and get

$$V(\mathbf{x},t) \le \int_t^\infty f(\beta(\|\mathbf{x}\|_2), \tau - t) d\tau = \int_0^\infty f(\beta(\|\mathbf{x}\|_2), \bar{\tau}) d\bar{\tau}$$

where we choose, in the notation of Lemma C.5 $u(\tau)=\beta(\|\mathbf{x}\|,\tau)$ and $g(\tau)=\beta(\|\mathbf{x}\|+\varepsilon,\tau)$, for some sufficiently small $\varepsilon>0$ such that $\|\mathbf{x}\|+\varepsilon<\rho$. Therefore $f(\cdot)$ can be chosen such that the integral is bounded, but dependent on $\|\mathbf{x}\|_2$:

$$V(\mathbf{x},t)\le\int_0^\infty f\left(\beta\left(\|\mathbf{x}_2\|,\bar{\tau}\right)\right)=:\alpha_2\left(\|\mathbf{x}\|_2\right)$$

$\alpha_2(\cdot)$ is clearly a class $\mathcal{K}$ function since it is continuous and, from the definition of class $\mathcal{KL}$ functions, $\beta(0,\tau)=0$ for all $\tau\ge 0$ and hence the integral equals to zero for $\mathbf{x}=0$. Also for some $\mathbf{x}_1,\mathbf{x}_2$ such that $\|\mathbf{x}_2\|_2>\|\mathbf{x}_1\|_2$, we have $\beta\left(\|\mathbf{x}_2\|_2,\tau\right)>\beta\left(\|\mathbf{x}_1\|_2,\tau\right)$ for all $\tau\ge 0$ and hence $\alpha_2\left(\|\mathbf{x}_2\|_2\right)>\alpha_2\left(\|\mathbf{x}_1\|_2\right)$.

Inequality 2

In order to proof the inequality 2 of the theorem, consider

$$\begin{aligned}\frac{\partial V(\mathbf{x},t)}{\partial t}+\frac{\partial V(\mathbf{x},t)}{\partial\mathbf{x}}\mathbf{f}(\mathbf{x},t)=&-f\left(\|\boldsymbol{\varphi}(t,\mathbf{x},t)\|_2\right)\\&+\int_t^\infty f'\left(\|\boldsymbol{\varphi}(\tau,\mathbf{x},t)\|_2\right)\frac{\boldsymbol{\varphi}^T(\tau,\mathbf{x},t)}{\|\boldsymbol{\varphi}(\tau,\mathbf{x},t)\|_2}\frac{\partial\boldsymbol{\varphi}(\tau,\mathbf{x},t)}{\partial t}d\tau\\&+\int_t^\infty f'\left(\|\boldsymbol{\varphi}(\tau,\mathbf{x},t)\|_2\right)\frac{\boldsymbol{\varphi}^T(\tau,\mathbf{x},t)}{\|\boldsymbol{\varphi}(\tau,\mathbf{x},t)\|_2}\frac{\partial\boldsymbol{\varphi}(\tau,\mathbf{x},t)}{\partial\mathbf{x}}\mathbf{f}(\mathbf{x},t)d\tau\\=&-f\left(\|\mathbf{x}\|_2\right)\\&+\int_t^\infty f'\left(\|\boldsymbol{\varphi}(\tau,\mathbf{x},t)\|_2\right)\frac{\boldsymbol{\varphi}^T(\tau,\mathbf{x},t)}{\|\boldsymbol{\varphi}(\tau,\mathbf{x},t)\|_2}\left\{\frac{\partial\boldsymbol{\varphi}(\tau,\mathbf{x},t)}{\partial t}+\frac{\partial\boldsymbol{\varphi}(\tau,\mathbf{x},t)}{\partial\mathbf{x}}\mathbf{f}(\mathbf{x},t)\right\}d\tau\end{aligned}$$

Note that the derivative w.r.t. t is taken in a similar manner as done for (C.51). We know from Lemma C.7 that the term in curly bracket equals zero and hence

$$\frac{\partial V(\mathbf{x},t)}{\partial t}+\frac{\partial V(\mathbf{x},t)}{\partial\mathbf{x}}\mathbf{f}(\mathbf{x},t)=-f\left(\|\mathbf{x}\|_2\right)$$

Since $f(\cdot)$ is a class $\mathcal{K}$ function the 2nd inequality is proved.

Inequality 3

For the 3rd inequality, consider

$$\begin{aligned}\left\|\frac{\partial V(\mathbf{x},t)}{\partial\mathbf{x}}\right\|_2&=\left\|\int_t^\infty f'\left(\|\boldsymbol{\varphi}(\tau,\mathbf{x},t)\|_2\right)\frac{\boldsymbol{\varphi}^T(\tau,\mathbf{x},t)}{\|\boldsymbol{\varphi}(\tau,\mathbf{x},t)\|_2}\frac{\partial\boldsymbol{\varphi}(\tau,\mathbf{x},t)}{\partial\mathbf{x}}d\tau\right\|_2\\&\le\int_t^\infty f'\left(\|\boldsymbol{\varphi}(\tau,\mathbf{x},t)\|_2\right)\left\|\frac{\partial\boldsymbol{\varphi}(\tau,\mathbf{x},t)}{\partial\mathbf{x}}\right\|_2 d\tau\end{aligned}\qquad\text{(C.39)}$$

With inequality (C.47) we obtain

$$\left\|\frac{\partial\boldsymbol{\varphi}(\tau,\mathbf{x},t)}{\partial\mathbf{x}}\right\|_2\le\sqrt{n}\left\|\frac{\partial\boldsymbol{\varphi}(t,\mathbf{x},t)}{\partial\mathbf{x}}\right\|_2 e^{L(\tau-t)}=\sqrt{n}\|\mathbf{I}\|_2 e^{L(\tau-t)}=\sqrt{n}e^{L(\tau-t)}$$

where we used equation (C.50). Using the latter result and (C.36) in (C.39) yields

$$\left\|\frac{\partial V(\mathbf{x},t)}{\partial\mathbf{x}}\right\|_2\le\int_t^\infty f'\left(\beta\left(\|\mathbf{x}\|_2,\tau-t\right)\right)\sqrt{n}\cdot e^{L(\tau-t)}d\tau=\int_0^\infty f'\left(\beta\left(\|\mathbf{x}\|_2,\bar{\tau}\right)\right)\sqrt{n}\cdot e^{L\bar{\tau}}d\tau$$

where we substituted $\bar{\tau}=t-t$. In terms of Lemma C.5, where we already defined $g(\tau)$, we now define

$$h(\tau)=\sqrt{n}e^{L\tau}.$$

The lemma assures that $f(\cdot)$ can be chosen such that the integral

$$\int_0^\infty f'\left(\beta\left(\|\mathbf{x}\|_2,\bar{\tau}\right)\right)h(\bar{\tau})d\bar{\tau}\le\alpha_4\left(\|\mathbf{x}\|_2\right)$$

is bounded, while the bound depends on $\|\mathbf{x}\|_2$ and $\alpha_4(\cdot)$ is a class $\mathcal{K}$ function for the same reasons as used for $\alpha_2(\cdot)$.

Autonomous Systems

If the system is autonomous, the solution only depends on $\tau-t$ than explicitly on τ and t and (C.37) becomes

$$V(\mathbf{x},t)=\int_t^\infty f\left(\boldsymbol{\varphi}(\tau-t,\mathbf{x})\right)d\tau$$

By changing the variable $\bar{\tau}=\tau-t$, the integral becomes independent of t and hence

$$V(\mathbf{x})=\int_0^\infty f\left(\boldsymbol{\varphi}(\bar{\tau},\mathbf{x})\right)d\bar{\tau}$$

□

The next three Lemmas are preliminary to Theorem C.4

Lemma C.5 Massera's Lemma [Kha02]

Let $g:[0,\infty)\to\mathbb{R}$ be a positive, continuous, strictly decreasing function with

$$\lim_{t\to\infty}(g(t))=0$$

Let $h:[0,\infty)\to\mathbb{R}$ be a positive, continuous nondecreasing function.

Then there exists as function $f(t)$ such that

1. $f(t)$ and its derivative $f'(t)$ belong to class $\mathcal{K}$
2. For any continuous function, satisfying $0\le u(t)\le g(t)$ for all $t\ge 0$ there are constants $k_1,k_2>0$ such that

$$\int_0^\infty f(u(t))dt\le k_1 \qquad \int_0^\infty f'(\tau)h(\tau)d\tau\le k_2$$

Proof

The proof is done by construction of such a function $f(t)$. Choose a sequence $\{t_n\}$, $n=1,2,\ldots$, such that

$$g(t_n)\le\frac{1}{n+1}.$$

Such a sequence exists since $g(t)$ is continuous and strictly decreasing. Using that same sequence we define a function $\eta(t)$, $t\in(0,\infty)$, such that

- $\eta(t_n)=\frac{1}{n}$
- $\eta(t)$ is linear between t_n and t_{n+1},
- for $0<t\le t_1$, $\eta(t)=\frac{t_1}{t}$

The function is continuous, strictly decreasing and $g(t)<\eta(t)$. Since $\eta(t)$ is strictly decreasing, the inverse $\eta^{-1}(t)$ exists, is continuous, strictly decreasing and

$$\lim_{t\to 0^+}\left(\eta^{-1}(t)\right)=\infty \tag{C.40}$$

For $t\ge t_1$ we have the following order

$$u(t)\le g(t)<\eta(t) \tag{C.41}$$

And, since $\eta^{-1}(t)$ is strictly decreasing, we have

$$\eta^{-1}(u(t))\ge\eta^{-1}(g(t))>\eta^{-1}(\eta(t))=t \tag{C.42}$$

At next, define

$$f'(t)=\begin{cases}0 & \text{for} \quad t=0\\ \dfrac{\exp\left(-\eta^{-1}(t)\right)}{h\left(\eta^{-1}(t)\right)} & \text{for} \quad t>0\end{cases} \tag{C.43}$$

$$f(t)=\int_0^t f'(\tau)d\tau \tag{C.44}$$

It follows from the definition of $f'(t)$ that

- $\lim_{t\to 0^+}(f'(t))=0$ which implies that $f'(t)$ is continuous on $[0,\infty)$.
- $f'(t)$ is strictly increasing

Hence, $f'(t)$ and $f(t)$ belong to class $\mathcal{K}$. In order to show the first inequality of the lemma note that, for $0<\tau\le\eta(t)$, we have $-\eta^{-1}(\tau)\le -t$ and $h(0)\ge h\left(\eta^{-1}(\tau)\right)$. Hence, with (C.41)

$$f(u(t))=\int_0^{u(t)} f'(\tau)d\tau=\int_0^{u(t)}\frac{\exp\left(-\eta^{-1}(\tau)\right)}{h\left(\eta^{-1}(\tau)\right)}d\tau\le\int_0^{\eta(t)}\frac{\exp\left(-\eta^{-1}(\tau)\right)}{h\left(\eta^{-1}(\tau)\right)}d\tau\le\int_0^{\eta(t)}\frac{\exp(-t)}{h(0)}d\tau=\frac{\exp(-t)}{h(0)}\eta(t)$$

If we assume $t\ge t_1$, then $h(t)\le 1$ and hence

$$f(u(t))\le\frac{e^{-t}}{h(0)} \tag{C.45}$$

Then

$$\int_0^\infty f(u(\tau))=\int_0^{t_1} f(u(\tau))d\tau+\int_{t_1}^\infty f(u(\tau))d\tau\le\int_0^{t_1} f(u(\tau))d\tau+\int_{t_1}^\infty\frac{e^{-\tau}}{h(0)}d\tau$$

The first integral is bounded, since it is an integral of a bounded function over a bounded interval, let us say $f(u(\tau)) \le M$ for $0 < \tau \le t_1$. The second integral can be computed explicitly and hence

$$\int_0^\infty f(u(\tau))d\tau \le Mt_1 + \frac{e^{-t_1}}{h(0)} =: k_1$$

The second inequality is proved by using (C.42) and noting that $\exp(\cdot)$ is strictly decreasing and $h(\cdot)$ is nondecreasing. Therefore

$$f'(u(t)) \cdot h(t) = \frac{\exp\left(-\eta^{-1}(u(t))\right)}{h\left(\eta^{-1}(u(t))\right)} h(t) \le \frac{\exp\left(-\eta^{-1}(\eta(t))\right)}{h\left(\eta^{-1}(\eta(t))\right)} h(t) = \frac{\exp(-t)}{h(t)} h(t) = e^{-t}$$

for $t \ge t_1$ and hence

$$\int_0^\infty f'(\tau)h(\tau)d\tau = \int_0^{t_1} f'(\tau)h(\tau)d\tau + \int_{t_1}^\infty f'(\tau)h(\tau)d\tau \le \int_0^{t_1} f'(\tau)h(\tau)d\tau + e^{-t_1}.$$

Again the integral is bounded, since $f'(t)h(t) \le M$ on $0 \le t \le t_1$ for some $M>0$ and therefore

$$\int_0^\infty f'(\tau)h(\tau)d\tau \le M_1 + e^{-t_1} =: k_2.$$

□

Lemma C.6 Exponentially Growing Systems [Kha02]

Assume that $\mathbf{f}(\mathbf{x},t)$ in (C.1) is Lipschitz on $\mathcal{D}$
uniformly in t, i.e. $\|\mathbf{f}(\mathbf{x},t)\|_2 \le L \cdot \|\mathbf{x}\|_2$ for all $\mathbf{x} \in \mathcal{D}$ and $t \in [0,\infty)$,
then the solution $\mathbf{x}(t)$ that belongs to $\mathcal{D}$ for all $t \ge t_0$ is bounded by

$$\|\mathbf{x}_0\|_2 e^{-L(t-t_0)} \le \|\mathbf{x}(t)\|_2 \le \|\mathbf{x}_0\|_2 e^{L(t-t_0)}$$

Remark

In case of a linear time invariant matrix differential equations of the form

$$\dot{\mathbf{X}}(t) = \mathbf{A}(t) \cdot \mathbf{X}(t) \quad \mathbf{x}(t_0) = \mathbf{x}_0$$

where $\mathbf{X} \in \mathbb{R}^{n \times m}$ and $\|\mathbf{A}(t)\|_2 \le L$ for all $t \ge t_0$ one can conclude exponential bounds. Let $\mathbf{x}_i(t)$ denote the columns of $\mathbf{X}$ (t), $i=1,\ldots,m$, then the matrix differential equation collapses to m vector differential equations of the form

$$\dot{\mathbf{x}}_i(t) = \mathbf{f}(\mathbf{x}_i, t)$$

where $\mathbf{f}(\mathbf{x}_i, t) = \mathbf{A}(t)\mathbf{x}_i$. Using the matrix induced norm, we get

$$\|\mathbf{f}(\mathbf{x}_0, t)\| \le \|\mathbf{A}(t)\|_2 \|\mathbf{x}_i\|_2 \le L\|\mathbf{x}_i\|_2$$

and Lemma C.6 delivers exponential bounds

$$\|\mathbf{x}_i(t_0)\|_2 e^{-L(t-t_0)} \le \|\mathbf{x}_i(t)\|_2 \le \|\mathbf{x}_i(t_0)\|_2 e^{L(t-t_0)}$$

The Frobenius norm of $\mathbf{X}$ can be computed as

$$\|\mathbf{X}(t)\|_F = \sqrt{\sum_{i=1}^{n} \|\mathbf{x}_i(t)\|_2^2}$$

Applying this result to the exponential bounds yields and inequality in terms of the Frobenius norm

$$\|\mathbf{X}(t_0)\|_F \cdot e^{-L(t-t_0)} \le \|\mathbf{X}(t)\|_F \le \|\mathbf{X}(t_0)\|_F \cdot e^{L(t-t_0)} \tag{C.46}$$

From the matrix norm inequalities (see e.g. [Lüt96]),

$$\|\mathbf{X}\|_2 \le \|\mathbf{X}\|_F \le \sqrt{\min(n,m)} \cdot \|\mathbf{X}\|_2$$

we also obtain the following inequalities in terms of matrix induced 2-norm

$$\frac{1}{\sqrt{\min(n,m)}} \|\mathbf{X}(t_0)\|_2 \cdot e^{-L(t-t_0)} \le \|\mathbf{X}(t)\|_2 \le \sqrt{\min(n,m)} \|\mathbf{X}(t_0)\|_2 \cdot e^{L(t-t_0)} \tag{C.47}$$

Proof

With

$$u(t) = \mathbf{x}^T(t)\mathbf{x}(t) = \|\mathbf{x}(t)\|_2^2$$

we get

$$\begin{aligned} |\dot{u}(t)| &= \left|\dot{\mathbf{x}}^T(t)\mathbf{x}(t) + \mathbf{x}^T(t)\dot{\mathbf{x}}(t)\right| = 2\left|\mathbf{x}^T(t)\dot{\mathbf{x}}(t)\right| \\ &= 2\left|\mathbf{x}^T(t)\mathbf{f}(\mathbf{x},t)\right| \end{aligned}$$

Using the Lipschitz condition

$$\|\mathbf{f}(\mathbf{x},t)\|_2 \le L\|\mathbf{x}\|_2$$

we arrive at a separable differential inequality

$$\left|\frac{d}{dt}(u(t))\right| \le 2L\|\mathbf{x}(t)\|^2 = 2Lu(t)$$

which reads explicitly

$$-2Lu(t) \le \frac{d(u(t))}{dt} \le 2Lu(t).$$

Separation of variables and integration yields

$$-2L(t-t_0) \le \ln\left(\frac{u(t)}{u_0}\right) \le 2L(t-t_0)$$

Finally back substitution $u(t) = \|\mathbf{x}(t)\|_2^2$ and solving for $\mathbf{x}(t)$ yields

$$\|\mathbf{x}_0\| \cdot e^{-L(t-t_0)} \le \|\mathbf{x}(t)\| \le \|\mathbf{x}_0\| e^{L(t-t_0)}$$

and the proof is complete.

□

Lemma C.7 Differential Equation of Sensitivity Functions

Assume that $f(\mathbf{x},t)$ in (C.1)

is continuously differentiable in its arguments on $\mathcal{D}\times[0,\infty)$.

Let $\mathbf{x}(t,\mathbf{x}_0,t_0)$ be the solution of (C.1) with initial value $\mathbf{x}_0$ and initial time t_0, observed at time t which resides in some compact set W for all $t \geq t_0$ and $W \subset \mathcal{D}$.

Then $\mathbf{X}_{t_0}(t,\mathbf{x}_0,t_0)+\mathbf{x}_{x_0}(t,\mathbf{x}_0,t_0)\cdot\mathbf{f}(\mathbf{x}_0,t_0)=\mathbf{0}$ for all $t \geq t_0$

Proof

(C.1) satisfies the following integral equation

$$\mathbf{x}(t,\mathbf{x}_0,t_0)=\mathbf{x}_0+\int_{t_0}^{t}\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)d\tau \tag{C.48}$$

Then deriving $\mathbf{x}(t,\mathbf{x}_0,t_0)$ w.r.t. $\mathbf{x}_0$ yields

$$\mathbf{X}_{x_0}(t,\mathbf{x}_0,t_0)=\mathbf{I}+\int_{t_0}^{t}\frac{\partial\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0))}{\partial\mathbf{x}}\cdot\mathbf{X}_{x_0}(\tau,\mathbf{x}_0,t_0)d\tau \tag{C.49}$$

where $\mathbf{I}$ denotes the identity matrix of appropriate dimension. If we derive (C.49) w.r.t. t, yields a matrix differential equation

$$\frac{\partial\mathbf{X}_{x_0}(t,\mathbf{x}_0,t_0)}{\partial t}=\frac{\partial\mathbf{f}(\mathbf{x}(t,\mathbf{x}_0,t_0))}{\partial\mathbf{x}}\cdot\mathbf{X}_{x_0}(t,\mathbf{x}_0,t_0) \tag{C.50}$$

with initial condition $\mathbf{X}_{x_0}(t_0,\mathbf{x}_0,t_0)=\mathbf{I}$. Since $\mathbf{x}(t,\mathbf{x}_0,t_0)$ remains in the compact set W for all times and the Jacobian of $\mathbf{f}(\mathbf{x},t)$ is continuous on W, we have

$$\left\|\frac{\partial\mathbf{f}(\mathbf{x},t)}{\partial\mathbf{x}}\right\|\leq L$$

for some $L>0$. Further let $\mathbf{h}_i(t,\mathbf{x}_0,t_0)$, $i=1,\ldots,n$ be the columns of $\mathbf{X}_{\mathbf{x}0}(t,\mathbf{x}_0,t_0)$, then we obtain n vector differential equations

$$\frac{\partial\mathbf{h}_i(t,\mathbf{x}_0,t_0)}{\partial t}=\frac{\partial\mathbf{f}(\mathbf{x}(t,\mathbf{x}_0,t_0),t)}{\partial\mathbf{x}}\cdot\mathbf{h}_i(t,\mathbf{x}_0,t_0)$$

The function on the right hand side is globally Lipschitz with constant L and therefore existence and uniqueness of $\mathbf{X}_{\mathbf{x}0}(t,\mathbf{x}_0,t_0)$ is assured for all $t\geq t_0$ (see Theorem 3.2 in [Kha02]). The derivative of (C.48) w.r.t. to t_0 is not so straight forward, since t_0 appears in the integrand as well as in the integration bound. In order to make the derivative plausible, we start with the definition of the derivative by its differential quotient

$$\mathbf{x}_{t_0}(t,\mathbf{x}_0,t_0)=\lim_{h\to 0}\left\{\frac{\mathbf{x}_0+\int_{t_0+h}^{t}\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0+h),\tau)d\tau-\mathbf{x}_0-\int_{t_0}^{t}\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)d\tau}{h}\right\}$$

$$=\lim_{h\to 0}\left\{\frac{\int_{t_0+h}^{t}\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0+h),\tau)-\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)d\tau-\int_{t_0}^{t_0+h}\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)d\tau}{h}\right\}$$

Since $\mathbf{f}(\mathbf{x},t)$ is continuously differentiable w.r.t. $\mathbf{x}$, we have

$$\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0+h),\tau)-\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)=\frac{\partial\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)}{\partial\mathbf{x}}\cdot\mathbf{x}_{t_0}(\tau,\mathbf{x}_0,t_0)\cdot h+O(h^2)$$

and

$$\frac{\partial\mathbf{x}(t,\mathbf{x}_0,t_0)}{\partial t_0}=\lim_{h\to 0}\left\{\frac{\int_{t_0+h}^{t}\frac{\partial\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)}{\partial\mathbf{x}}\frac{\partial\mathbf{x}(\tau,\mathbf{x}_0,t_0)}{\partial t_0}d\tau\cdot h+O(h^2)-\int_{t_0}^{t_0+h}\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)d\tau}{h}\right\}$$

$$=\lim_{h\to 0}\left\{\int_{t_0+h}^{t}\frac{\partial\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)}{\partial\mathbf{x}}\frac{\partial\mathbf{x}(\tau,\mathbf{x}_0,t_0)}{\partial t_0}d\tau+O(h)\right\}-\lim_{h\to 0}\left\{\frac{\int_{t_0}^{t_0+h}\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)d\tau}{h}\right\}$$

while the first limit converges to

$$\int_{t_0}^{t}\frac{\partial\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)}{\partial\mathbf{x}}\frac{\partial\mathbf{x}(\tau,\mathbf{x}_0,t_0)}{\partial t_0}d\tau$$

for the second term denote the antiderivative of $\mathbf{f}(t,\mathbf{x}_0,t_0)$ w.r.t. τ as $\mathbf{F}(t,\mathbf{x}_0,t_0)$, then

$$\lim_{h\to 0}\left\{h^{-1}\left[\int_{t_0}^{t_0+h}\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)d\tau\right]\right\}=\lim_{h\to 0}\left\{h^{-1}\left[\mathbf{F}(t_0+h,\mathbf{x}_0,t_0)-\mathbf{F}(t_0,\mathbf{x}_0,t_0)\right]\right\}=\mathbf{f}(\mathbf{x}(t_0,\mathbf{x}_0,t_0),t_0)=\mathbf{f}(\mathbf{x}_0,t_0)$$

and hence

$$\mathbf{x}_{t_0}(t,\mathbf{x}_0,t_0)=\int_{t_0}^{t}\frac{\partial\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)}{\partial\mathbf{x}}\mathbf{x}_{t_0}(\tau,\mathbf{x}_0,t_0)d\tau-\mathbf{f}(\mathbf{x}_0,t_0).\qquad\text{(C.51)}$$

With (C.50) and (C.51), we get

$$\mathbf{x}_{t_0}(t,\mathbf{x}_0,t_0)+\mathbf{x}_{x_0}(t,\mathbf{x}_0,t_0)\cdot\mathbf{f}(\mathbf{x}_0,t_0)=\int_{t_0}^{t}\frac{\partial\mathbf{f}(\mathbf{x}(\tau,\mathbf{x}_0,t_0),\tau)}{\partial\mathbf{x}}\left\{\mathbf{x}_{t_0}(\tau,\mathbf{x}_0,t_0)+\mathbf{x}_{x_0}(\tau,\mathbf{x}_0,t_0)\mathbf{f}(\mathbf{x}_0,t_0)\right\}d\tau\qquad\text{(C.52)}$$

The derivation of (C.52) w.r.t. t, $\mathbf{y}(t,\mathbf{x}_0,t_0)=\mathbf{x}_{t_0}(t,\mathbf{x}_0,t_0)+\mathbf{x}_{x_0}(t,\mathbf{x}_0,t_0)\mathbf{f}(\mathbf{x}_0,t_0)$ satisfies the following differential equation.

$$\frac{\partial\mathbf{y}(t,\mathbf{x}_0,t_0)}{\partial t}=\frac{\partial\mathbf{f}(\mathbf{x}(t,\mathbf{x}_0,t_0),t)}{\partial\mathbf{x}}\mathbf{y}(t,\mathbf{x}_0,t_0)\qquad\text{(C.53)}$$

where the initial condition is $\mathbf{y}(t_0,\mathbf{x}_0,t_0)=\mathbf{0}$ in light of (C.52). Obviously $\mathbf{y}=\mathbf{0}$ is an equilibrium and hence

$$\mathbf{y}(t,\mathbf{x}_0,t_0)=\mathbf{0}$$

for all $t\geq t_0$ which completes the proof.

□

Appendix D

Robustness Modifications

D.1 Projection Operator

In adaptive systems, unmatched uncertainties cause problems in stability analysis, when the standard law is used for the parameter update. The specific problem is that parameters could drift and hence grow unbounded, if no measure for robustification is taken.

A possible remedy is a simple limitation of the parameters, by stopping integration when the parameter estimate hits a predefined boundary. Drawback of this method is that the time derivative of the adaptive parameters is discontinuous at time instances, the bound is hit. A more sophisticated concept is given by the *projection operator* ([Hov10], [Ioa06]). It confines the parameter matrix to a convex set by reducing the very component of the time derivative, pointing normal to the bound when the parameter estimate is approaching the boundary. On the boundary itself, the normal component is completely faded out and only the tangential component is used, forcing the parameter to stride along the boundary. The following explanations are inspired the ones given in [Hov10], however the definitions and propositions, which are given for vectors are extended to the matrix case. Additionally to [Hov10], the framework allows the incorporation of the weighted Frobenius norm, which allows a customization of the convex set such that it fits better to the set where the ideal weights are located, leading to potentially less conservative bounds in the stability analysis of the adaptive system.

Definition D.1 Convex Set on $\mathbb{R}^{nxm}$

A set $\mathcal{U} \subset \mathbb{R}^{n\times m}$ is said to be convex if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{U}$,

also $\mathbf{z}(\lambda) = \lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in \mathcal{U}$ for all $\lambda \in [0,1]$.

Remark

A set is convex, if the straight line, connecting two arbitrary points in $\mathcal{U}$, completely lies within $\mathcal{U}$. In $\mathbb{R}^2$ a convex set has to be such that is does not have any dint (Figure D.1).

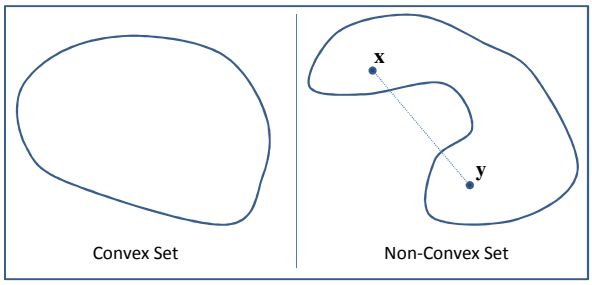


Figure D.1 Convex Set

Definition D.2 Convex Function

Let $\mathcal{U} \subset \mathbb{R}^{n\times m}$ be convex. Then a function $f : \mathcal{U} \to \mathbb{R}$ is said to be convex if, for any $\mathbf{x}, \mathbf{y} \in \mathcal{U}$:

$$f[\lambda\mathbf{x} + (1-\lambda)\mathbf{y}] \leq \lambda f[\mathbf{x}](1-\lambda) f[\mathbf{y}] \text{ for all } \lambda \in [0,1]$$

Remark

A function is convex, if a straight line, connecting two arbitrary points in the image space lies above the function itself. For a scalar function, i.e. $\mathcal{U} \subset \mathbb{R}$, Figure D.2 illustrates a convex function. Hence a convex function must not have any dint.

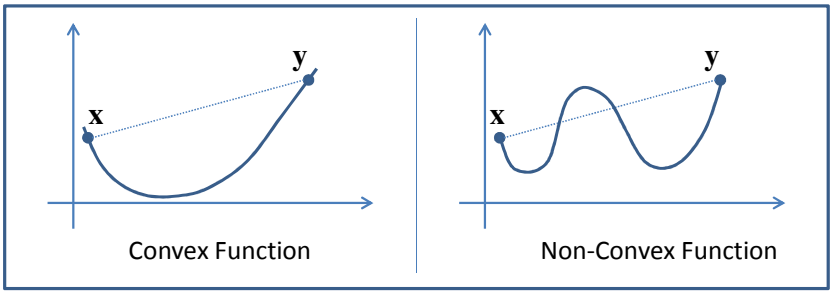


Figure D.2 Convex Function

The next lemma states an important property of convex functions, namely that level sets of convex functions define convex sets.

Lemma D.1 Convex Function Level Set

Let $f : \mathcal{U} \to \mathbb{R}$, $\mathcal{U} \subset \mathbb{R}^{n\times m}$ be a convex function.

Then any level set $\Omega_\delta = \{\mathbf{x} \in \mathcal{U} | f(\mathbf{x}) \leq \delta\}$, $\delta \in \mathbb{R}$, is convex

Now getting precise, the projection operator employs the following convex function

$$f_{\varepsilon,\theta_{\max},\mathbf{\Gamma}}(\mathbf{\Theta}) = \max\left[0, \frac{(\varepsilon+1)\|\mathbf{\Theta}\|_\Gamma^2 - \theta_{\max}^2}{\varepsilon\theta_{\max}^2}\right] \tag{D.1}$$

where $\varepsilon \in (0,1)$ and $\theta_{\max} > 0$, $\mathbf{\Theta} \in \mathcal{U} \subset \mathbb{R}^{n\times m}$ and $\|\cdot\|_\Gamma$ denotes the weighted Frobenius norm with a positive definite weighting matrix $\mathbf{\Gamma} \in \mathbb{R}^{n\times n}$ (refer to section 1.3 for more information). Further the gradient of (D.1) is defined as:

$$\nabla f_{\varepsilon,\theta_{\max},\Gamma}(\boldsymbol{\Theta}) = \begin{cases} \mathbf{0} & \text{for} \left\|\boldsymbol{\Theta}\right\|_\Gamma < \dfrac{\theta_{\max}}{1+\varepsilon} \\ \dfrac{2(\varepsilon+1)\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}}{\varepsilon\theta_{\max}^2} & \text{for} \left\|\boldsymbol{\Theta}\right\|_\Gamma \geq \dfrac{\theta_{\max}}{1+\varepsilon} \end{cases} \tag{D.2}$$

In the following, the indices $\varepsilon, \theta_{\max}, \boldsymbol{\Gamma}$ will be dropped for readability, where appropriate.

For convexity of (D.1), it has to be shown that

$$g(\lambda, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) := f\left[\lambda\boldsymbol{\Theta}_1 + (1-\lambda)\boldsymbol{\Theta}_2\right] - \lambda f\left[\boldsymbol{\Theta}_1\right] - (1-\lambda) f(\boldsymbol{\Theta}_2) \leq 0$$

for some $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2 \in \mathcal{U}$, if $0 \leq \lambda \leq 1$.

CASE 1: $\left\|\boldsymbol{\Theta}_1\right\|_\Gamma^2 \geq \dfrac{\theta_{\max}^2}{1+\varepsilon}, \left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2 \geq \dfrac{\theta_{\max}^2}{1+\varepsilon}$

$$g(\lambda, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) = \frac{(\varepsilon+1)\left\|\lambda\boldsymbol{\Theta}_1 + (1-\lambda)\boldsymbol{\Theta}_2\right\|_\Gamma^2 - \theta_{\max}^2}{\varepsilon\theta_{\max}^2} - \frac{(\varepsilon+1)\left[\lambda\left\|\boldsymbol{\Theta}_1\right\|_\Gamma^2 + (1-\lambda)\left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2\right] - \theta_{\max}^2}{\varepsilon\theta_{\max}^2}$$

Since $\varepsilon, \theta_{\max} > 0$, we can equivalently investigate

$$\frac{\varepsilon\theta_{\max}^2}{\varepsilon+1} g(\lambda, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) = \left\|\lambda\boldsymbol{\Theta}_1 + (1-\lambda)\boldsymbol{\Theta}_2\right\|_\Gamma^2 - \lambda\left\|\boldsymbol{\Theta}_1\right\|_\Gamma^2 - (1-\lambda)\left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2 \leq 0$$

Then, if we employ the trace notation of weighted Frobenius norm, we obtain:

$$\begin{aligned} \frac{\varepsilon\theta_{\max}^2}{\varepsilon+1} g(\lambda, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) &= \lambda^2 \operatorname{tr}\left[\boldsymbol{\Theta}_1^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}_1\right] + 2\lambda(1-\lambda)\operatorname{tr}\left[\boldsymbol{\Theta}_1^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}_2\right] + (1-\lambda)^2 \operatorname{tr}\left[\boldsymbol{\Theta}_2^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}_2\right] - \lambda\left\|\boldsymbol{\Theta}_1\right\|_\Gamma^2 - (1-\lambda)\left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2 \\ &= \lambda^2\left\|\boldsymbol{\Theta}_1\right\|_\Gamma^2 + 2\lambda(1-\lambda)\operatorname{tr}\left[\boldsymbol{\Theta}_1^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}_2\right] + (1-\lambda)^2\left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2 - \lambda\left\|\boldsymbol{\Theta}_1\right\|_\Gamma^2 - (1-\lambda)\left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2 \\ &= \lambda(\lambda-1)\left\|\boldsymbol{\Theta}_1\right\|_\Gamma^2 + 2\lambda(1-\lambda)\operatorname{tr}\left[\boldsymbol{\Theta}_1^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}_2\right] + \lambda(\lambda-1)\left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2 \end{aligned}$$

Applying Cauchy-Schwartz inequality to the mixed term, further yields

$$\begin{aligned} \frac{\varepsilon\theta_{\max}^2}{\varepsilon+1} g(\lambda, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) &\leq \lambda(\lambda-1)\left\|\boldsymbol{\Theta}_1\right\|_\Gamma^2 + 2\lambda(1-\lambda)\left\|\boldsymbol{\Theta}_1\right\|_\Gamma\left\|\boldsymbol{\Theta}_2\right\|_\Gamma + \lambda(\lambda-1)\left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2 \\ &= \lambda(\lambda-1)\left(\left\|\boldsymbol{\Theta}_1\right\|_\Gamma - \left\|\boldsymbol{\Theta}_2\right\|_\Gamma\right)^2 \end{aligned}$$

and since $0 \leq \lambda \leq 1$, we obtain $g(\lambda, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) \leq 0$.

CASE 2: $\left\|\boldsymbol{\Theta}_1\right\|_F^2 < \dfrac{\theta_{\max}^2}{1+\varepsilon}, \left\|\boldsymbol{\Theta}_2\right\|_F^2 \geq \dfrac{\theta_{\max}^2}{1+\varepsilon}$

(Without loss of generality, the analog case $\left\|\boldsymbol{\Theta}_1\right\|_\Gamma \geq \theta_{\max}^2/(\varepsilon+1)^2$), $\left\|\boldsymbol{\Theta}_2\right\|_\Gamma < \theta_{\max}^2/(\varepsilon+1)$ is done along similar lines.)

$$g(\lambda, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) = \frac{(\varepsilon+1)\left\|\lambda\boldsymbol{\Theta}_1 + (1-\lambda)\boldsymbol{\Theta}_2\right\|_\Gamma^2 - \theta_{\max}^2}{\varepsilon\theta_{\max}^2} - \frac{(\varepsilon+1)(1-\lambda)\left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2 - (1-\lambda)\theta_{\max}^2}{\varepsilon\theta_{\max}^2}$$

With $(1+\varepsilon)\left\|\boldsymbol{\Theta}_1\right\|_F^2 < \theta_{\max}^2$, following inequality holds.

$$g(\lambda, \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) \leq \frac{(\varepsilon+1)\left\|\lambda\boldsymbol{\Theta}_1 + (1-\lambda)\boldsymbol{\Theta}_2\right\|_\Gamma^2 - \theta_{\max}^2}{\varepsilon\theta_{\max}^2} - \frac{(\varepsilon+1)(1-\lambda)\left\|\boldsymbol{\Theta}_2\right\|_\Gamma^2 + (\varepsilon+1)\lambda\left\|\boldsymbol{\Theta}_1\right\|_\Gamma^2 - \theta_{\max}^2}{\varepsilon\theta_{\max}^2}$$

Since $\varepsilon, \theta_{\max} > 0$, we can equivalently investigate

$$\frac{\varepsilon\theta_{\max}^2}{\varepsilon+1} g(\lambda,\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2) \le \|\lambda\boldsymbol{\Theta}_1+(1-\lambda)\boldsymbol{\Theta}_2\|_\Gamma^2 - (1-\lambda)\|\boldsymbol{\Theta}_2\|_\Gamma^2 - \lambda\|\boldsymbol{\Theta}_1\|_\Gamma^2$$

The, using trace notation and applying Cauchy-Schwartz inequality to the mixed term yields

$$\begin{aligned}\frac{\varepsilon\theta_{\max}^2}{\varepsilon+1} g(\lambda,\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2) &\le \lambda^2\|\boldsymbol{\Theta}_1\|_\Gamma^2 + 2\lambda(1-\lambda)\|\boldsymbol{\Theta}_1\|_\Gamma\|\boldsymbol{\Theta}_2\|_\Gamma + (1-\lambda)^2\|\boldsymbol{\Theta}_2\|_\Gamma^2 - (1-\lambda)\|\boldsymbol{\Theta}_2\|_\Gamma^2 - \lambda\|\boldsymbol{\Theta}_1\|_\Gamma^2 \\ &= \lambda(\lambda-1)\|\boldsymbol{\Theta}_1\|_\Gamma^2 + 2\lambda(1-\lambda)\|\boldsymbol{\Theta}_1\|_\Gamma\|\boldsymbol{\Theta}_2\|_\Gamma + \lambda(\lambda-1)\|\boldsymbol{\Theta}_2\|_\Gamma^2 \\ &= \lambda(\lambda-1)\left(\|\boldsymbol{\Theta}_1\|_\Gamma - \|\boldsymbol{\Theta}_2\|_\Gamma\right)\end{aligned}$$

and since $0 \le \lambda \le 1$, one can conclude, that $g(\lambda,\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2) \le 0$

CASE 3: $\|\boldsymbol{\Theta}_1\|_F^2 < \frac{\theta_{\max}^2}{1+\varepsilon}, \|\boldsymbol{\Theta}_2\|_F^2 < \frac{\theta_{\max}^2}{1+\varepsilon}$

This case is trivial, since then $f(\boldsymbol{\Theta}_1) = f(\boldsymbol{\Theta}_2) = f(\lambda\boldsymbol{\Theta}_1 + (1-\lambda)\boldsymbol{\Theta}_2) = 0$.

The next lemma is needed in the Lyapunov stability proof, to obtain bounds for the projection operator

Lemma D.2 Bound on Convex Function

Let the function $f_{\varepsilon,\theta_{\max},\Gamma}(\boldsymbol{\Theta})$ with its gradient $\nabla f_{\varepsilon,\theta_{\max},\Gamma}(\boldsymbol{\Theta})$, defined in (D.1) and (D.2),

a constant $\delta > 0$,

a convex level set $\Omega_\delta = \left\{\boldsymbol{\Theta} \in \mathbb{R}^{n\times m} \middle| f_{\varepsilon,\theta_{\max},\Gamma}(\boldsymbol{\Theta}) \le \delta\right\}$

a $\boldsymbol{\Theta}^*$ such that $f_{\varepsilon,\theta_{\max},\Gamma}(\boldsymbol{\Theta}^*) \le \delta$ ($\boldsymbol{\Theta}^*$ in Ω_δ)

a $\boldsymbol{\Theta}$ such that $f_{\varepsilon,\theta_{\max},\Gamma}(\boldsymbol{\Theta}^*) = \delta$ ($\boldsymbol{\Theta}$ on boundary of Ω_δ)

Then $\mathrm{tr}\left[(\boldsymbol{\Theta}^* - \boldsymbol{\Theta})^T \nabla f_{\varepsilon,\theta_{\max},\Gamma}(\boldsymbol{\Theta})\right] \le 0$

Proof

$$\mathrm{tr}\left[(\boldsymbol{\Theta}^* - \boldsymbol{\Theta})^T \nabla f(\boldsymbol{\Theta})\right] = \frac{2(\varepsilon+1)}{\varepsilon\theta_{\max}^2}\mathrm{tr}\left[\boldsymbol{\Theta}^{*T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta} - \boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}\right] = \frac{2(\varepsilon+1)\mathrm{tr}\left[\boldsymbol{\Theta}^{*T}\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}\right]}{\varepsilon\theta_{\max}^2} - \frac{2(\varepsilon+1)\mathrm{tr}\left[\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}\right]}{\varepsilon\theta_{\max}^2}$$

Then, by applying Cauchy-Schwartz inequality:

$$\mathrm{tr}\left[(\boldsymbol{\Theta}^* - \boldsymbol{\Theta})^T \nabla f(\boldsymbol{\Theta})\right] \le \frac{2(\varepsilon+1)\|\boldsymbol{\Theta}^*\|_\Gamma\|\boldsymbol{\Theta}\|_\Gamma}{\varepsilon\theta_{\max}^2} - \frac{2(\varepsilon+1)\|\boldsymbol{\Theta}\|_\Gamma\|\boldsymbol{\Theta}\|_\Gamma}{\varepsilon\theta_{\max}^2}.$$

Then, from the definition of $\boldsymbol{\Theta}$ and $\boldsymbol{\Theta}^*$, the following inequality can be concluded.

$$f(\boldsymbol{\Theta}) = \frac{(\varepsilon+1)\|\boldsymbol{\Theta}\|_{\Gamma}^2 - \theta_{\max}^2}{\varepsilon\theta_{\max}^2} = \delta \rightarrow \qquad \|\boldsymbol{\Theta}\|_{\Gamma}^2 = \frac{\delta\varepsilon+1}{\varepsilon+1}\theta_{\max}^2$$

$$f(\boldsymbol{\Theta}^*) = \frac{(\varepsilon+1)\|\boldsymbol{\Theta}^*\|_{\Gamma}^2 - \theta_{\max}^2}{\varepsilon\theta_{\max}^2} \leq \delta \qquad \rightarrow \qquad \|\boldsymbol{\Theta}^*\|_{\Gamma}^2 \leq \frac{\delta\varepsilon+1}{\varepsilon+1}\theta_{\max}^2$$

$$\|\boldsymbol{\Theta}^*\|_{\Gamma} < \|\boldsymbol{\Theta}\|_{\Gamma}$$

Therefore

$$\operatorname{tr}\left[(\boldsymbol{\Theta}^* - \boldsymbol{\Theta})^T \nabla f(\boldsymbol{\Theta})\right] \leq \frac{2(\varepsilon+1)\|\boldsymbol{\Theta}\|_{\Gamma}\|\boldsymbol{\Theta}\|_{\Gamma}}{\varepsilon\theta_{\max}^2} - \frac{2(\varepsilon+1)\|\boldsymbol{\Theta}\|_{\Gamma}\|\boldsymbol{\Theta}\|_{\Gamma}}{\varepsilon\theta_{\max}^2} = 0$$

□

Using the results so far, the projection operator can now be defined.

Definition D.3 Projection Operator

Let some $\boldsymbol{\Theta}, \mathbf{Y} \in \mathbb{R}^{n\times m}$ and $f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\boldsymbol{\Theta})$, $\nabla f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\boldsymbol{\Theta})$ as defined in (D.1), (D.2).

The projection operator is defined as

$$\operatorname{Proj}(\boldsymbol{\Theta},\mathbf{Y})_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}} = \begin{cases} \mathbf{Y} & \text{if } \operatorname{tr}\left[\left[\nabla f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\boldsymbol{\Theta})\right]^T \mathbf{Y}\right] < 0 \\ \mathbf{Y} - \dfrac{\boldsymbol{\Gamma}\nabla f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\boldsymbol{\Theta})}{\left\|\boldsymbol{\Gamma}\nabla f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\boldsymbol{\Theta})\right\|_{\Gamma}} \operatorname{tr}\left[\dfrac{\left[\nabla f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\boldsymbol{\Theta})\right]^T \mathbf{Y}}{\left\|\boldsymbol{\Gamma}\nabla f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\boldsymbol{\Theta})\right\|_{\Gamma}}\right] \cdot f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\boldsymbol{\Theta}) & \text{if } \operatorname{tr}\left[\left[\nabla f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\boldsymbol{\Theta})\right]^T \mathbf{Y}\right] \geq 0 \end{cases}$$

Remark 1

The projection operator projects $\mathbf{Y}$ dependent on $\boldsymbol{\Theta}$. If $f(\boldsymbol{\Theta}) = 0$, i.e.

$$\|\boldsymbol{\Theta}\|_{\Gamma}^2 < \frac{\theta_{\max}^2}{\varepsilon+1}$$

and $\operatorname{tr}\left[\left[\nabla f(\boldsymbol{\Theta})\right]^T \mathbf{Y}\right] < 0$ the projection operator leaves $\mathbf{Y}$ unchanged. Otherwise the projection operator is activated. In order to explain functionality, we at first assume that the projection operator is applied to column vectors $\boldsymbol{\theta} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^n$ and the norm weighting equals identity matrix.

$$\operatorname{Proj}(\boldsymbol{\theta},\mathbf{y}) = \mathbf{y} - \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|_2}\frac{\boldsymbol{\theta}^T\mathbf{y}}{\|\boldsymbol{\theta}\|_2} f(\boldsymbol{\theta})$$

$\boldsymbol{\theta}^T\mathbf{y}$ is the standard scalar product. It is well-known that the second term in the expression above defines the projection of $\mathbf{y}$ onto $\boldsymbol{\theta}$, scaled by $f(\boldsymbol{\theta})$. If $f(\boldsymbol{\theta}) = 1$, with corresponds to $\|\boldsymbol{\theta}\|_2 = \theta_{\max}$, the projection is completely subtracted from $\mathbf{y}$ and only the perpendicular part is left and if $\boldsymbol{\theta}$ moves along a trajectory, i.e. $\dot{\boldsymbol{\theta}} = \operatorname{Proj}(\boldsymbol{\theta},\mathbf{y})$, $\boldsymbol{\theta}$ remains at constant Euclidean length, as can be shown, by computing:

$$\frac{d}{dt}\|\boldsymbol{\theta}\|_2^2 = 2\boldsymbol{\theta}^T\dot{\boldsymbol{\theta}} = 2\boldsymbol{\theta}^T\,\mathrm{Proj}(\boldsymbol{\theta},\mathbf{y}) = 2\boldsymbol{\theta}^T\left[\mathbf{y} - \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|_2}\frac{\boldsymbol{\theta}^T\mathbf{y}}{\|\boldsymbol{\theta}\|_2}\right]$$

$$= 2\boldsymbol{\theta}^T\mathbf{y} - \frac{2\boldsymbol{\theta}^T\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|_2}\frac{\boldsymbol{\theta}^T\mathbf{y}}{\|\boldsymbol{\theta}\|_2} = 2\boldsymbol{\theta}^T\mathbf{y} - \frac{2\|\boldsymbol{\theta}\|_2^2}{\|\boldsymbol{\theta}\|_2}\frac{\boldsymbol{\theta}^T\mathbf{y}}{\|\boldsymbol{\theta}\|_2} = 0$$

For matrices $\mathbf{Y}$ and $\boldsymbol{\Theta}$, the projection operator reads as

$$\mathrm{Proj}(\boldsymbol{\Theta},\mathbf{Y}) = \mathbf{Y} - \frac{\boldsymbol{\Theta}}{\|\boldsymbol{\Theta}\|_\Gamma}\frac{\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y}\right)}{\|\boldsymbol{\Theta}\|_\Gamma} f(\boldsymbol{\Theta})$$

and analogously, if $f(\boldsymbol{\Theta}) = 1$, corresponding to $\|\boldsymbol{\Theta}\|_\Gamma = \theta_{\max}$, it can be shown that $\|\boldsymbol{\Theta}\|_\Gamma$ is cannot exceed $\theta_{\max}$, if it moves on a trajectory prescribed by $\mathrm{Proj}(\boldsymbol{\Theta},\mathbf{Y})$. To this end, assume $\dot{\boldsymbol{\Theta}} = \mathrm{Proj}(\boldsymbol{\Theta},\mathbf{Y})$, $\|\boldsymbol{\Theta}\|_\Gamma = \theta_{\max}$ and compute

$$\frac{d}{dt}\|\boldsymbol{\Theta}\|_\Gamma^2 = 2\,\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\dot{\boldsymbol{\Theta}}\right) = 2\,\mathrm{tr}\left[\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\,\mathrm{Proj}(\boldsymbol{\Theta},\mathbf{Y})\right].$$

Obviously, if $\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y}\right) < 0$, the projection operator is inactive and we have

$$\frac{d}{dt}\|\boldsymbol{\Theta}\|_\Gamma^2 = 2\,\mathrm{tr}\left[\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y}\right] < 0$$

However, if $\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y}\right) \geq 0$, we have

$$\frac{d}{dt}\|\boldsymbol{\Theta}\|_\Gamma^2 = 2\,\mathrm{tr}\left[\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\left(\mathbf{Y} - \frac{\boldsymbol{\Theta}}{\|\boldsymbol{\Theta}\|_\Gamma}\frac{\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y}\right)}{\|\boldsymbol{\Theta}\|_\Gamma}\right)\right]$$

$$= 2\,\mathrm{tr}\left[\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y} - \frac{\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}}{\|\boldsymbol{\Theta}\|_\Gamma}\frac{\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y}\right)}{\|\boldsymbol{\Theta}\|_\Gamma}\right] = 2\,\mathrm{tr}\left[\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y}\right] - 2\frac{\|\boldsymbol{\Theta}\|_\Gamma^2}{\|\boldsymbol{\Theta}\|_\Gamma}\frac{\mathrm{tr}\left(\boldsymbol{\Theta}^T\boldsymbol{\Gamma}^{-1}\mathbf{Y}\right)}{\|\boldsymbol{\Theta}\|_\Gamma} = 0$$

It reveals that the time derivative of $\|\boldsymbol{\Theta}\|_\Gamma^2$ is not greater than zero and hence $\|\boldsymbol{\Theta}\|_\Gamma \leq \theta_{\max}$. Taking into account the argument so far, if $\boldsymbol{\Theta}$ is moving in the direction of $\mathrm{Proj}\,(\boldsymbol{\Theta},\mathbf{Y})$, it cannot leave the level set

$$\Omega_1 = \left\{\boldsymbol{\Theta} \in \mathbb{R}^{n\times m} \,\middle|\, f(\boldsymbol{\Theta}) \leq 1\right\} \tag{D.3}$$

which corresponds to $\|\boldsymbol{\Theta}\|_\Gamma \leq \theta_{\max}$. Figure D.3 illustrates the projection operator for $\mathbf{Y},\boldsymbol{\Theta} \in \mathbb{R}^2$ and identity norm weighting matrix. It also illustrates why the projection operator is inactive if $\mathrm{tr}\left[[\nabla f(\boldsymbol{\Theta})]^T\mathbf{Y}\right] < 0$. It corresponds to the case where $\mathbf{Y}$ points inbound the contour lines of $f(\boldsymbol{\Theta})$ and hence does not leave Ω_1 anyway.

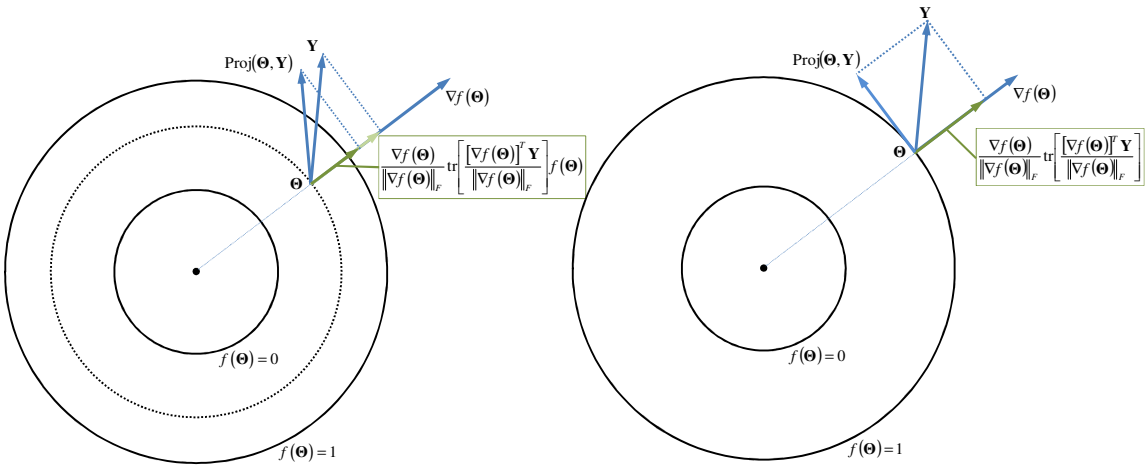


Figure D.3 Projection Operator

Remark 2

The Frobenius norm of the projection can be bounded from above by the Frobenius norm of its argument, if $\boldsymbol{\Theta} \in \Omega_1$.

$$\left\|\mathrm{Proj}(\boldsymbol{\Theta}, \mathbf{Y})\right\|_\Gamma \le \left\|\mathbf{Y}\right\|_\Gamma \tag{D.4}$$

I.e. the projection operator does not increase the Frobenius norm of $\mathbf{Y}$. In order to show this, we only have to investigate the case $f(\boldsymbol{\Theta}) \ge 0,\ \mathrm{tr}\left[\boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y}\right] \ge 0$.

$$\left\|\mathrm{Proj}(\boldsymbol{\Theta}, \mathbf{Y})\right\|_\Gamma^2 = \mathrm{tr}\left[\left(\mathbf{Y} - \frac{\boldsymbol{\Theta}}{\left\|\boldsymbol{\Theta}\right\|_\Gamma} \frac{\mathrm{tr}\left(\boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y}\right)}{\left\|\boldsymbol{\Theta}\right\|_\Gamma} f(\boldsymbol{\Theta})\right)^T \boldsymbol{\Gamma}^{-1} \left(\mathbf{Y} - \frac{\boldsymbol{\Theta}}{\left\|\boldsymbol{\Theta}\right\|_\Gamma} \frac{\mathrm{tr}\left(\boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y}\right)}{\left\|\boldsymbol{\Theta}\right\|_\Gamma} f(\boldsymbol{\Theta})\right)\right]$$

$$= \mathrm{tr}\left[\mathbf{Y}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y} - \mathbf{Y}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\Theta} \frac{\mathrm{tr}\left[\boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y}\right]}{\left\|\boldsymbol{\Theta}\right\|_\Gamma^2} f(\boldsymbol{\Theta}) - \boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y} \frac{\mathrm{tr}\left[\boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y}\right]}{\left\|\boldsymbol{\Theta}\right\|_\Gamma^2} f(\boldsymbol{\Theta}) + \boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \boldsymbol{\Theta} \frac{\mathrm{tr}\left[\boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y}\right]}{\left\|\boldsymbol{\Theta}\right\|_\Gamma^4} f^2(\boldsymbol{\Theta})\right]$$

Separating the summands, we get

$$\left\|\mathrm{Proj}(\boldsymbol{\Theta}, \mathbf{Y})\right\|_\Gamma^2 = \mathrm{tr}\left[\mathbf{Y}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y}\right] - \frac{\mathrm{tr}\left[\boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y}\right]^2}{\left\|\boldsymbol{\Theta}\right\|_\Gamma^2} \left(2 - f(\boldsymbol{\Theta})\right) f(\boldsymbol{\Theta})$$

$$= \left\|\mathbf{Y}\right\|_\Gamma^2 - \frac{\mathrm{tr}\left[\boldsymbol{\Theta}^T \boldsymbol{\Gamma}^{-1} \mathbf{Y}\right]^2}{\left\|\boldsymbol{\Theta}\right\|_\Gamma^2} \left(2 - f(\boldsymbol{\Theta})\right) f(\boldsymbol{\Theta})$$

where the second term is positive, since $\boldsymbol{\Theta}$ is restricted such that $0 \le f(\boldsymbol{\Theta}) \le 1$ and hence (D.4) is proved.

Remark 3

For some $\boldsymbol{\Theta}^*$ within the level set

$$\Omega_0 = \left\{\boldsymbol{\Theta}^* \in \mathbb{R}^{n \times m} \middle| f\left(\boldsymbol{\Theta}^*\right) \le 0\right\}$$

the following inequality holds and appears to be useful in the stability analysis or adaptive systems, using projection operator.

$$\mathrm{tr}\left[\left(\boldsymbol{\Theta}-\boldsymbol{\Theta}^{*}\right)^{T}\boldsymbol{\Gamma}^{-1}\left(\mathrm{Proj}\left[\boldsymbol{\Theta},\mathbf{Y}\right]-\mathbf{Y}\right)\right]\leq 0 \tag{D.5}$$

In order to show this, only the case

$$\mathrm{tr}\left[\left[\nabla f\left(\boldsymbol{\Theta}\right)\right]^{T}\mathbf{Y}\right]\geq 0 \text{ and } f\left(\boldsymbol{\Theta}\right)\geq 0 \tag{D.6}$$

has to be considered since, otherwise (D.5) evaluates to zero (refer to Definition D.3). Then

$$\begin{aligned}&\mathrm{tr}\left[\left(\boldsymbol{\Theta}-\boldsymbol{\Theta}^{*}\right)^{T}\boldsymbol{\Gamma}^{-1}\left(\mathrm{Proj}\left[\boldsymbol{\Theta},\mathbf{Y}\right]-\mathbf{Y}\right)\right]\\ &=\mathrm{tr}\left[\left(\boldsymbol{\Theta}-\boldsymbol{\Theta}^{*}\right)^{T}\boldsymbol{\Gamma}^{-1}\left(\mathbf{Y}-\frac{\boldsymbol{\Gamma}\nabla f\left(\boldsymbol{\Theta}\right)}{\left\|\boldsymbol{\Gamma}\nabla f\left(\boldsymbol{\Theta}\right)\right\|_{\Gamma}}\mathrm{tr}\left[\frac{\left[\nabla f\left(\boldsymbol{\Theta}\right)\right]^{T}\mathbf{Y}}{\left\|\boldsymbol{\Gamma}\nabla f\left(\boldsymbol{\Theta}\right)\right\|_{\Gamma}}\right]\cdot f\left(\boldsymbol{\Theta}\right)-\mathbf{Y}\right)\right]\\ &=\frac{\mathrm{tr}\left[\left[\nabla f\left(\boldsymbol{\Theta}\right)\right]^{T}\mathbf{Y}\right]}{\left\|\nabla f\left(\boldsymbol{\Theta}\right)\right\|_{\Gamma}^{2}}\cdot f\left(\boldsymbol{\Theta}\right)\mathrm{tr}\left[\left(\boldsymbol{\Theta}^{*}-\boldsymbol{\Theta}\right)^{T}\nabla f\left(\boldsymbol{\Theta}\right)\right]\end{aligned}$$

and, using assumption (D.6) and Lemma D.2, we obtain (D.5).

D.2 Switching σ-Modification

In the idealized situation of purely matched uncertainties, the MRAC approach with standard update law is sufficient to proof stability. However in presence of unmatched uncertainty, an indefinite terms remains in the time derivative of the Lyapunov function candidate. There are several modifications available for the update law, that render the system robust against limited unmatched uncertainty and at least boundedness of the system states can be concluded. σ and e modification are commonly known as robustness modification but they have a drawback that they deteriorate the performance of adaptation. Besides these, which are often the first choice, in order to make the adaptive system robust against unmatched uncertainties, there are also several others such as “dead zone”, bound for parameter estimates, or parameter projection. The problem of unmatched uncertainties in Lyapunov stability analysis is briefly recapitulated in the following.

In the standard MRAC approach with unmatched uncertainties, the error dynamics typically takes the following form (refer to equation (4.127))

$$\dot{\mathbf{e}}=\mathbf{A}\mathbf{e}+\mathbf{B}\left[\tilde{\boldsymbol{\Theta}}^{T}\boldsymbol{\varphi}\left(\mathbf{e},t\right)+\boldsymbol{\Delta}\left(\mathbf{e},t\right)\right] \tag{D.7}$$

where $\mathbf{A}$ is a stable matrix, $\tilde{\boldsymbol{\Theta}}=\hat{\boldsymbol{\Theta}}-\boldsymbol{\Theta}$ parameter-estimation-error and $\boldsymbol{\Delta}\left(\mathbf{e},t\right)$ is the unmatched uncertainty. For stability analysis, consider the following Lyapunov function candidate

$$V\left(\mathbf{e},\tilde{\boldsymbol{\Theta}}\right)=\mathbf{e}^{T}\mathbf{P}\mathbf{e}+\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}^{T}\boldsymbol{\Gamma}^{-1}\tilde{\boldsymbol{\Theta}}\right]$$

where $\mathbf{P}$ is the symmetric positive definite solution to the Lyapunov equation

$$\mathbf{A}^{T}\mathbf{P}+\mathbf{P}\mathbf{A}=-\mathbf{Q}$$

for some symmetric positive definite $\mathbf{Q}$ and $\boldsymbol{\Gamma}$ is a symmetric positive definite matrix. Then, the time derivative along the system trajectories is, after a few lines of computation

$$\dot{V} = -\mathbf{e}^T\mathbf{Q}\mathbf{e} + 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Delta}(\mathbf{e},t) + 2\,\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}^T\boldsymbol{\Gamma}^{-1}\dot{\tilde{\boldsymbol{\Theta}}} + \tilde{\boldsymbol{\Theta}}^T\boldsymbol{\varphi}(\mathbf{e},t)\mathbf{e}^T\mathbf{P}\mathbf{B}\right].$$

The standard update law

$$\dot{\hat{\boldsymbol{\Theta}}} = -\boldsymbol{\Gamma}\boldsymbol{\varphi}(\mathbf{e},t)\mathbf{e}^T\mathbf{P}\mathbf{B} \tag{D.8}$$

fails to render the Lyapunov time derivative negative definite.

$$\dot{V} = -\mathbf{e}^T\mathbf{Q}\mathbf{e} + 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Delta}(\mathbf{e},t)$$

The problem thereby is that $\dot{V}$ remains indefinite even if $\tilde{\boldsymbol{\Theta}}$ grows unbounded, if the error is small enough. In order to see this, an upper bound on $\dot{V}$ is

$$\dot{V} \le -\underline{\lambda}_Q\|\mathbf{e}\|_2^2 + 2D\|\mathbf{P}\mathbf{B}\|_2\|\mathbf{e}\|_2$$

where $\underline{\lambda}_Q$ denotes the minimum eigenvalues of $\mathbf{Q}$ and we assumed that the unmatched uncertainty is bounded by some $D>0$.

$$\|\boldsymbol{\Delta}(\mathbf{e},t)\|_2 \le D$$

Therefore $\dot{V}<0$ can only be concluded, if

$$\|\mathbf{e}\|_2 > \frac{D\|\mathbf{P}\mathbf{B}\|_2}{\underline{\lambda}_Q}$$

This in turn implies, that $\dot{V}$ is indefinite in the set

$$\Omega = \left\{(\mathbf{e},\tilde{\boldsymbol{\Theta}}) \,\middle|\, \|\mathbf{e}\|_2 \le \frac{2D\|\mathbf{P}\mathbf{B}\|_2}{\underline{\lambda}_Q}\right\}$$

Figure D.4 illustrates the situation. Hence the set, where $\dot{V}$ is indefinite cannot be encapsulated by some bounded set, but this is preliminary to the proof of ultimate boundedness (Theorem 3.1).

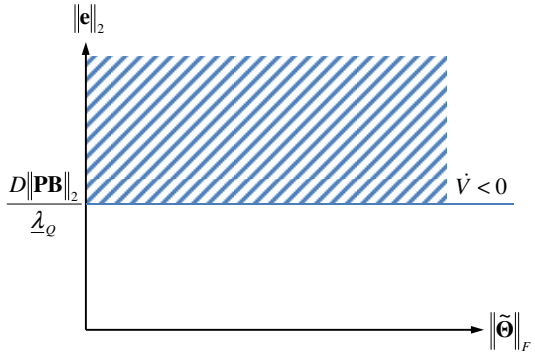


Figure D.4 Lyapunov Stability Situation without Robustness Modifications

Robustness modifications pursue the same intention, to render Ω, i.e. the set where $\dot{V}$ is indefinite, a bounded set. The σ–modification adds a feedback of the form

$$\dot{\hat{\Theta}} = -\Gamma\varphi(\mathbf{e},t)\mathbf{e}^T\mathbf{PB} - \sigma\hat{\Theta} \tag{D.9}$$

while the Lyapunov function derivative is changed to the following expression.

$$\dot{V} = -\mathbf{e}^T\mathbf{Q}\mathbf{e} + 2\mathbf{e}^T\mathbf{PB}\Delta(\mathbf{e},t) - 2\sigma\,\mathrm{tr}\left[\tilde{\Theta}^T\Gamma^{-1}\hat{\Theta}\right]$$

With $\hat{\Theta} = \tilde{\Theta} + \Theta$ and using Cauchy Schwartz inequality, an upper bound on $\dot{V}$ is given by

$$\dot{V} \le -\underline{\lambda}_Q\|\mathbf{e}\|_2^2 + 2D\|\mathbf{PB}\|_2\|\mathbf{e}\|_2 - 2\sigma\|\tilde{\Theta}\|_\Gamma^2 + 2\sigma\|\tilde{\Theta}\|_\Gamma\|\Theta\|_\Gamma \tag{D.10}$$

.

Now, the Lyapunov function derivative bound consists of two parts.

$$h_e\left(\|\mathbf{e}\|_2\right) = -\underline{\lambda}_Q\|\mathbf{e}\|_2^2 + 2D\|\mathbf{PB}\|_2\|\mathbf{e}\|_2 \text{ and } h_\theta\left(\|\tilde{\Theta}\|_\Gamma\right) = -2\sigma\|\tilde{\Theta}\|_\Gamma^2 + 2\sigma\|\tilde{\Theta}\|_\Gamma\|\Theta\|_\Gamma$$

Both are inverted parabolas and hence, are bounded from above.

$$h_e\left(\|\mathbf{e}\|_2\right) \le \frac{D\|\mathbf{PB}\|_2}{2\underline{\lambda}_Q} \text{ and } h_\theta\left(\|\tilde{\Theta}\|_F\right) \le \frac{\sigma\|\Theta\|_\Gamma}{2}$$

It can be concluded from these inequalities and (D.10) that $\dot{V} \le 0$, if either

$$\|\mathbf{e}\|_2 \ge \frac{D\|\mathbf{PB}\|_2}{\underline{\lambda}_Q} + \sqrt{\left(\frac{D\|\mathbf{PB}\|_2}{\underline{\lambda}_Q}\right)^2 + \frac{\sigma\|\Theta\|_F}{2\underline{\lambda}_Q}} =: e_0 \text{ or } \|\tilde{\Theta}\|_\Gamma \ge \frac{\|\Theta\|_\Gamma}{2} + \sqrt{\left(\frac{\|\Theta\|_\Gamma}{2}\right)^2 + \frac{D\|\mathbf{PB}\|_2}{4\sigma\underline{\lambda}_Q}} =: \theta_0 .$$

Hence the set of indefinite $\dot{V}$ is now

$$\Omega = \left\{(\mathbf{e},\tilde{\Theta}) \middle| \|\mathbf{e}\|_2 < e_0, \|\Theta\|_\Gamma < \theta_0\right\}$$

which is clearly bounded.

The σ-modification term unfortunately deteriorates the adaptation process. Suppose that there is no unmatched uncertainty in the error dynamics (D.7) and the estimated parameter converged to the true one, $\hat{\Theta} = \Theta$. Then the error dynamic reduces to a homogeneous stable system and thus $\mathbf{e} \to \mathbf{0}$ for large times. This in turn implies that, using update (D.8), the estimated parameters remain at their ideal value. If σ-modification is enabled, equation (D.9), the estimated parameter is pulled back to zero and hence away from the ideal parameter.

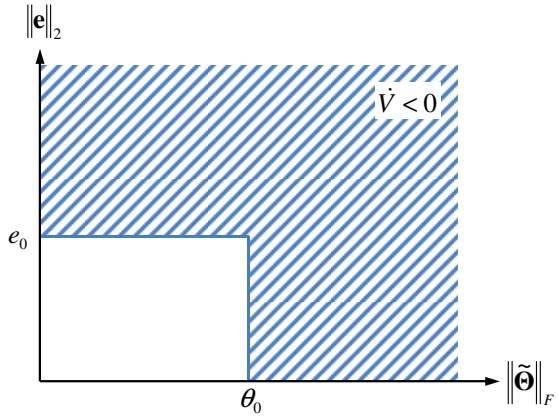


Figure D.5 Lyapunov Stability Situation with σ-Modification

In order to circumvent the drawback, imposed by σ-modification Ioannou introduced a modification called *switching* σ-*modification* ([Ioa06]). If there is any knowledge about the set, where the true parameters lie, is it reasonable to fade out the feedback term associated with σ-modification, if the parameter is moving within that set. Only, if it is imminent to run out of bound, the feedback term is activated and will pull the estimated parameter back to the admissible set. Hence parameter update with switching σ-modification is

$$\dot{\hat{\boldsymbol{\Theta}}} = -\boldsymbol{\Gamma}\boldsymbol{\varphi}(\mathbf{e},t)\mathbf{e}^T\mathbf{P}\mathbf{B} - \sigma\bar{f}_{\varepsilon,\theta_{\max}}(\hat{\boldsymbol{\Theta}})\boldsymbol{\Theta} \tag{D.11}$$

where

$$\bar{f}_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\hat{\boldsymbol{\Theta}}) = \min\left[1, f_{\varepsilon,\theta_{\max},\boldsymbol{\Gamma}}(\hat{\boldsymbol{\Theta}})\right] \tag{D.12}$$

and $f_{\varepsilon,\theta_{\max}}(\hat{\boldsymbol{\Theta}})$ is defined in (D.1). Then, the time derivative of the Lyapunov function candidate becomes

$$\dot{V} = -\mathbf{e}^T\mathbf{Q}\mathbf{e} + 2\mathbf{e}^T\mathbf{P}\mathbf{B}\boldsymbol{\Delta}(\mathbf{e},t) - 2\sigma\bar{f}(\hat{\boldsymbol{\Theta}})\mathrm{tr}\left[\tilde{\boldsymbol{\Theta}}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\Theta}\right]$$

where we have dropped the indices $\varepsilon,\theta_{\max}$. The σ-modification feedback is hence faded in if

$$\frac{\theta_{\max}^2}{\varepsilon+1} \le \left\|\hat{\boldsymbol{\Theta}}\right\|_{\boldsymbol{\Gamma}}^2 \le \theta_{\max}^2$$

and, of course $\theta_{\max}$ has to be chosen, such that $(\varepsilon+1)\|\boldsymbol{\Theta}\|_{\boldsymbol{\Gamma}}^2 \le \theta_{\max}^2$. Hence, if $\left\|\hat{\boldsymbol{\Theta}}\right\|_{\boldsymbol{\Gamma}} \ge \theta_{\max}$ the Lyapunov function time derivative can be bounded from above by (D.10) and hence $\dot{V} < 0$, if

$$\|\mathbf{e}\|_2 > e_0 \text{ and } \|\boldsymbol{\Theta}\|_F > \max[\theta_{\max}, \theta_0].$$

Thus switching σ-modification preserves the property that Ω is compact and, additionally, it does not deteriorate adaptation performance if the estimated parameter is within the admissible set.

Appendix E

Decoupling Control of Falb/Wolovich

Decoupling Control is a well-known control concept and thus described briefly here. More information can be found in [Wol74].

Consider a linear state space model.

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{A}\mathbf{x}+\mathbf{B}\mathbf{u}\\ \mathbf{y} &= \mathbf{C}\mathbf{x}\end{aligned} \tag{E.1}$$

with $\mathbf{A}\in\mathbb{R}^{n\times n}$, $\mathbf{B}\in\mathbb{R}^{n\times m}$, $\mathbf{C}\in\mathbb{R}^{m\times n}$.

The aim of the control algorithm is the decoupling of the MIMO system into *m* SISO systems. This approach is in fact a special case of feedback linearization, applied to linear systems. Therefore the relative degree is determined by differentiating each output until the input vector appears for the first time.

$$\begin{aligned} y_i &= \mathbf{c}_i^T\mathbf{x}\\ \dot{y}_i &= \mathbf{c}_i^T\mathbf{A}\mathbf{x}+\underbrace{\mathbf{c}_i^T\mathbf{B}\mathbf{u}}_{=0}\\ &\vdots\\ y_i^{(d_i)} &= \mathbf{c}_i^T\mathbf{A}^{d_i}\mathbf{x}+\underbrace{\mathbf{c}_i^T\mathbf{B}\mathbf{A}^{d_i-1}\mathbf{u}}_{\neq 0}\end{aligned} \tag{E.2}$$

where d_i is the relative degree of the *i*th output, $i=1,\ldots,m$.

Hence we get the following input-output relationship:

$$\left[y_i^{(d_i)}\right]=\mathbf{F}\mathbf{x}+\mathbf{G}\mathbf{u} \tag{E.3}$$

where

$$\mathbf{F}=\begin{bmatrix}\mathbf{c}_1^T\mathbf{A}^{d_1}\\ \vdots\\ \mathbf{c}_m^T\mathbf{A}^{d_m}\end{bmatrix} \qquad \mathbf{G}=\begin{bmatrix}\mathbf{c}_1^T\mathbf{A}^{d_1-1}\mathbf{B}\\ \vdots\\ \mathbf{c}_m^T\mathbf{A}^{d_m-1}\mathbf{B}\end{bmatrix} \tag{E.4}$$

The matrix $\mathbf{G}$ is called decoupling matrix and the system can be decoupled, if $\mathbf{G}$ is regular. If the decoupling control law is defined as

$$\mathbf{u}=\mathbf{G}^{-1}\left(\mathbf{v}-\mathbf{F}\mathbf{x}\right) \tag{E.5}$$

where $\boldsymbol{\nu}$ is the pseudo control vector of same dimension as $\mathbf{u}$, the input output dynamic reduces to *m* decoupled SISO integrator chains.

$$y_i^{(d_i)} = \nu_i \tag{E.6}$$

If the reference command to y_i is defined as r_i, $\mathbf{r}^T = \begin{pmatrix} r_1 & \dots & r_m \end{pmatrix}$, a desired steady state accurate transfer behavior of order d_i is achieved for each SISO system, by defining the pseudo control as

$$\nu_i = -M_{i,d_i-1} y_i^{(d_i-1)} - \dots - M_{i,0} y_i + M_{i,0} r_i \tag{E.7}$$

and the decoupled input-output dynamics becomes:

$$y_i(s) = \frac{M_{i,0}}{s^{d_i} + M_{i,d_i-1} s^{d_i-1} + \dots + M_{i,0}} r_i(s) \tag{E.8}$$

By inserting (E.2) into (E.7) the pseudo control can be written in terms of the state vector:

$$\boldsymbol{\nu} = -\tilde{\mathbf{F}}\mathbf{x} + \mathbf{M}_0 \mathbf{r} \tag{E.9}$$

where

$$\tilde{\mathbf{F}} = \begin{bmatrix} M_{1,d_1-1}\mathbf{c}_1^T \mathbf{A}^{d_1-1} + \dots + M_{1,0}\mathbf{c}_1^T \\ \vdots \\ M_{m,d_m-1}\mathbf{c}_m^T \mathbf{A}^{d_m-1} + \dots + M_{m,0}\mathbf{c}_m^T \end{bmatrix} \qquad \mathbf{M}_0 = \operatorname{diag}\left[M_{i,0}\right] \tag{E.10}$$

and thus written as state feedback and prefilter, the control law becomes:

$$\mathbf{u} = -\mathbf{K}\mathbf{x} + \mathbf{H}\mathbf{r} \tag{E.11}$$

with $\mathbf{K} = \mathbf{G}^{-1}\left(\mathbf{F} + \tilde{\mathbf{F}}\right)$ and $\mathbf{H} = \mathbf{G}^{-1}\mathbf{M}_0$.

Appendix F

Proofs and Bounds

F.1 Proof of Theorem 4.1

The proof is a little lengthy and therefore split up into two parts. In the first part it is shown that the rows of a specially constructed matrix are independent. Then the second part proves, that this special matrix can be written as product of two matrices, where one matrix consist of the rows of the Jacobi matrix of $\mathbf{\Phi}_\zeta$, from which the claim of Theorem 4.1 is proved.

PART 1

Initially we define the following block-structured matrix (dropping the argument $\mathbf{x}$)

$$\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \cdots & \mathbf{M}_{1r_1} \\ \vdots & & \vdots \\ \mathbf{M}_{m1} & \cdots & \mathbf{M}_{mr_1} \end{bmatrix}$$

$$\mathbf{M}_{i,j} = \begin{bmatrix} L_{\mathrm{ad}_f^{j-1}\mathbf{g}_1} L_\mathbf{f}^0 h_i & L_{\mathrm{ad}_f^{j-1}\mathbf{g}_2} L_\mathbf{f}^0 h_i & \cdots & L_{\mathrm{ad}_f^{j-1}\mathbf{g}_m} L_\mathbf{f}^0 h_i \\ L_{\mathrm{ad}_f^{j-1}\mathbf{g}_1} L_\mathbf{f}^1 h_i & L_{\mathrm{ad}_f^{j-1}\mathbf{g}_2} L_\mathbf{f}^1 h_i & \cdots & L_{\mathrm{ad}_f^{j-1}\mathbf{g}_m} L_\mathbf{f}^1 h_i \\ \vdots & \vdots & & \vdots \\ L_{\mathrm{ad}_f^{j-1}\mathbf{g}_1} L_\mathbf{f}^{r_i-1} h_i & L_{\mathrm{ad}_f^{j-1}\mathbf{g}_2} L_\mathbf{f}^{r_i-1} h_i & \cdots & L_{\mathrm{ad}_f^{j-1}\mathbf{g}_m} L_\mathbf{f}^{r_i-1} h_i \end{bmatrix}$$

Here we assume without loss of generality that $r_1 \geq r_2 \geq \ldots \geq r_m$. For clarity note, that the index i is associated with h_i, the mapping from the states to the i^{th} output, while the index j is associated with the Lie product of order *j-1* in the argument of the left Lie derivative in each expression. While the order of the right Lie derivative in the expressions increases with the *rows* in $\mathbf{M}_{i,j}$, each *column* is thereby associated with an input vector $\mathbf{g}_k$. Therefore each $\mathbf{M}_{i,j} \in \mathbb{R}^{r_i \times m}$. Consequently the composed matrix $\mathbf{M}$ has dimension $r \times m \cdot r_1$, where $r = r_1 + \ldots + r_m$. In the following if the argument $\mathbf{x}$ is omitted it is meant that $\mathbf{x}=\mathbf{x}_0$. If the system (4.1) has a relative degree, the matrix $\overline{\mathbf{B}}$ has full rank by definition. I.e. each row has at least one nonzero entry, which means, that for each output i, there is at least one input which has relative degree r_i, when the respective input/output pair is considered as SISO system. All other inputs have a relative degree

$\geq r_i$ w.r.t. to the considered output. This in turn implies that $L_{\mathbf{g}_k} L_{\mathbf{f}}^{r_i-2} h_i = 0$ for all $k = 1,\ldots,m$ and $L_{\mathbf{g}_k} L_{\mathbf{f}}^{r_i-1} h_i \neq 0$ for at least one k.

The first part of Lemma B.2 shows that, in this case $L_{\mathrm{ad}_{\mathbf{f}}^{j-1}\mathbf{g}_k} L_{\mathbf{f}}^{n} h_i = 0$, for $n<r_i-1-(j-1)=r_i-j$ for all k and further, the second part implies $L_{\mathrm{ad}_{\mathbf{f}}^{j-1}\mathbf{g}_k} L_{\mathbf{f}}^{r_i-1-(j-1)} h_i = L_{\mathbf{g}_k} L_{\mathbf{f}}^{r_i-1} h_i$ (which differs from 0 for at least one k)

Therefore

$$\mathbf{M}_{i,j} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \\ L_{\mathbf{g}_1} L_{\mathbf{f}}^{r_i-1} h_i & L_{\mathbf{g}_2} L_{\mathbf{f}}^{r_i-1} h_i & \cdots & L_{\mathbf{g}_m} L_{\mathbf{f}}^{r_i-1} h_i \\ L_{\mathrm{ad}_{\mathbf{f}}^{j-1}\mathbf{g}_1} L_{\mathbf{f}}^{r_i-1-j} h_i & L_{\mathrm{ad}_{\mathbf{f}}^{j-1}\mathbf{g}_2} L_{\mathbf{f}}^{r_i-1-j} & \cdots & \\ \vdots & \vdots & & \vdots \\ L_{\mathrm{ad}_{\mathbf{f}}^{j-1}\mathbf{g}_1} L_{\mathbf{f}}^{r_i-1} h_i & L_{\mathrm{ad}_{\mathbf{f}}^{j-1}\mathbf{g}_2} L_{\mathbf{f}}^{r_i-1} h_i & \cdots & L_{\mathrm{ad}_{\mathbf{f}}^{j-1}\mathbf{g}_m} L_{\mathbf{f}}^{r_i-1} h_i \end{bmatrix}$$

Evidently $\mathbf{M}_{i,j}$ has zeros in its upper r_i-j rows, while row r_i-j+1 obviously equals to the i^{th} row of $\overline{\mathbf{B}}(\mathbf{x}_0)$, which is, for better readability, written as

$$\mathbf{M}_{i,j} = \begin{bmatrix} \mathbf{0}^{(r_i-j)\times m} \\ \overline{\mathbf{b}}_i^T(\mathbf{x}_0) \\ * \end{bmatrix}$$

Inserting this interim result into $\mathbf{M}$ yields

$$\mathbf{M} = \begin{bmatrix} \begin{bmatrix} \mathbf{0}^T \\ \vdots \\ \mathbf{0}^T \\ \overline{\mathbf{b}}_1^T \end{bmatrix} & \begin{bmatrix} \mathbf{0}^T \\ \vdots \\ \overline{\mathbf{b}}_1^T \\ * \end{bmatrix} & \cdots & \begin{bmatrix} * \\ * \\ \vdots \\ * \end{bmatrix} \\ \vdots & \vdots & & \vdots \\ \begin{bmatrix} \mathbf{0}^T \\ \vdots \\ \mathbf{0}^T \\ \overline{\mathbf{b}}_m^T \end{bmatrix} & \begin{bmatrix} \mathbf{0}^T \\ \vdots \\ \overline{\mathbf{b}}_m^T \\ * \end{bmatrix} & \cdots & \begin{bmatrix} * \\ * \\ \vdots \\ * \end{bmatrix} \end{bmatrix}$$

Note that, since we assumed that $r_1 \geq r_i, i = 2,\ldots,m$, the last row of block matrices, say $j=r_1$, does potentially not contain some $\overline{\mathbf{b}}_i^T$, since it is "pushed" out at the top.

Now $\mathbf{M}$ is resorted in a way, that at first the last rows of each block are moved to the first m rows of a new matrix $\tilde{\mathbf{M}}$. Hence in the upper left block a matrix is created which equals to the decoupling matrix $\overline{\mathbf{B}}$. Then the same is done for the second last rows of $\mathbf{M}_{i,j}$, which are placed in the rows $m+1$ to $2m$ of $\tilde{\mathbf{M}}$. The procedure is repeated until r_m,

which equals to the smallest number of the relative degree vector. The upper $mr_m \times mr_m$ block of $\tilde{\mathbf{M}}$ looks like

$$\tilde{\mathbf{M}}_1 = \begin{bmatrix} \overline{\mathbf{B}} & * & * \\ & \ddots & * \\ \mathbf{0} & & \overline{\mathbf{B}} \end{bmatrix}$$

I.e. we have a block diagonal structure, in which the diagonal blocks equal to $\overline{\mathbf{B}}$ which is non-singular by assumption of the existence of a vector relative degree.
The rest of $\mathbf{M}$ is resorted in a similar manner, with the difference that now the block diagonal element consist only of parts of $\overline{\mathbf{B}}$. Summing up, the result is a block diagonal matrix whose block diagonal elements consist of rows of the decoupling matrix $\overline{\mathbf{B}}$.

$$\tilde{\mathbf{M}} = \begin{bmatrix} \tilde{\mathbf{M}}_1 & * \\ \mathbf{0} & \tilde{\mathbf{M}}_2 \end{bmatrix}$$

where

$$\tilde{\mathbf{M}}_2 = \begin{bmatrix} \begin{bmatrix} \vdots \\ \overline{\mathbf{b}}_{l_1} \\ \vdots \end{bmatrix} & * & * \\ * & \ddots & * \\ * & * & \begin{bmatrix} \vdots \\ \overline{\mathbf{b}}_{l_r} \\ \vdots \end{bmatrix} \end{bmatrix} \in \mathbb{R}^{m(r_1 - r_m) \times m(r_1 - r_m)}$$

The non-singularity of $\overline{\mathbf{B}}$ implies that its rows are linearly independent. Since the reordering of the rows of a matrix does not change its rank and noting that $\tilde{\mathbf{M}}$ is a block triangular matrix whose block diagonal elements consists of rows of the decoupling matrix $\mathbf{M}$ clearly has full row rank *r*.

PART 2
The composed matrix $\mathbf{M}$ can also be written as a product of matrices $\mathbf{M} = \mathbf{QP}$, where

$$\mathbf{Q} = \begin{bmatrix} \begin{bmatrix} \dfrac{\partial L_{\mathbf{f}}^0 h_1}{\partial \mathbf{x}} \\ \vdots \\ \dfrac{\partial L_{\mathbf{f}}^{r_1 - 1} h_1}{\partial \mathbf{x}} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \dfrac{\partial L_{\mathbf{f}}^0 h_m}{\partial \mathbf{x}} \\ \vdots \\ \dfrac{\partial L_{\mathbf{f}}^{r_m - 1} h_m}{\partial \mathbf{x}} \end{bmatrix} \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \begin{bmatrix} \mathrm{ad}_{\mathbf{f}}^0 \mathbf{g}_1 & \cdots & \mathrm{ad}_{\mathbf{f}}^0 \mathbf{g}_m \end{bmatrix} & \cdots & \begin{bmatrix} \mathrm{ad}_{\mathbf{f}}^{r_1 - 1} \mathbf{g}_1 & \cdots & \mathrm{ad}_{\mathbf{f}}^{r_1 - 1} \mathbf{g}_m \end{bmatrix} \end{bmatrix}$$

This is due to the fact that

$$L_{\mathrm{ad}_{\mathbf{f}}^{j-1}\mathbf{g}_k} L_{\mathbf{f}}^{n} h_i = \frac{\partial\left(L_{\mathbf{f}}^{n} h_i\right)}{\partial \mathbf{x}} \cdot \mathrm{ad}_{\mathbf{f}}^{j-1}\mathbf{g}_k \,.$$

Since, from the first part, $\mathrm{rank}(\mathbf{M}) = r$, necessarily the matrix $\mathbf{Q} \in \mathbb{R}^{r \times n}$ also has rank r, which implies that its rows are linearly independent. By considering $\mathbf{Q}$ in further detail, it reveals that its rows obviously equal to the rows of $\partial \boldsymbol{\Phi}_{\zeta} / \partial \mathbf{x}$ and hence its rows are linearly independent at $\mathbf{x}_0$. Continuity of $\boldsymbol{\Phi}_{\zeta}(\mathbf{x})$ implies that the rows of its Jacobian are linearly independent also in a vicinity of $\mathbf{x}_0$, which completes the proof.

□

F.2 Proof of Theorem 4.2

It is first shown that $\boldsymbol{\Phi}_{\eta}(\mathbf{x})$ exists. Therefore note that the rows of

$$\left.\frac{\partial \boldsymbol{\Phi}_{\zeta}}{\partial \mathbf{x}}\right|_{\mathbf{x}_0} = \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix}$$

are linearly independent by Theorem 4.1. Since $\mathbf{v}_i \in \mathcal{O}^n$ it is always possible to find additional $\mathbf{u}_i \in \mathbb{R}^n$, $i = 1, \ldots, n-r$ such that

$$\{\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_r \quad \mathbf{u}_1 \quad \cdots \quad \mathbf{u}_{n-r}\}$$

forms a basis of $\mathbb{R}^n$. For $\mathbf{u}_i^T = (u_{i,1} \quad \cdots \quad u_{i,n})$ a possible choice for the additional mappings making $\partial \boldsymbol{\Phi} / \partial \mathbf{x}$ regular is

$$\boldsymbol{\Phi}_{\eta}(\mathbf{x}) = \begin{pmatrix} u_{1,1} \cdot x_1 + \ldots + u_{1,n} \cdot x_n \\ \\ u_{n-r,1} \cdot x_1 + \ldots + u_{n-r,n} \cdot x_n \end{pmatrix}$$

which proves the first claim of Theorem 4.2.

For the second claim note that $\mathbf{g}_1(\mathbf{x}), \ldots, \mathbf{g}_m(\mathbf{x})$ are linearly independent in a vicinity of $\mathbf{x}_0$, which can be concluded from the non-singularity of $\overline{\mathbf{B}}(\mathbf{x}_0)$:

From the smoothness of the involved functions one can also conclude the non-singularity of $\overline{\mathbf{B}}(\mathbf{x})$ on $\mathcal{U}$ (which is implied by the existence of a relative degree). Moreover $\overline{\mathbf{B}}(\mathbf{x})$ also can be written as a product of matrices

$$\overline{\mathbf{B}}(\mathbf{x})=\begin{bmatrix} L_{\mathbf{g}_1}L_{\mathbf{f}}^{r_1-1}h_1(\mathbf{x}) & \cdots & L_{\mathbf{g}_m}L_{\mathbf{f}}^{r_1-1}h_1(\mathbf{x}) \\ \vdots & & \vdots \\ L_{\mathbf{g}_1}L_{\mathbf{f}}^{r_m-1}h_m(\mathbf{x}) & \cdots & L_{\mathbf{g}_m}L_{\mathbf{f}}^{r_m-1}h_m(\mathbf{x}) \end{bmatrix}=\begin{bmatrix} \frac{\partial\left(L_{\mathbf{f}}^{r_1-1}h_i(\mathbf{x})\right)}{\partial\mathbf{x}} \\ \vdots \\ \frac{\partial\left(L_{\mathbf{f}}^{r_m-1}h_m(\mathbf{x})\right)}{\partial\mathbf{x}} \end{bmatrix}\cdot\begin{bmatrix}\mathbf{g}_1(\mathbf{x}) & \cdots & \mathbf{g}_m(\mathbf{x})\end{bmatrix}$$

which highlights that necessarily $\mathbf{G}(\mathbf{x})$ is nonsingular on $\mathcal{U}$.

Therefore also the distribution G is nonsingular with constant dimension m. If additionally G is involutive Theorem B.10 suggests that there exist n-m mappings $\lambda_1(\mathbf{x}),\ldots,\lambda_{n-m}(\mathbf{x})$, such that

$$G^{\perp}(\mathbf{x})=\operatorname{span}\left\{\frac{\partial\lambda_1(\mathbf{x})}{\partial\mathbf{x}} \quad \cdots \quad \frac{\partial\lambda_{n-m}(\mathbf{x})}{\partial\mathbf{x}}\right\} \tag{F.1}$$

Consider now the codistribution spanned by the row vectors of $\partial\mathbf{\Phi}_{\zeta}(\mathbf{x})/\partial\mathbf{x}$.

$$\phi(\mathbf{x})=\operatorname{span}\left\{\frac{\partial\left(L_{\mathbf{f}}^{k}h_i(\mathbf{x})\right)}{\partial\mathbf{x}}, i=1,\ldots,m\ ;\ k=1,\ldots,r_i-1\right\} \tag{F.2}$$

which has dimension r by Theorem 4.1. Defining a codistribution

$$H(\mathbf{x})=\phi(\mathbf{x})+G^{\perp}(\mathbf{x}) \tag{F.3}$$

it is assured that one can find n linearly independent row vectors in.

$$\frac{\partial\mathbf{\Phi}_{\zeta}(\mathbf{x})}{\partial\mathbf{x}} \text{ and } \frac{\partial\lambda_1(\mathbf{x})}{\partial\mathbf{x}},\ldots,\frac{\partial\lambda_{n-m}(\mathbf{x})}{\partial\mathbf{x}}$$

if

$$\dim\left(H(\mathbf{x})\right)=n \tag{F.4}$$

Using the relations between annihilators stated in the remarks after Definition B.36, it suffices to show that

$$H^{\perp}(\mathbf{x})=\phi^{\perp}(\mathbf{x})\cap G(\mathbf{x})=\mathbf{0} \tag{F.5}$$

since $\dim\left(H(\mathbf{x})\right)+\dim\left(H^{\perp}(\mathbf{x})\right)=n$. Expressing (F.5) in words, there is no vector except the zero vector in $G(\mathbf{x})$ which annihilates all covectors in $\phi(\mathbf{x})$. In order to show this, consider a vector belonging to $G(\mathbf{x})$

$$\mathbf{g}(\mathbf{x})=c_1(\mathbf{x})\cdot\mathbf{g}_1(\mathbf{x})+\ldots+c_m(\mathbf{x})\cdot\mathbf{g}_m(\mathbf{x})$$

where $c_1(\mathbf{x}),\ldots,c_m(\mathbf{x})$ are real valued scalar functions. Then, as the rows of $\partial\mathbf{\Phi}_{\zeta}/\partial\mathbf{x}$ belong to $\phi(\mathbf{x})$, consider

$$\begin{bmatrix} \dfrac{\partial L_{\mathbf{f}}^{r_1-1} h_1(\mathbf{x})}{\partial \mathbf{x}} \\ \vdots \\ \dfrac{\partial L_{\mathbf{f}}^{r_m-1} h_m(\mathbf{x})}{\partial \mathbf{x}} \end{bmatrix} \cdot \mathbf{g}(\mathbf{x}) = \mathbf{0}$$

$$= \begin{bmatrix} \dfrac{\partial L_{\mathbf{f}}^{r_1-1} h_1(\mathbf{x})}{\partial \mathbf{x}} \\ \vdots \\ \dfrac{\partial L_{\mathbf{f}}^{r_m-1} h_m(\mathbf{x})}{\partial \mathbf{x}} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{g}_1(\mathbf{x}) & \cdots & \mathbf{g}_m(\mathbf{x}) \end{bmatrix} \begin{pmatrix} c_1(\mathbf{x}) \\ \vdots \\ c_m(\mathbf{x}) \end{pmatrix} \qquad \text{(F.6)}$$

$$= \underbrace{\begin{bmatrix} L_{\mathbf{g}_1} L_{\mathbf{f}}^{r_1-1} h_1(\mathbf{x}) & \cdots & L_{\mathbf{g}_m} L_{\mathbf{f}}^{r_1-1} h_1(\mathbf{x}) \\ \vdots & & \vdots \\ L_{\mathbf{g}_1} L_{\mathbf{f}}^{r_m-1} h_m(\mathbf{x}) & \cdots & L_{\mathbf{g}_m} L_{\mathbf{f}}^{r_m-1} h_m(\mathbf{x}) \end{bmatrix}}_{\overline{\mathbf{B}}(\mathbf{x})} \cdot \begin{pmatrix} c_1(\mathbf{x}) \\ \vdots \\ c_m(\mathbf{x}) \end{pmatrix}$$

Since however the existence of a relative degree implies the non-singularity of $\overline{\mathbf{B}}(\mathbf{x})$, $c_1(\mathbf{x}) = \ldots = c_m(\mathbf{x}) = 0$ is the only solution and consequently $\mathbf{g}(\mathbf{x}) = \mathbf{0}$. Therefore (F.5) and (F.4) are true.

It has hence been shown that in $\lambda_1(\mathbf{x}), \ldots, \lambda_{n-m}(\mathbf{x})$ there are n-r mappings (note that $r \geq m$) - without loss of generality assume that these are the first n-r $\lambda_i(\mathbf{x})$ - such that

$$\frac{\partial \mathbf{\Phi}(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial \mathbf{\Phi}_\zeta(\mathbf{x})}{\partial \mathbf{x}} \\ \dfrac{\partial \lambda_1(\mathbf{x})}{\partial \mathbf{x}} \\ \\ \dfrac{\partial \lambda_{n-r}(\mathbf{x})}{\partial \mathbf{x}} \end{bmatrix}$$

is nonsingular and

$$\frac{\partial \lambda_i(\mathbf{x})}{\partial \mathbf{x}} \cdot \mathbf{g}_k(\mathbf{x}) = 0$$

since $\partial \lambda_i(\mathbf{x}) / \partial \mathbf{x}$ belong to $G^\perp$ and the proof is complete.

□

F.3 Proof of Lemma 4.1

Remark

The theorem is formulated for the case, when system (4.1) is transformed to input-normalized Byrnes-Isidori normal form. In case this is not possible, the internal dynamics of (4.27) have to be taken instead. Using the stabilizing feedback (4.36), the internal dynamics is

$$\dot{\boldsymbol{\eta}}(t) = \mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta}) - \mathbf{P}(\boldsymbol{\zeta},\boldsymbol{\eta}) \cdot \left(\mathbf{B}^{-1}(\boldsymbol{\zeta},\boldsymbol{\eta}) \cdot \left[\mathbf{C} \cdot \boldsymbol{\zeta} + \mathbf{a}(\boldsymbol{\zeta},\boldsymbol{\eta})\right]\right)$$

and the proof can be done along the same lines.

Proof

Since the zero dynamics are assumed to be asymptotically stable, they are also uniformly asymptotically stable since the system is autonomous and by Theorem C.4 there is a continuously differentiable Lyapunov function $V_0(\boldsymbol{\eta})$ and class $\mathcal{K}$ functions $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ such that

$$\left\| \frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right\|_2 \le \alpha_1\left(\|\boldsymbol{\eta}\|_2\right) \tag{F.7}$$

$$\frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \cdot \mathbf{q}(\mathbf{0},\boldsymbol{\eta}) \le -\alpha_2\left(\|\boldsymbol{\eta}\|_2\right) \tag{F.8}$$

for $\|\boldsymbol{\eta}\|_2 < \overline{\rho}$. Further, since $\mathbf{J}\text{-}\mathbf{H}\mathbf{C}^T$ is Hurwitz, by Lemma C.4 there is a symmetric positive definite matrix $\mathbf{P}$, that satisfies the Lyapunov equation

$$\left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right)^T \mathbf{P} + \mathbf{P}\left(\mathbf{J} - \mathbf{H}\mathbf{C}^T\right) = -\mathbf{I}\,.$$

Consider the Lyapunov function candidate

$$V(\boldsymbol{\zeta},\boldsymbol{\eta}) = V_0(\boldsymbol{\eta}) + k\sqrt{\boldsymbol{\zeta}^T \mathbf{P} \boldsymbol{\zeta}}$$

where $k>0$ is determined later. The time derivative long the system trajectories is

$$\begin{aligned}\dot{V} &= \frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \cdot \mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta}) + \frac{k}{2\sqrt{\boldsymbol{\zeta}^T \mathbf{P} \boldsymbol{\zeta}}} \boldsymbol{\zeta}^T \cdot \left(\left(\mathbf{J} - \mathbf{H}\mathbf{C}\right)^T \mathbf{P} + \mathbf{P}\left(\mathbf{J} - \mathbf{H}\mathbf{C}\right)\right) \cdot \boldsymbol{\zeta} \\ &= \frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \cdot \mathbf{q}(\mathbf{0},\boldsymbol{\eta}) + \frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \left(\mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta}) - \mathbf{q}(\mathbf{0},\boldsymbol{\eta})\right) - \frac{k}{2\sqrt{\boldsymbol{\zeta}^T \mathbf{P} \boldsymbol{\zeta}}} \boldsymbol{\zeta}^T \boldsymbol{\zeta}\end{aligned}$$

Remark

Note that $\dot{V}$ is not defined on $\boldsymbol{\zeta}=\mathbf{0}$. Therefore this case has to be considered separately. Since the dynamics of $\boldsymbol{\zeta}(t)$ is a linear one, there are only 2 possibilities:

- $\boldsymbol{\zeta}_0 \neq \mathbf{0}$, then $\boldsymbol{\zeta}(t) \neq \mathbf{0}$ for all $t \ge t_0$
- $\boldsymbol{\zeta}_0 = \mathbf{0}$ then $\boldsymbol{\zeta}(t) \equiv \mathbf{0}$ for all $t \ge t_0$

In the second case, the internal dynamics reduce to the zero dynamics, which are asymptotically stable by assumption. Therefore also

$$\lim_{t\to\infty}(\boldsymbol{\eta}(t)) = \mathbf{0}$$

and for the remainder we assume that $||\boldsymbol{\zeta}(t)||>0$ for all $t\geq t_0$.

Since $\mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta})$ is continuously differentiable on the compact set $\overline{\mathcal{B}_r}$, Theorem B.8 implies that

$$\left\| \frac{\partial \mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial \boldsymbol{\zeta}} \right\|_2 \leq M$$

on $\overline{\mathcal{B}_r}$ for some $M>0$. Since $\overline{\mathcal{B}_r}$ is convex, it is shown by Lemma 3.1 in [Kha02] that

$$\|\mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta}) - \mathbf{q}(\mathbf{0},\boldsymbol{\eta})\| \leq M \cdot \|\boldsymbol{\zeta}\| \qquad \text{(F.9)}$$

on $\overline{\mathcal{B}_r}$. With bounds (F.7) and (F.9) the Lyapunov function derivative is bounded from above by

$$\dot{V} \leq -\alpha_2(\|\boldsymbol{\eta}\|_2) + M \cdot \alpha_1(\|\boldsymbol{\eta}\|_2) \cdot \|\boldsymbol{\zeta}\|_2 - \frac{k}{\overline{\lambda}_P} \cdot \|\boldsymbol{\zeta}\|_2 .$$

Moreover, on the compact set

$$\Omega_\rho = \left\{ \boldsymbol{\eta} \in \mathbb{R}^{n-r} \,\middle|\, \|\boldsymbol{\eta}\|_2 \leq \rho \right\}$$

for some $0 < \rho < \overline{\rho}$, $\alpha_1(\|\boldsymbol{\eta}\|_2) \leq \alpha_1(\rho)$ and hence

$$\dot{V} \leq -\alpha_2(\|\boldsymbol{\eta}\|_2) - \left(\frac{k}{\sqrt{\overline{\lambda}_P}} - M \cdot \alpha_1(\rho) \right) \cdot \|\boldsymbol{\zeta}\|_2$$

on Ω_ρ. If we choose

$$k = \sqrt{\overline{\lambda}_P}(M \cdot \alpha_1(\rho) + k_1)$$

for some $k_1>0$, then

$$\dot{V} \leq -\alpha_2(\|\boldsymbol{\eta}\|_2) - k_1 \|\boldsymbol{\zeta}\|_2$$

and the conditions of Theorem C.1 are fulfilled with the positive definite functions

$$W_1(\boldsymbol{\zeta},\boldsymbol{\eta}) = V_0(\boldsymbol{\eta}) + \sqrt{\overline{\lambda}_P} \cdot \|\boldsymbol{\zeta}\|_2$$

$$W_2(\boldsymbol{\zeta},\boldsymbol{\eta}) = V_0(\boldsymbol{\eta}) + \sqrt{\underline{\lambda}_P} \cdot \|\boldsymbol{\zeta}\|_2$$

$$W_3(\boldsymbol{\zeta},\ ) = \alpha_2(\|\boldsymbol{\eta}\|_2) + k_1 \|\boldsymbol{\zeta}\|_2$$

Hence the system is (locally asymptotically stable.

□

F.4 Proof of Theorem 4.3

The set $\overline{\mathcal{B}}_{r_0}$ is defined in terms of the vector $\mathbf{z}^T = \begin{pmatrix}\boldsymbol{\zeta}^T & \boldsymbol{\eta}^T\end{pmatrix}$ but the proof will employ an equivalent set in terms of the error vector $\mathbf{z}_e^T = \begin{pmatrix}\mathbf{e}^T & \boldsymbol{\eta}^T\end{pmatrix}$, which is defined as

$$\overline{\Omega}_{r_0,t} = \left\{ \mathbf{z}_e := \begin{pmatrix}\mathbf{e}\\ \boldsymbol{\eta}\end{pmatrix} \in \mathbb{R}^n \;\middle|\; \mathbf{z} = \begin{pmatrix}\boldsymbol{\zeta}_R(t)+\mathbf{e}\\ \boldsymbol{\eta}\end{pmatrix} \in \overline{\mathcal{B}}_{r_0} \right\}$$

This set is time varying since it depends on the current value of the reference state $\boldsymbol{\zeta}_R(t)$. Since $\mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta})$ continuously differentiable on the compact set $\overline{\mathcal{B}}_{r0}$, its Jacobian adopts some maximum (Theorem B.8), i.e.

$$\left\| \frac{\partial \mathbf{q}(\boldsymbol{\zeta},\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right\|_2 \le M \tag{F.10}$$

for some $M>0$. As $\mathbf{q}(\mathbf{0},\boldsymbol{\eta})$is asymptotically stable, by Theorem C.4 there is a continuously differentiable Lyapunov function $V_0(\boldsymbol{\eta})$ and class $\mathcal{K}$ functions $\alpha_1(\cdot)$, $\alpha_2(\cdot)$, $\alpha_3(\cdot)$ and $\alpha_4(\cdot)$ such that

$$\alpha_1\left(\|\boldsymbol{\eta}\|_2\right) \le V_0(\boldsymbol{\eta}) \le \alpha_2\left(\|\boldsymbol{\eta}\|_2\right) \tag{F.11}$$

$$\left\| \frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \right\|_2 \le \alpha_3\left(\|\boldsymbol{\eta}\|_2\right) \tag{F.12}$$

$$\frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \cdot \mathbf{q}(\mathbf{0},\boldsymbol{\eta}) \le -\alpha_4\left(\|\boldsymbol{\eta}\|_2\right) \tag{F.13}$$

for $\|\boldsymbol{\eta}\|_2<\bar{\rho}$ and some $\bar{\rho} \in (0, r_\eta]$. Moreover, since $\mathbf{J}-\mathbf{H}\mathbf{C}^T$ is Hurwitz, by Lemma C.4 there is a symmetric positive definite matrix $\mathbf{P}$, that satisfies the Lyapunov equation

$$\left(\mathbf{J}-\mathbf{H}\mathbf{C}^T\right)^T\mathbf{P} + \mathbf{P}\left(\mathbf{J}-\mathbf{H}\mathbf{C}^T\right) = -\mathbf{I} \tag{F.14}$$

where $\mathbf{I}$ is the identity matrix. Then consider the Lyapunov function candidate for the whole system

$$V(\mathbf{e},\boldsymbol{\eta}) = V_0(\boldsymbol{\eta}) + k\sqrt{\mathbf{e}^T\mathbf{P}\mathbf{e}} \tag{F.15}$$

where $k>0$ is determined later. Note that, by (F.11), $V(\boldsymbol{\zeta},\boldsymbol{\eta})$ is bounded by

$$\alpha_1\left(\|\boldsymbol{\eta}\|_2\right) + \sqrt{\underline{\lambda}_P}\cdot\|\mathbf{e}\|_2 \le V(\mathbf{e},\boldsymbol{\eta}) \le \alpha_2\left(\|\boldsymbol{\eta}\|_2\right) + \sqrt{\overline{\lambda}_P}\cdot\|\mathbf{e}\|_2 \tag{F.16}$$

if $\|\boldsymbol{\eta}\|_2\le\rho$ for some $\rho<\bar{\rho}$ and all $\mathbf{e}\in\mathbb{R}^r$ where $\underline{\lambda}_P$ and $\overline{\lambda}_P$ denote the minimum and maximum eigenvalue of $\mathbf{P}$ respectively. The time derivative along the system trajectories is

$$\dot{V} = \frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \cdot \mathbf{q}(\zeta_R + \mathbf{e}, \boldsymbol{\eta}) + \frac{k}{2\sqrt{\mathbf{e}^T \mathbf{P} \mathbf{e}}} \mathbf{e}^T \cdot \left(\left(\mathbf{J} - \mathbf{H}\mathbf{C}^T \right)^T \mathbf{P} + \mathbf{P}\left(\mathbf{J} - \mathbf{H}\mathbf{C}^T \right) \right) \cdot \mathbf{e}$$
$$= \frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \cdot \mathbf{q}(\mathbf{0}, \boldsymbol{\eta}) + \frac{\partial V_0(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}} \cdot (\mathbf{q}(\zeta_R + \mathbf{e}, \boldsymbol{\eta}) - \mathbf{q}(\mathbf{0}, \boldsymbol{\eta})) - \frac{k}{2\sqrt{\mathbf{e}^T \mathbf{P} \mathbf{e}}} \mathbf{e}^T \mathbf{e}$$

which is well-defined on a set of vectors $\mathbf{z}_e^T \in \overline{\Omega}_{r_0,t}$ such that $\|\boldsymbol{\eta}\|_2 < \rho$ since V_0 is defined for $\|\boldsymbol{\eta}\|_2 \leq \rho$ and $\mathbf{q}(\,,\,)$ is defined on $\overline{\Omega}_{r_0,t}$.

$$\overline{\Omega}_{\rho,t} = \left\{ \mathbf{z}_e := \begin{pmatrix} \mathbf{e} \\ \boldsymbol{\eta} \end{pmatrix} \in \overline{\Omega}_{r_0} \;\middle|\; \|\boldsymbol{\eta}\|_2 \leq \rho \right\}$$

Remark

Note that $\dot{V}$ is not defined on $\mathbf{e}=\mathbf{0}$. Therefore this case has to be considered separately. Since the dynamics of $\mathbf{e}(t)$ is a linear one, there are only 2 possibilities:

- $\mathbf{e}_0 \neq \mathbf{0}$, then $\mathbf{e}(t) \neq \mathbf{0}$ for all $t \geq t_0$
- $\mathbf{e}_0 = \mathbf{0}$ then $\mathbf{e}(t) \equiv \mathbf{0}$ for all $t \geq t_0$

In the second case the internal dynamics can be considered separately by using $V_0(\boldsymbol{\eta})$. But we will postpone this case and at first assume that $\|\mathbf{e}\|_2 > 0$ for all $t \geq t_0$.

By (F.10) and mean value theorem, we get

$$\|\mathbf{q}(\zeta, \boldsymbol{\eta}) - \mathbf{q}(\mathbf{0}, \boldsymbol{\eta})\|_2 \leq M \cdot \|\zeta\|_2 \leq M \cdot \|\zeta_R\|_2 + M \cdot \|\mathbf{e}\|_2 \qquad \text{(F.17)}$$

on $\overline{\Omega}_{\rho,t}$ and with bounds (F.12), (F.13) and (F.17) the Lyapunov function derivative becomes

$$\dot{V} \leq -\alpha_4(\|\boldsymbol{\eta}\|_2) + M \cdot \alpha_3(\|\boldsymbol{\eta}\|_2) \cdot \|\zeta_R\|_2 - \left(\frac{k}{2\sqrt{\bar{\lambda}_P}} - M \cdot \alpha_3(\|\boldsymbol{\eta}\|_2) \right) \cdot \|\mathbf{e}\|_2 \qquad \text{(F.18)}$$

on $\overline{\Omega}_{\rho,t}$. Moreover, if the reference states are bounded such that $\|\zeta_R\|_2 \leq \overline{\zeta}$ for some

$$\overline{\zeta} < r_\zeta \qquad \text{(F.19)}$$

then $\mathbf{z}_e \in \overline{\Omega}_{\rho,t}$, if $\|\mathbf{e}\|_2 \leq r_\zeta - \overline{\zeta}$ and $\|\boldsymbol{\eta}\|_2 \leq \rho$, independent of $\zeta_R(t)$. Figure F.1 illustrates the situation.

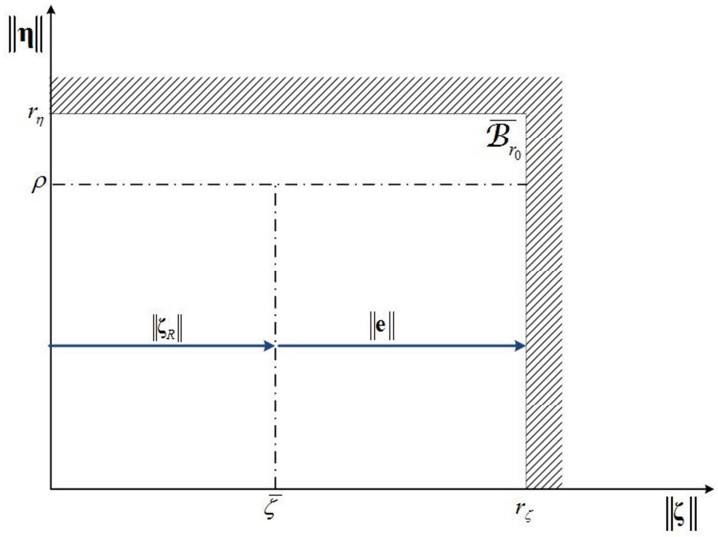


Figure F.1 Situation Lyapunov Function

I.e. the set

$$\overline{\Omega}_\rho = \left\{ \mathbf{z}_e = \begin{pmatrix} \mathbf{e} \\ \boldsymbol{\eta} \end{pmatrix} \in \mathbb{R}^n \,\middle|\, \|\mathbf{e}\|_2 \le r_\zeta - \bar{\zeta} \,,\, \|\boldsymbol{\eta}\|_2 \le \rho \right\} \tag{F.20}$$

is time invariant and lies within $\overline{\Omega}_{\rho,t}$ for all $t \geq t_0$. On $\overline{\Omega}_\rho$, a more conservative bound on $\dot{V}$ is given by

$$\dot{V} \le -\alpha_4\left(\|\boldsymbol{\eta}\|_2\right) + M \cdot \alpha_3(\rho) \cdot \bar{\zeta} - \left(\frac{k}{2\bar{\lambda}_P} - M \cdot \alpha_3(\rho) \right) \cdot \|\mathbf{e}\|_2 \tag{F.21}$$

and the last term is negative definite if we choose

$$k = 2\sqrt{\bar{\lambda}_P} \cdot \left(k_1 + M \cdot \alpha_3(\rho)\right) \tag{F.22}$$

for some $k_1 > 0$ and we arrive at

$$\dot{V} \le -\alpha_4\left(\|\boldsymbol{\eta}\|_2\right) + M \cdot \alpha_3(\rho) \cdot \bar{\zeta} - k_1 \cdot \|\mathbf{e}\|_2 \tag{F.23}$$

Furthermore, if

$$\|\mathbf{e}\|_2 \ge \frac{M \cdot \alpha_3(\rho) \cdot \bar{\zeta} + k_2}{k_1} =: \mu_e \text{ or } \|\boldsymbol{\eta}\|_2 \ge \alpha_4^{-1}\left(M \cdot \alpha_3(\rho) \cdot \bar{\zeta} + k_2\right) =: \mu_\eta \tag{F.24}$$

for some $k_2 > 0$, then

$$\dot{V} \le -k_2 \tag{F.25}$$

In other words, (F.25) holds, if $\mathbf{z}_e$ is in $\overline{\Omega}_\rho \setminus \Omega_\mu$, where

$$\Omega_\mu = \left\{ \begin{pmatrix} \mathbf{e} \\ \boldsymbol{\eta} \end{pmatrix} \in \mathbb{R}^n \,\middle|\, \|\mathbf{e}\|_2 < \mu_e, \|\boldsymbol{\eta}\|_2 < \mu_\eta \right\}$$

Switching to the nomenclature of Corollary 3.1, we define class $\mathcal{K}$ functions:

$$\gamma_\eta\left(\|\boldsymbol{\eta}\|_2\right)=\begin{cases}\dfrac{k_2}{\mu_\eta}\|\boldsymbol{\eta}\|_2 & \text{for } 0\le\|\boldsymbol{\eta}\|_2<\mu_\eta\\ \alpha_4\left(\|\boldsymbol{\eta}\|_2\right)-M\alpha_3(\rho)\overline{\zeta} & \text{for } \mu_\eta\le\|\boldsymbol{\eta}\|_2\le\rho\end{cases} \tag{F.26}$$

$$\gamma_e\left(\|\mathbf{e}\|_2\right)=\begin{cases}\dfrac{k_2}{\mu_e}\|\mathbf{e}\|_2 & \text{for } 0\le\|\mathbf{e}\|_2<\mu_e\\ k_1\|\mathbf{e}\|_2-M\alpha_3(\rho)\overline{\zeta} & \text{for } \mu_e\le\|\mathbf{e}\|_2\le r_\zeta-\overline{\zeta}\end{cases} \tag{F.27}$$

(which are continuous by the definition of μ_e and μ_η) and get

$$\dot{V}(\mathbf{x},t)\le-\max\left(\gamma_\eta\left(\|\boldsymbol{\eta}\|_2\right),\gamma_e\left(\|\mathbf{e}\|_2\right)\right) \tag{F.28}$$

on the set $\overline{\Omega}_\rho\setminus\Omega_\mu$ (respect (F.23) and Figure F.2).

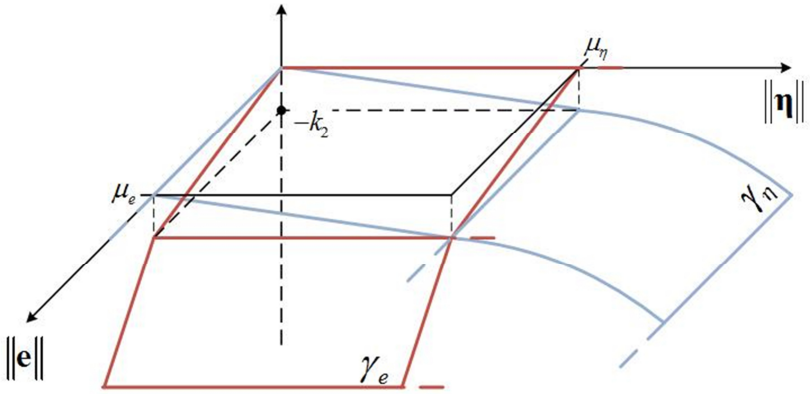


Figure F.2 Class $\mathcal{K}$ Bounds on Lyapunov Function Derivative

In light of (F.16) and by considering Corollary 3.1, for ultimate boundedness we need to fulfill

$$\mu_e<\frac{\underline{u}}{2\sqrt{\overline{\lambda}_P}}\ \text{and}\ \mu_\eta<\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right) \tag{F.29}$$

where

$$\underline{u}=\min\left(\sqrt{\underline{\lambda}_P}\cdot\left(r_\zeta-\overline{\zeta}\right),\alpha_1(\rho)\right)$$

By definition of (F.24), we obtain the following inequality constraints:

$$\frac{M\cdot\alpha_3(\rho)\cdot\overline{\zeta}+k_2}{k_1}<\frac{\underline{u}}{2\sqrt{\overline{\lambda}_P}}\ \text{and}\ \alpha_4^{-1}\left(M\cdot\alpha_3(\rho)\cdot\overline{\zeta}+k_2\right)<\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)$$

The first inequality can be easily fulfilled by setting k_1 sufficiently large

$$k_1>2\sqrt{\overline{\lambda}_P}\cdot\frac{M\cdot\alpha_3(\rho)\cdot\overline{\zeta}+k_2}{\underline{u}} \tag{F.30}$$

while the second inequality is fulfilled if

$$0 < \overline{\zeta} < \frac{\alpha_4\left(\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right)}{M \cdot \alpha_3(\rho)} \tag{F.31}$$

and

$$k_2 < \alpha_4\left(\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right) - M \cdot \alpha_3(\rho) \cdot \overline{\zeta} \tag{F.32}$$

Notice that (F.31) is an implicit inequality since the definition of $\underline{u}$ also contains $\overline{\zeta}$. With condition (F.32), the constraint on k_1 in (F.30) is fulfilled independently of k_2, if we choose

$$k_1 \geq 2\sqrt{\overline{\lambda}_P} \cdot \frac{M \cdot \alpha_3(\rho) \cdot \overline{\zeta} + \left(\alpha_4\left(\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right) - M \cdot \alpha_3(\rho) \cdot \overline{\zeta}\right)}{\underline{u}} = \frac{2\sqrt{\overline{\lambda}_P}}{\underline{u}} \cdot \alpha_4\left(\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right)$$

Summing up so far, if (F.31) holds then we can choose:

- $k_2 \in \left(0, \alpha_4\left(\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right) - M \cdot \alpha_3(\rho) \cdot \overline{\zeta}\right)$
- $k_1 \geq \frac{2\sqrt{\overline{\lambda}(\mathbf{P})}}{\underline{u}} \cdot \alpha_4\left(\alpha_2^{-1}\left(\frac{\underline{u}}{2}\right)\right)$

such that (F.29) is fulfilled and Corollary 3.1 guarantees uniform ultimate boundedness such that

$$\|\mathbf{e}(t)\| \leq b_e$$

for all $t \geq t_0 + T_e(\delta_e, b_e)$, $\|\mathbf{e}(t_0)\|_2 \leq \delta_e$, $\delta_e \in \left(0, \left(2\sqrt{\overline{\lambda}_P}\right)^{-1} \underline{u}\right]$ and

$\|\boldsymbol{\eta}(t)\| \leq b_\eta$ for all $t \geq t_0 + T_\eta(\delta_\eta, b_\eta)$, $\|\boldsymbol{\eta}(t_0)\|_2 \leq \delta_\eta$, $\delta_\eta \in \left[0, \alpha_2^{-1}(0.5\underline{u})\right]$, where

$$b_e = \frac{\sqrt{\overline{\lambda}_P} \cdot \mu_e + \alpha_2(\mu_\eta)}{\sqrt{\underline{\lambda}_P}}, \tag{F.33}$$

$$b_\eta = \alpha_1^{-1}\left(\sqrt{\overline{\lambda}_P} \cdot \mu_e + \alpha_2(\mu_\eta)\right) \tag{F.34}$$

μ_e, μ_η defined in (F.24) and

$$T_e(\delta_e, b_e) = \frac{\sqrt{\overline{\lambda}_P} \cdot (\delta_e - \mu_e) + \frac{\underline{u}}{2} - \alpha_2(\mu_\eta)}{k_2}, \tag{F.35}$$

$$T_\eta(\delta_\eta, b_\eta) = \frac{\frac{u}{2} - \sqrt{\bar{\lambda}_P} \cdot \mu_e + \alpha_2(\delta_\eta) - \alpha_2(\mu_\eta)}{k_2} \tag{F.36}$$

Note that we used $\max(\gamma_1(\mu_1), \gamma_2(\mu_2)) = k_2$ (see Figure F.2) for computation of T_e and T_η.

It is left to analyze the case $\mathbf{e}(t) = \mathbf{0}$ for all $t \geq t_0$. Then the system reduces to $\dot{\boldsymbol{\eta}}(t) = \mathbf{q}(\boldsymbol{\zeta}_R, \boldsymbol{\eta})$ and we can use $V_0(\boldsymbol{\eta})$ as Lyapunov function whose time derivative is, according to (F.21), bounded by,

$$\dot{V}_0 \leq -\alpha_4(\|\boldsymbol{\eta}\|_2) + M \cdot \alpha_3(\rho) \cdot \bar{\zeta}$$

provided $\zeta_R(t) \leq \bar{\zeta}$ for some $\bar{\zeta} < r_\zeta$. Then if

$$\|\boldsymbol{\eta}\|_2 \geq \alpha_4^{-1}(M\alpha_3(\rho)\bar{\zeta} + k_2) := \mu_\eta \tag{F.37}$$

for some $k_2 > 0$, then $\dot{V} \leq -k_2$ and we can employ class $\mathcal{K}$ function (F.26) such that

$$\dot{V}_0 \leq -\gamma_\eta(\|\boldsymbol{\eta}\|_2)$$

for $\mu_\eta \leq \|\boldsymbol{\eta}\|_2 \leq \rho$. By (F.11) and in light of Corollary 3.2, $\|\boldsymbol{\eta}(t)\|_2$ is uniformly ultimately bounded if

$$\mu_\eta < \alpha_2^{-1}(\alpha_1(\rho)). \tag{F.38}$$

If

$$\bar{\zeta} < \frac{\alpha_4(\alpha_2^{-1}(\alpha_1(\rho)))}{M\alpha_3(\rho)} \tag{F.39}$$

we can choose a $k_2 > 0$ such that

$$k_2 < \alpha_4(\alpha_2^{-1}(\alpha_1(\rho))) - M\alpha_3(\rho)\bar{\zeta} \tag{F.40}$$

and (F.38) is fulfilled. Note that (F.39) and (F.40) are fulfilled if (F.31) and (F.32) are fulfilled, which means that the case $\mathbf{e}(t) = \mathbf{0}$ does not impose any stronger conditions on ultimate boundedness. Then, by Corollary 3.2:

$\|\boldsymbol{\eta}(t)\|_2 \leq b_\eta$ for all $t \geq t_0 + T_{\eta,0}(\delta_\eta, b_\eta)$, $\|\boldsymbol{\eta}(t_0)\| \leq \delta_{\eta,0}$ and $\delta_{\eta,0} \in [0, \alpha_2^{-1}(\alpha_1(\rho))]$ where

$$b_{\eta,0} = \alpha_2^{-1}(\alpha_1(\mu_\eta)) \text{ and } T_{\eta,0}(\delta_\eta, b_\eta) = \frac{\alpha_2(\delta_\eta) - \alpha_2(\mu_\eta)}{k_2}$$

Finally remark that $b_{\eta,0} \leq b_\eta$, b_η defined in (F.34), since $\mu_e \geq 0$. Also $T_{\eta,0} \leq T_\eta$, T_η defined in (F.36), since (F.29) implies that $\frac{u}{2} - \sqrt{\bar{\lambda}_P} \cdot \mu_e \geq 0$. In other words, the case $\mathbf{e}(t) = \mathbf{0}$ assures a lower ultimate bound in a shorter ultimate time than the case $\mathbf{e}(t) \neq \mathbf{0}$ and hence the former case is incorporated in the latter one.

Although Theorem 4.3 only proves ultimate boundedness for $\mathbf{e}(t)$ and $\boldsymbol{\eta}(t)$ it is known from linear system theory that

$$\lim_{t\to\infty}\left(\left\|\mathbf{e}(t)\right\|_2\right)=0$$

since $\mathbf{J}-\mathbf{H}\mathbf{C}^T$ is Hurwitz (provided the solution $\mathbf{z}_e(t)$) exists for all $t\geq t_0$ which is assured by ultimate boundedness). Moreover, Lemma C.4 even implies that $\mathbf{e}(t)$ is exponentially stable such that

$$\left\|\mathbf{e}(t)\right\|_2\leq\left\|\mathbf{e}_0\right\|_2\sqrt{\frac{\overline{\lambda}_P}{\underline{\lambda}_P}}\exp\left(-\frac{1}{2\overline{\lambda}_P}\cdot(t-t_0)\right)$$

where $\mathbf{P}$ solves (F.14).

□

F.5 Bounds for Lyapunov Analysis of Variant 2 Part 2

Bound on Indefinite Term p

This section is concerned with computation of an upper bound on the indefinite term (4.334) in section 4.6.2. It consists of 5 expressions, on which upper bounds are computed in the following.

Upper bound on expression 1

By application of Lemma B.4 and assumption F, we obtain

$$\left\|\mathbf{g}\left(\mathbf{u}_N^*+\tilde{\mathbf{u}}_N\right)-\mathbf{w}^*\right\|_2\leq k_2\left\|\tilde{\mathbf{u}}_N\right\|_2. \tag{F.41}$$

Upper bound on expression 2

Analogous considerations, as for equation (4.146), yield

$$\left\|\mathbf{H}^T\mathbf{P}_R\mathbf{e}\right\|_2\leq\overline{\lambda}_{P_R}\left\|\mathbf{e}\right\|_2 \tag{F.42}$$

Upper bound on expression 3

An upper bound on $\dot{\zeta}$ is obtained from equation (4.271).

$$\left\|\dot{\boldsymbol{\zeta}}\right\|_2\leq\left\|\mathbf{J}\right\|_2\left\|\boldsymbol{\zeta}\right\|_2+\left\|\mathbf{H}\right\|_2\left[\left\|\mathbf{v}\right\|_2+\left\|\boldsymbol{\Theta}_x\right\|_2\left\|\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})\right\|_2+b(\boldsymbol{\zeta},\boldsymbol{\eta})\left\|\mathbf{B}_L\right\|_2\left\|\tilde{\boldsymbol{\Lambda}}_L\right\|_2\left\|\mathbf{u}\right\|_2+\left\|\hat{\mathbf{g}}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})-\mathbf{w}^*\right\|_2+\left\|\boldsymbol{\delta}(\boldsymbol{\zeta},\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right\|_2\right]$$

Then, by equations (4.18), (4.20), (4.112), assumption D, I, J, equation (F.41) and if the plant states ζ are restricted to their valid domain $\overline{\mathcal{B}_{\bar{\zeta}}}$ we obtain

$$\left\|\dot{\boldsymbol{\zeta}}\right\|_2 \le \bar{\zeta} + \left\|\mathbf{v}\right\|_2 + \left\|\boldsymbol{\Theta}_x\right\|_2 F + \bar{b}\left(\left\|\hat{\boldsymbol{\Lambda}}_L\right\|_2 + \left\|\boldsymbol{\Lambda}_L\right\|_2\right)\left\|\mathbf{B}_L\right\|_2\left\|\mathbf{u}\right\|_2 + k_2\left\|\tilde{\mathbf{u}}_N\right\|_2 + D. \tag{F.43}$$

Next, we have to find upper bounds on $\mathbf{v}$ and $\mathbf{u}$. With (4.122), (4.273), (4.277), (4.278), (4.281), (4.282), we get

$$\begin{aligned}
\left\|\mathbf{v}\right\|_2 &\le \left\|\hat{\mathbf{v}}_R\right\|_2 + \left\|\mathbf{v}_E\right\|_2 + \left\|\mathbf{v}_A\right\|_2 \\
&\le \left\|\mathbf{K}\right\|_2\left\|\hat{\boldsymbol{\zeta}}\right\|_2 + \left\|\mathbf{A}_0\right\|_2\left\|\mathbf{y}_C\right\|_2 + \left\|\mathbf{C}\right\|_2\left\|\hat{\mathbf{e}}\right\|_2 + \left\|\hat{\boldsymbol{\Theta}}_x\right\|_2\left\|\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})\right\|_2 \\
&\le \left\|\mathbf{K}\right\|_2\left(\left\|\boldsymbol{\zeta}\right\|_2 + \left\|\hat{\mathbf{e}}\right\|_2\right) + \left\|\mathbf{A}_0\right\|_2\left\|\mathbf{y}_C\right\|_2 + \left\|\mathbf{C}\right\|_2\left\|\hat{\mathbf{e}}\right\|_2 + \left\|\hat{\boldsymbol{\Theta}}_x\right\|_2\left\|\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})\right\|_2.
\end{aligned}$$

According to assumption H for the upper bound on the state dependent regressor, bound on the exogenous input $\mathbf{y}_C$ in equation (4.285) and if the plant states $\boldsymbol{\zeta}$ are restricted to their valid domain $\overline{\mathcal{B}_{\bar{\zeta}}}$, we get

$$\left\|\mathbf{v}\right\|_2 \le \left(\left\|\mathbf{K}\right\|_2 + \left\|\mathbf{C}\right\|_2\right)\left\|\hat{\mathbf{e}}\right\|_2 + \left\|\mathbf{K}\right\|_2\bar{\zeta} + \left\|\mathbf{A}_0\right\|_2\bar{y} + \left\|\hat{\boldsymbol{\Theta}}_x\right\|_2 F\,. \tag{F.44}$$

Further, by definition of weighted Frobenius norm (1.13), and relation (1.15), we have

$$\bar{\lambda}_{\Gamma_x}^{-1}\left\|\hat{\boldsymbol{\Theta}}_x\right\|_F^2 \le \mathrm{tr}\left[\hat{\boldsymbol{\Theta}}_x^T\boldsymbol{\Gamma}_x^{-1}\hat{\boldsymbol{\Theta}}_x\right] = \left\|\hat{\boldsymbol{\Theta}}_x\right\|_{\Gamma_x}^2 \qquad \Leftrightarrow \qquad \left\|\hat{\boldsymbol{\Theta}}_x\right\|_F^2 \le \bar{\lambda}_{\Gamma_x}\left\|\hat{\boldsymbol{\Theta}}_x\right\|_{\Gamma_x}^2$$

with (1.4) and bound (4.133), imposed by the projection operator, we obtain

$$\left\|\hat{\boldsymbol{\Theta}}_x\right\|_2 \le \left\|\hat{\boldsymbol{\Theta}}_x\right\|_F \le \sqrt{\bar{\lambda}_{\Gamma_x}}\left\|\hat{\boldsymbol{\Theta}}_x\right\|_{\Gamma_x} \le \sqrt{\bar{\lambda}_{\Gamma_x}}\,\theta_{x,\max}\,. \tag{F.45}$$

An analogous consideration, together with assumption D yields

$$\left\|\boldsymbol{\Theta}_x\right\|_2 \le \left\|\boldsymbol{\Theta}_x\right\|_F \le \sqrt{\bar{\lambda}_{\Gamma_x}}\left\|\boldsymbol{\Theta}_x\right\|_{\Gamma_x} \le \sqrt{\bar{\lambda}_{\Gamma_x}}\,\theta_x\,. \tag{F.46}$$

Using (F.45) for (F.44) yields

$$\left\|\mathbf{v}\right\|_2 \le \left(\left\|\mathbf{K}\right\|_2 + \left\|\mathbf{C}\right\|_2\right)\left\|\hat{\mathbf{e}}\right\|_2 + \left\|\mathbf{K}\right\|_2\bar{\zeta} + \left\|\mathbf{A}_0\right\|_2\bar{y} + \theta_{x,\max}\sqrt{\bar{\lambda}_{\Gamma_x}}\,F\,. \tag{F.47}$$

In order to find an upper bound on $\mathbf{u}$, take 2-norm of (4.267) and use assumption I.

$$\left\|\mathbf{u}\right\|_2 \le \left\|\hat{\mathbf{B}}(\boldsymbol{\zeta},\boldsymbol{\eta})\right\|_2\left\|\left[\hat{\mathbf{B}}(\boldsymbol{\zeta},\boldsymbol{\eta})\hat{\mathbf{B}}^T(\boldsymbol{\zeta},\boldsymbol{\eta}) + \mathbf{I}\right]^{-1}\right\|_2\left[\left\|\mathbf{v}\right\|_2 + \left\|\hat{\mathbf{K}}_x\right\|_2 F\right] \tag{F.48}$$

Now, since $\hat{\mathbf{B}}(\boldsymbol{\zeta},\boldsymbol{\eta})\hat{\mathbf{B}}^T(\boldsymbol{\zeta},\boldsymbol{\eta})$ is symmetric positive definite, we have for its minimum eigenvalue ([Lüt96] p.153-6)

$$\underline{\lambda}\left[\hat{\mathbf{B}}(\boldsymbol{\zeta},\boldsymbol{\eta})\hat{\mathbf{B}}^T(\boldsymbol{\zeta},\boldsymbol{\eta}) + \mathbf{I}\right] \ge 1$$

which renders $\hat{\mathbf{B}}(\boldsymbol{\zeta},\boldsymbol{\eta})\hat{\mathbf{B}}^T(\boldsymbol{\zeta},\boldsymbol{\eta}) + \mathbf{I}$ a positive definite symmetric matrix. Further, since singular values and eigenvalues of symmetric positive definite matrices are equal (Theorem B.17), we have, by [Lüt96] p.79-4,

$$\overline{\sigma}\left(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]^{-1}\right)=\left(\underline{\lambda}\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]\right)^{-1}.$$

and hence the following inequality holds for the matrix induced 2-norm.

$$\left\|\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]^{-1}\right\|_2\leq 1 \tag{F.49}$$

For computation of an upper bound on $\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})$, we have to insert either of absolute scaling (4.111) or relative scaling (4.114), dependent on which one is implemented. Together with (4.184), we have

$$\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})=b(\zeta,\boldsymbol{\eta})\mathbf{B}_L\hat{\boldsymbol{\Lambda}}_L \quad , \quad \hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})=b(\zeta,\boldsymbol{\eta})\mathbf{B}_L\left(\mathbf{I}+\hat{\boldsymbol{\Lambda}}_L\right).$$

Further, by analogous considerations, as for inequalities (F.45), (F.46), we get

$$\left\|\hat{\boldsymbol{\Lambda}}_L\right\|_2\leq\left\|\hat{\boldsymbol{\Lambda}}_L\right\|_F\leq\sqrt{\overline{\lambda}_{\Gamma_L}}\left\|\hat{\boldsymbol{\Lambda}}_L\right\|_{\Gamma_L}\leq\sqrt{\overline{\lambda}_{\Gamma_L}}\lambda_{L,\max} \tag{F.50}$$

$$\left\|\boldsymbol{\Lambda}_L\right\|_2\leq\left\|\boldsymbol{\Lambda}_L\right\|_F\leq\sqrt{\overline{\lambda}_{\Gamma_L}}\left\|\boldsymbol{\Lambda}_L\right\|_{\Gamma_L}\leq\sqrt{\overline{\lambda}_{\Gamma_L}}\lambda_L \tag{F.51}$$

where $\lambda_{L,\max}$, defined in (4.135), is the bound, imposed by projection operator and λ_L is an upper bound on the ideal parameter in assumption D. With (1.4) and assumption I for a bound on $b(\zeta,\boldsymbol{\eta})$, we get

$$\left\|\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\right\|_2\leq\overline{b}\left\|\mathbf{B}_L\right\|_2\sqrt{\overline{\lambda}_{\Gamma_L}}\lambda_{L,\max} \quad , \quad \left\|\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\right\|_2\leq\overline{b}\left\|\mathbf{B}_L\right\|_2\left(1+\sqrt{\overline{\lambda}_{\Gamma_L}}\lambda_{L,\max}\right). \tag{F.52}$$

Defining constants

$$k_4=\sqrt{\overline{\lambda}_{\Gamma_L}}\lambda_{L,\max} \text{ or } k_4=1+\sqrt{\overline{\lambda}_{\Gamma_L}}\lambda_{L,\max} \tag{F.53}$$

dependent on, whether absolute or relative scaling has been used, and using results (F.47), (F.49), and (F.52), we obtain for (F.48)

$$\left\|\mathbf{u}\right\|_2\leq\overline{b}\left\|\mathbf{B}_L\right\|_2k_4\left[\left(\left\|\mathbf{K}\right\|_2+\left\|\mathbf{C}\right\|_2\right)\left\|\hat{\mathbf{e}}\right\|_2+\left\|\mathbf{K}\right\|_2\overline{\zeta}+\left\|\mathbf{A}_0\right\|_2\overline{y}+\left(\theta_{x,\max}\sqrt{\overline{\lambda}_{\Gamma_x}}+\left\|\hat{\mathbf{K}}_x\right\|_2\right)F\right]. \tag{F.54}$$

Applying results (F.46), (F.47), (F.50), (F.51), (F.54) and assumptions H, I to (F.43) yields

$$\left\|\dot{\boldsymbol{\zeta}}\right\|_2\leq Z_{\hat{e}}\left\|\hat{\mathbf{e}}\right\|_2+k_2\left\|\tilde{\mathbf{u}}_N\right\|_2+Z_0\,. \tag{F.55}$$

Thereby:

$$\begin{aligned}
Z_{\hat{e}}&=\left(1+\sqrt{\overline{\lambda}_{\Gamma_L}}\left(\lambda_{L,\max}+\lambda_L\right)\left(\left\|\mathbf{B}_L\right\|_2\overline{b}\right)^2k_4\right)\left(\left\|\mathbf{K}\right\|_2+\left\|\mathbf{C}\right\|_2\right)\\
Z_0&=\left(1+\left\|\mathbf{K}\right\|_2\right)\overline{\zeta}+\left\|\mathbf{A}_0\right\|_2\overline{y}+\sqrt{\overline{\lambda}_{\Gamma_x}}\left(\theta_x+\theta_{x,\max}\right)F+D\\
&\quad+\sqrt{\overline{\lambda}_{\Gamma_L}}\left(\lambda_{L,\max}+\lambda_L\right)\left(\left\|\mathbf{B}_L\right\|_2\overline{b}\right)^2k_4\left(\left\|\mathbf{K}\right\|_2\overline{\zeta}+\left\|\mathbf{A}_0\right\|_2\overline{y}+\left(\theta_{x,\max}\sqrt{\overline{\lambda}_{\Gamma_x}}+\left\|\hat{\mathbf{K}}_x\right\|_2\right)F\right)
\end{aligned} \tag{F.56}$$

Finally, with assumption H, we get for expression 3

$$\left\|\mathbf{J}_{\zeta}\left(\mathbf{u}_N^{*}+\tilde{\mathbf{u}}_N\right)\dot{\zeta}\right\|_2 \le M_{\zeta} Z_{\hat{e}} \left\|\hat{\mathbf{e}}\right\|_2 + M_{\zeta} k_2 \left\|\tilde{\mathbf{u}}_N\right\|_2 + M_{\zeta} Z_0 \,. \tag{F.57}$$

Upper bound on expression 4

With bound on $\mathbf{u}$, (F.55) and assumption J, an upper bound on the internal dynamics is

$$\left\|\dot{\boldsymbol{\eta}}\right\|_2 \le Y_{\hat{e}} \left\|\hat{\mathbf{e}}\right\|_2 + Y_0 \tag{F.58}$$

where

$$\begin{aligned} Y_{\hat{e}} &= \bar{p}\bar{b}\left\|\mathbf{B}_L\right\|_2 k_4 \left(\left\|\mathbf{K}\right\|_2 + \left\|\mathbf{C}\right\|_2\right) \\ Y_0 &= \bar{p}\bar{b}\left\|\mathbf{B}_L\right\|_2 k_4 \left(\left\|\mathbf{K}\right\|_2 \bar{\zeta} + \left\|\mathbf{A}_0\right\|_2 \bar{y} + \left(\theta_{x,\max}\sqrt{\bar{\lambda}_{\boldsymbol{\Gamma}_x}} + \left\|\hat{\mathbf{K}}_x\right\|_2\right) F\right) + \bar{q}. \end{aligned} \tag{F.59}$$

By assumption G, we obtain

$$\left\|\mathbf{J}_{\eta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\dot{\boldsymbol{\eta}}\right\|_2 \le M_{\eta} Y_{\hat{e}} \left\|\hat{\mathbf{e}}\right\|_2 + M_{\eta} Y_0 \,. \tag{F.60}$$

Upper bound on expression 5

For this expression, (4.323) has to be derived w.r.t. time.

$$\begin{aligned} \dot{\mathbf{w}}^{*} = {} & \frac{d}{dt}\left(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^{T}(\zeta,\boldsymbol{\eta}) + \mathbf{I}\right]^{-1}\right)\left[\mathbf{v} - \hat{\mathbf{K}}_x \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})\right] \\ & + \left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^{T}(\zeta,\boldsymbol{\eta}) + \mathbf{I}\right]^{-1}\left[\dot{\mathbf{v}} - \hat{\mathbf{K}}_x \frac{\partial \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})}{\partial \zeta}\dot{\zeta} - \hat{\mathbf{K}}_x \frac{\partial \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\dot{\boldsymbol{\eta}}\right] \end{aligned} \tag{F.61}$$

An upper bound is obtained using matrix induced 2-norms.

$$\begin{aligned} \left\|\dot{\mathbf{w}}^{*}\right\|_2 \le {} & \overbrace{\left\|\frac{d}{dt}\left(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^{T}(\zeta,\boldsymbol{\eta}) + \mathbf{I}\right]^{-1}\right)\right\|_2}^{5a}\left[\left\|\mathbf{v}\right\|_2 + \left\|\hat{\mathbf{K}}_x\right\|_2 \left\|\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})\right\|_2\right] \\ & + \left\|\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^{T}(\zeta,\boldsymbol{\eta}) + \mathbf{I}\right]^{-1}\right\|_2 \left[\overbrace{\left\|\dot{\mathbf{v}}\right\|_2}^{5b} + \left\|\hat{\mathbf{K}}_x\right\|_2 \left(\left\|\frac{\partial \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})}{\partial \zeta}\right\|_2 \left\|\dot{\zeta}\right\|_2 + \left\|\frac{\partial \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right\|_2 \left\|\dot{\boldsymbol{\eta}}\right\|\right)\right] \end{aligned} \tag{F.62}$$

Computation of bounds on expressions 5a and 5b is quite involved and hence will be accomplished before within auxiliary computations.

Auxiliary Computation 5a

The time derivative of the inverse of the decoupling matrix is obtained by derivation of the identity matrix.

$$\frac{d(\mathbf{I})}{dt} = \frac{d}{dt}\left(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^{T}(\zeta,\boldsymbol{\eta}) + \mathbf{I}\right]^{-1}\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^{T}(\zeta,\boldsymbol{\eta}) + \mathbf{I}\right]\right) = \mathbf{0}$$

$$\frac{d}{dt}\Big(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]^{-1}\Big)\cdot\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]+\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]^{-1}\cdot\frac{d}{dt}\Big(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]\Big)=\mathbf{0}$$

$$\frac{d}{dt}\Big(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]^{-1}\Big)=-\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]^{-1}\frac{d}{dt}\Big(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]\Big)\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]^{-1}$$

An upper bound on the matrix induced 2-norm is obtained by (1.2) and inequality (F.49).

$$\left\|\frac{d}{dt}\Big(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]^{-1}\Big)\right\|_2\le\left\|\frac{d}{dt}\Big(\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]\Big)\right\|_2 \tag{F.63}$$

Further, with absolute or relative scaling ((4.111), (4.114)) together with (4.184) the decoupling matrix reads

$$\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]=b^2(\zeta(t),\boldsymbol{\eta}(t))\left[\mathbf{B}_L\hat{\boldsymbol{\Lambda}}_L(t)\hat{\boldsymbol{\Lambda}}_L^T(t)\mathbf{B}_L^T+\mathbf{I}\right] \tag{F.64}$$

$$\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]=b^2(\zeta(t),\boldsymbol{\eta}(t))\left[\mathbf{B}_L\left(\hat{\boldsymbol{\Lambda}}_L(t)+\mathbf{I}\right)\left(\hat{\boldsymbol{\Lambda}}_L(t)+\mathbf{I}\right)^T\mathbf{B}_L^T+\mathbf{I}\right] \tag{F.65}$$

All dependencies on time have been highlighted. The time derivative evaluates to

$$\begin{aligned}&\frac{d}{dt}\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]=\\&2b(\zeta,\boldsymbol{\eta})\left[\mathbf{B}_L\hat{\boldsymbol{\Lambda}}_L\hat{\boldsymbol{\Lambda}}_L^T\mathbf{B}_L^T+\mathbf{I}\right]\cdot\left(\frac{\partial b(\zeta,\boldsymbol{\eta})}{\partial\zeta}\dot{\zeta}+\frac{\partial b(\zeta,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\dot{\boldsymbol{\eta}}\right)+b^2(\zeta,\boldsymbol{\eta})\cdot\left[\mathbf{B}_L\dot{\hat{\boldsymbol{\Lambda}}}_L\hat{\boldsymbol{\Lambda}}_L^T\mathbf{B}_L^T+\mathbf{B}_L\hat{\boldsymbol{\Lambda}}_L\dot{\hat{\boldsymbol{\Lambda}}}_L^T\mathbf{B}_L^T\right]\end{aligned} \tag{F.66}$$

$$\begin{aligned}&\frac{d}{dt}\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]=\\&2b(\zeta,\boldsymbol{\eta})\cdot\left[\mathbf{B}_L\left(\hat{\boldsymbol{\Lambda}}_L+\mathbf{I}\right)\left(\hat{\boldsymbol{\Lambda}}_L+\mathbf{I}\right)^T\mathbf{B}_L^T+\mathbf{I}\right]\cdot\left(\frac{\partial b(\zeta,\boldsymbol{\eta})}{\partial\zeta}\dot{\zeta}+\frac{\partial b(\zeta,\boldsymbol{\eta})}{\partial\boldsymbol{\eta}}\dot{\boldsymbol{\eta}}\right)+b^2(\zeta,\boldsymbol{\eta})\cdot\left[\mathbf{B}_L\dot{\hat{\boldsymbol{\Lambda}}}_L\hat{\boldsymbol{\Lambda}}_L^T\mathbf{B}_L^T+\mathbf{B}_L\hat{\boldsymbol{\Lambda}}_L\dot{\hat{\boldsymbol{\Lambda}}}_L^T\mathbf{B}_L^T\right]\end{aligned} \tag{F.67}$$

and with (F.50), (F.52), assumption I and definition of k_4 in (F.53), we obtain.

$$\begin{aligned}&\left\|\frac{d}{dt}\left[\hat{\mathbf{B}}(\zeta,\boldsymbol{\eta})\hat{\mathbf{B}}^T(\zeta,\boldsymbol{\eta})+\mathbf{I}\right]\right\|_2\le\\&2\bar{b}\left[k_4^2\|\mathbf{B}_L\|_2^2+1\right]\left(\bar{b}_\zeta\|\dot{\boldsymbol{\xi}}\|_2+\bar{b}_\eta\|\dot{\boldsymbol{\eta}}\|_2\right)+2\bar{b}^2\|\mathbf{B}_L\|_2^2\sqrt{\bar{\lambda}_{\Gamma_L}}\lambda_{L,\max}\left\|\dot{\hat{\boldsymbol{\Lambda}}}_L\right\|_2\end{aligned} \tag{F.68}$$

Next, we have to establish a bound on $\dot{\hat{\boldsymbol{\Lambda}}}_L$. Considering update law (4.294) with modification (4.296)

$$\dot{\hat{\boldsymbol{\Lambda}}}_L^T=\mathrm{Proj}\left[\begin{matrix}\hat{\boldsymbol{\Lambda}}_L^T, & -\boldsymbol{\Gamma}_L b(\zeta,\boldsymbol{\eta})\mathbf{u}\hat{\mathbf{e}}^T\mathbf{P}_E\mathbf{H}\mathbf{B}_L\\ & -\sigma_L\bar{f}_{\varepsilon_L,\lambda_L,\boldsymbol{\Gamma}_L}\left(\hat{\boldsymbol{\Lambda}}_L^T\right)\hat{\boldsymbol{\Lambda}}_L^T-\kappa\mathbf{B}_L^T\mathbf{H}^T\tilde{\mathbf{C}}\frac{\mathbf{Q}_L^T}{\xi}\boldsymbol{\Gamma}_L\end{matrix}\right]_{\varepsilon_L,\lambda_{L,\max},\boldsymbol{\Gamma}_L} \tag{F.69}$$

From (D.4) in Appendix D.1 and inequalities (1.4), (1.15), we have

$$\left\|\dot{\hat{\mathbf{\Lambda}}}_L^T\right\|_2 \leq \sqrt{\bar{\lambda}_{\mathbf{\Gamma}_L}}\left\|-\mathbf{\Gamma}_L b(\zeta,\eta)\mathbf{u}\hat{\mathbf{e}}^T\mathbf{P}_E\mathbf{H}\mathbf{B}_L - \sigma_L \bar{f}_{\varepsilon_L,\lambda_L,\mathbf{\Gamma}_L}\left(\hat{\mathbf{\Lambda}}_L^T\right)\hat{\mathbf{\Lambda}}_L^T - \kappa\mathbf{B}_L^T\mathbf{H}^T\tilde{\mathbf{C}}\frac{\mathbf{Q}_L^T}{\xi}\mathbf{\Gamma}_L\right\|_{\mathbf{\Gamma}_L}.$$

Then, with triangle inequality, bound (4.135), imposed by projection operator, definition of switching function $\bar{f}_{\varepsilon_L,\lambda_L,\mathbf{\Gamma}_L}(\cdot)$ in equation (D.12) and assumption I, we get

$$\left\|\dot{\hat{\mathbf{\Lambda}}}_L\right\|_2 \leq \sqrt{\bar{\lambda}_{\mathbf{\Gamma}_L}}\bar{b}\left\|\mathbf{\Gamma}_L\mathbf{u}\hat{\mathbf{e}}^T\mathbf{P}_E\mathbf{H}\mathbf{B}_L\right\|_{\mathbf{\Gamma}_L} + \sigma_L\sqrt{\bar{\lambda}_{\mathbf{\Gamma}_L}}\lambda_{L,\max} + \kappa\sqrt{\bar{\lambda}_{\mathbf{\Gamma}_L}}\left\|\mathbf{B}_L^T\mathbf{H}^T\tilde{\mathbf{C}}\frac{\mathbf{Q}_L^T}{\xi}\mathbf{\Gamma}_L\right\|_{\mathbf{\Gamma}_L}.$$

According to (1.15), the weighted Frobenius norms are replaced by their respective Frobenius norms, we insert the definition of $\tilde{\mathbf{C}}$ from (4.305) and cancel the term $\mathbf{H}^T\mathbf{H}$ according to (4.17).

$$\left\|\dot{\hat{\mathbf{\Lambda}}}_L\right\|_2 \leq \sqrt{\frac{\bar{\lambda}_{\mathbf{\Gamma}_L}}{\underline{\lambda}_{\mathbf{\Gamma}_L}}}\bar{b}\left\|\mathbf{\Gamma}_L\mathbf{u}\hat{\mathbf{e}}^T\mathbf{P}_E\mathbf{H}\mathbf{B}_L\right\|_F + \sigma_L\sqrt{\bar{\lambda}_{\mathbf{\Gamma}_L}}\lambda_{L,\max} + \kappa\sqrt{\frac{\bar{\lambda}_{\mathbf{\Gamma}_L}}{\underline{\lambda}_{\mathbf{\Gamma}_L}}}\left\|\mathbf{B}_L^T\left(\tilde{\mathbf{W}}^T\mathbf{Q}-\mathbf{Q}_\delta\right)\frac{\mathbf{Q}_L^T}{\xi}\mathbf{\Gamma}_L\right\|_F$$

By help of (1.8), we split up the single matrices, while we account for the fact that, for vectors, Frobenius norm and Euclidean vector norm are equal according to (1.6).

$$\begin{aligned}\left\|\dot{\hat{\mathbf{\Lambda}}}_L\right\|_2 \leq{}& \sqrt{\frac{\bar{\lambda}_{\mathbf{\Gamma}_L}}{\underline{\lambda}_{\mathbf{\Gamma}_L}}}\bar{b}\left\|\mathbf{\Gamma}_L\mathbf{u}\right\|_2\left\|\hat{\mathbf{e}}^T\mathbf{P}_E\mathbf{H}\mathbf{B}_L\right\|_2 + \sigma_L\sqrt{\bar{\lambda}_{\mathbf{\Gamma}_L}}\lambda_{L,\max}\\ &+ \kappa\sqrt{\frac{\bar{\lambda}_{\mathbf{\Gamma}_L}}{\underline{\lambda}_{\mathbf{\Gamma}_L}}}\left\|\mathbf{B}_L\right\|_F\left(\left\|\tilde{\mathbf{W}}\right\|_F\frac{\left\|\mathbf{Q}\right\|_F\left\|\mathbf{Q}_L\right\|_F}{\xi} + \frac{\left\|\mathbf{Q}_\delta\right\|_F\left\|\mathbf{Q}_L\right\|_F}{\xi}\right)\left\|\mathbf{\Gamma}_L\right\|_F\end{aligned} \tag{F.70}$$

Now we need relations, involving some terms within this expression.

1. Relation between filtered regressors

$$\left\|\mathbf{Q}\right\|_F^2 = \mathrm{tr}\left(\begin{bmatrix}\mathbf{Q}_x^T & \mathbf{Q}_L^T\end{bmatrix}\begin{bmatrix}\mathbf{Q}_x\\ \mathbf{Q}_L\end{bmatrix}\right) = \mathrm{tr}\left(\mathbf{Q}_x^T\mathbf{Q}_x + \mathbf{Q}_L^T\mathbf{Q}_L\right) = \left\|\mathbf{Q}_L\right\|_F^2 + \left\|\mathbf{Q}_x\right\|_F^2 \geq \left\|\mathbf{Q}_L\right\|_F^2$$

$$\left\|\mathbf{Q}_L\right\|_F \leq \left\|\mathbf{Q}\right\|_F \tag{F.71}$$

2. Normed expression 1

$$f_1\left(\left\|\mathbf{Q}\right\|_F\right) := \frac{\left\|\mathbf{Q}\right\|_F}{\xi} = \frac{\left\|\mathbf{Q}\right\|_F}{1+\left\|\mathbf{Q}\right\|_F^2}$$

$$f_1'\left(\left\|\mathbf{Q}\right\|_F\right) = \frac{1-\left\|\mathbf{Q}\right\|_F^2}{\left(1+\left\|\mathbf{Q}\right\|_F^2\right)^2} \overset{!}{=} 0 \Leftrightarrow \left(\left\|\mathbf{Q}\right\|_F\right)_{opt} = 1 \Leftrightarrow f_1\left(\left(\left\|\mathbf{Q}\right\|_F\right)_{opt}\right) = \frac{1}{2}$$

$$f_1\left(\left\|\mathbf{Q}\right\|_F\right) := \frac{\left\|\mathbf{Q}\right\|_F}{1+\left\|\mathbf{Q}\right\|_F^2} \leq \frac{1}{2} \tag{F.72}$$

3. Normed expression 2

$$f_2\left(\left\|\mathbf{Q}\right\|_F\right) := \frac{\left\|\mathbf{Q}\right\|_F^2}{\xi} = \frac{\left\|\mathbf{Q}\right\|_F^2}{1+\left\|\mathbf{Q}\right\|_F^2} < 1 \tag{F.73}$$

4. Bound on $\tilde{\mathbf{W}}$

$$\begin{aligned}\left\|\tilde{\mathbf{W}}\right\|_F^2 &= \mathrm{tr}\left(\tilde{\mathbf{W}}^T\tilde{\mathbf{W}}\right) = \mathrm{tr}\left(\begin{bmatrix}\tilde{\boldsymbol{\Theta}}_x^T & \mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\end{bmatrix}\begin{bmatrix}\tilde{\boldsymbol{\Theta}}_x \\ \tilde{\boldsymbol{\Lambda}}_L^T\mathbf{B}_L^T\end{bmatrix}\right) \\ &= \mathrm{tr}\left(\tilde{\boldsymbol{\Theta}}_x^T\tilde{\boldsymbol{\Theta}}_x + \mathbf{B}_L\tilde{\boldsymbol{\Lambda}}_L\tilde{\boldsymbol{\Lambda}}_L^T\mathbf{B}_L^T\right) \le \left\|\tilde{\boldsymbol{\Theta}}_x\right\|_F^2 + \left\|\mathbf{B}_L\right\|_F^2\left\|\tilde{\boldsymbol{\Lambda}}_L\right\|_F^2 \\ &\le \left(\left\|\hat{\boldsymbol{\Theta}}_x\right\|_F + \left\|\boldsymbol{\Theta}_x\right\|_F\right)^2 + \left\|\mathbf{B}_L\right\|_F^2\left(\left\|\hat{\boldsymbol{\Lambda}}_L\right\|_F + \left\|\boldsymbol{\Lambda}_L\right\|_F\right)^2\end{aligned}$$

With (F.45), (F.46), (F.50), (F.51), we get

$$\left\|\tilde{\mathbf{W}}\right\|_F^2 \le \bar{\lambda}_{\Gamma_x}\left(\theta_{x,\max} + \theta_x\right)^2 + \left\|\mathbf{B}_L\right\|_F^2\bar{\lambda}_{\Gamma_L}\left(\lambda_{L,\max} + \lambda_L\right)^2 =: \bar{w}^2 \tag{F.74}$$

5. Bound on $\left\|\boldsymbol{\Gamma}_L\mathbf{u}\right\|_2$

According to (1.1) and Theorem B.17, the matrix induced 2-norm of the symmetric positive definite matrix $\boldsymbol{\Gamma}_L$ evaluates to $\left\|\boldsymbol{\Gamma}_L\right\|_2 = \bar{\lambda}_{\Gamma_L}$ and hence

$$\left\|\boldsymbol{\Gamma}_L\mathbf{u}\right\|_2 \le \bar{\lambda}_{\Gamma_L}\left\|\mathbf{u}\right\|_2 \tag{F.75}$$

6. Bound on $\left\|\hat{\mathbf{e}}^T\mathbf{P}_E\mathbf{H}\mathbf{B}_L\right\|_2$

Analogously to (F.42), we obtain

$$\left\|\hat{\mathbf{e}}^T\mathbf{P}_E\mathbf{H}\mathbf{B}_L\right\|_2 \le \left\|\mathbf{B}_L\right\|_2\left\|\mathbf{P}_E\mathbf{H}\right\|_2\left\|\hat{\mathbf{e}}\right\|_2$$

$$\left\|\hat{\mathbf{e}}^T\mathbf{P}_E\mathbf{H}\mathbf{B}_L\right\|_2 \le \bar{\lambda}_{P_E}\left\|\mathbf{B}_L\right\|_2\left\|\hat{\mathbf{e}}\right\|_2 \tag{F.76}$$

With relations 1 – 6, bound on $\mathbf{u}$ – equation (F.54) – and assumption C together with (4.233) – we obtain for (F.70)

$$\left\|\dot{\hat{\boldsymbol{\Lambda}}}_L\right\|_2 \le \Lambda_{\hat{e}^2}\left\|\hat{\mathbf{e}}\right\|_2^2 + \Lambda_{\hat{e}}\left\|\hat{\mathbf{e}}\right\|_2 + \Lambda_0 \tag{F.77}$$

where

$$\Lambda_{\hat{e}^2} = \sqrt{\frac{\overline{\lambda}_{\Gamma_L}^3}{\underline{\lambda}_{\Gamma_L}}} \overline{\lambda}_{P_E} \left(\|\mathbf{B}_L\|_2 \overline{b} \right)^2 k_4 \left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2 \right)$$

$$\Lambda_{\hat{e}} = \sqrt{\frac{\overline{\lambda}_{\Gamma_L}^3}{\underline{\lambda}_{\Gamma_L}}} \overline{\lambda}_{P_E} \left(\|\mathbf{B}_L\|_2 \overline{b} \right)^2 k_4 \left(\|\mathbf{K}\|_2 \overline{\zeta} + \|\mathbf{A}_0\|_2 \overline{y} + \left(\theta_{x,\max} \sqrt{\overline{\lambda}_{\Gamma_x}} + \left\| \hat{\mathbf{K}}_x \right\|_2 \right) F \right) \quad \text{(F.78)}$$

$$\Lambda_0 = \sigma_L \sqrt{\overline{\lambda}_{\Gamma_L}} \lambda_{L,\max} + \kappa \sqrt{\frac{\overline{\lambda}_{\Gamma_L}}{\underline{\lambda}_{\Gamma_L}}} \|\mathbf{B}_L\|_F \left(\overline{w} + \frac{mLD\sqrt{N}}{2} \right) \|\mathbf{\Gamma}_L\|_F .$$

Inserting bounds on $\dot{\zeta}$, $\dot{\boldsymbol{\eta}}$ and $\dot{\hat{\mathbf{\Lambda}}}_L$ in (F.55), (F.58) and (F.77) and inequality (F.68) into (F.63) yields

$$\left\| \frac{d}{dt} \left(\left[\hat{\mathbf{B}}(\zeta, \boldsymbol{\eta}) \hat{\mathbf{B}}^T(\zeta, \boldsymbol{\eta}) + \mathbf{I} \right]^{-1} \right) \right\|_2 \leq B_{\hat{e}} \|\hat{\mathbf{e}}\|_2 + B_{\hat{e}^2} \|\hat{\mathbf{e}}\|_2^2 + B_u \|\tilde{\mathbf{u}}_N\|_2 + B_0 \quad \text{(F.79)}$$

where

$$\begin{aligned}
B_{\hat{e}} &= 2\overline{b} \left[\left(k_4^2 \|\mathbf{B}_L\|_2^2 + 1 \right) \left(\overline{b}_\zeta Z_{\hat{e}} + b_\eta Y_{\hat{e}} \right) + \overline{b} \|\mathbf{B}_L\|_2^2 \sqrt{\overline{\lambda}_{\Gamma_L}} \lambda_{L,\max} \Lambda_{\hat{e}} \right] \\
B_{\hat{e}^2} &= 2\overline{b}^2 \|\mathbf{B}_L\|_2^2 \sqrt{\overline{\lambda}_{\Gamma_L}} \lambda_{L,\max} \Lambda_{\hat{e}^2} \\
B_0 &= 2\overline{b} \left[\overline{b} \|\mathbf{B}_L\|_2^2 \sqrt{\overline{\lambda}_{\Gamma_L}} \lambda_{L,\max} \Lambda_0 + \left(k_4^2 \|\mathbf{B}_L\|_2^2 + 1 \right) \left(\overline{b}_\zeta Z_0 + \overline{b}_\eta Y_0 \right) \right] \\
B_u &= 2\overline{b} \left[k_4^2 \|\mathbf{B}_L\|_2^2 + 1 \right] \overline{b}_\zeta k_2
\end{aligned} \quad \text{(F.80)}$$

Auxiliary computation 5b

For time derivative of pseudo control, consider (4.277) with (4.122), (4.273), (4.278), (4.281), (4.282). Taking the time derivative yields

$$\dot{\mathbf{v}} = -\mathbf{K} \left(\dot{\zeta} + \dot{\hat{\mathbf{e}}} \right) + \mathbf{A}_0 \dot{\mathbf{y}}_C + \mathbf{C} \dot{\hat{\mathbf{e}}} - \dot{\hat{\mathbf{\Theta}}}_x^T \boldsymbol{\varphi}(\zeta, \boldsymbol{\eta}) - \hat{\mathbf{\Theta}}_x^T \left(\frac{\partial \boldsymbol{\varphi}(\zeta, \boldsymbol{\eta})}{\partial \zeta} \dot{\zeta} + \frac{\partial \boldsymbol{\varphi}(\zeta, \boldsymbol{\eta})}{\partial \zeta} \dot{\boldsymbol{\eta}} \right). \quad \text{(F.81)}$$

Additionally to assumption in section 4.6.2, the time derivative of the exogenous command $\dot{\mathbf{y}}_C$ has to be bounded such that

$$\|\dot{\mathbf{y}}_C(t)\|_2 \leq \overline{y}_d \text{ for all } t \geq t_0 \quad \text{(F.82)}$$

for some $\overline{y}_d > 0$. With assumption H, inequality (F.45), an upper bound for (F.81) is

$$\|\dot{\mathbf{v}}\|_2 \leq \left(\|\mathbf{K}\|_2 + \sqrt{\overline{\lambda}_{\Gamma_x}} \theta_{x,\max} F_\zeta \right) \|\dot{\zeta}\|_2 + \|\mathbf{A}_0\|_2 \overline{y}_d + \left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2 \right) \|\dot{\hat{\mathbf{e}}}\|_2 + \left\| \dot{\hat{\mathbf{\Theta}}}_x \right\|_2 F + \sqrt{\overline{\lambda}_{\Gamma_x}} \theta_{x,\max} F_\eta \|\dot{\boldsymbol{\eta}}\|_2 \quad \text{(F.83)}$$

Next, we compute an upper bound on $\dot{\hat{\mathbf{\Theta}}}_x$. Therefore consider update law (4.294) with modification (4.296).

$$\dot{\hat{\boldsymbol{\Theta}}}_x = \mathrm{Proj}\left[\hat{\boldsymbol{\Theta}}_x, -\boldsymbol{\Gamma}_x \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})\hat{\mathbf{e}}^T \mathbf{P}_E \mathbf{H} - \sigma_x \bar{f}_{\varepsilon_x,\theta_x,\boldsymbol{\Gamma}_x}\left(\hat{\boldsymbol{\Theta}}_x\right)\hat{\boldsymbol{\Theta}}_x - \kappa \boldsymbol{\Gamma}_x \frac{\mathbf{Q}_x}{\xi}\tilde{\mathbf{C}}^T \mathbf{H}\right]_{\varepsilon_x,\theta_{x,\max},\boldsymbol{\Gamma}_x} \tag{F.84}$$

By analogous considerations as in auxiliary computation 5a, we arrive at an expression, similar to (F.70).

$$\begin{aligned}\left\|\dot{\hat{\boldsymbol{\Theta}}}_x\right\|_2 \le{}& \sqrt{\frac{\bar{\lambda}_{\Gamma_x}}{\underline{\lambda}_{\Gamma_x}}}\left\|\boldsymbol{\Gamma}_x \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})\right\|_2 \left\|\hat{\mathbf{e}}^T \mathbf{P}_E \mathbf{H}\right\|_2 + \sigma_x \sqrt{\bar{\lambda}_{\Gamma_x}}\theta_{x,\max} \\ &+ \kappa \sqrt{\frac{\bar{\lambda}_{\Gamma_x}}{\underline{\lambda}_{\Gamma_x}}}\left(\left\|\tilde{\mathbf{W}}\right\|_F \frac{\|\mathbf{Q}\|_F \|\mathbf{Q}_x\|_F}{\xi} + \frac{\|\mathbf{Q}_x\|_F \|\mathbf{Q}_\delta\|_F}{\xi}\right)\left\|\boldsymbol{\Gamma}_x\right\|_F\end{aligned} \tag{F.85}$$

With the analog to (F.71)

$$\left\|\mathbf{Q}_x\right\|_F \le \left\|\mathbf{Q}\right\|_F \tag{F.86}$$

analogues to inequalities (F.72), (F.73), inequality (F.74), the analog to (F.75)

$$\left\|\boldsymbol{\Gamma}_x \boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})\right\| \le \bar{\lambda}_{\Gamma_x} F \tag{F.87}$$

where we have used assumption H, inequality (F.76) and assumption C together with (4.233) we obtain for (F.85)

$$\left\|\dot{\hat{\boldsymbol{\Theta}}}_x\right\|_2 \le \sqrt{\frac{\bar{\lambda}_{\Gamma_x}^3}{\underline{\lambda}_{\Gamma_x}}}\bar{\lambda}_{P_E} F \left\|\hat{\mathbf{e}}\right\|_2 + \sigma_x \sqrt{\bar{\lambda}_{\Gamma_x}}\theta_{x,\max} + \kappa \sqrt{\frac{\bar{\lambda}_{\Gamma_x}}{\underline{\lambda}_{\Gamma_x}}}\left(\bar{w} + \frac{mLD\sqrt{N}}{2}\right)\left\|\boldsymbol{\Gamma}_x\right\|_F. \tag{F.88}$$

Further, a bound on $\dot{\hat{\mathbf{e}}}$ is obtained from (4.280) by use of matrix induced 2-norm, (4.18) and (4.20)

$$\left\|\dot{\hat{\mathbf{e}}}\right\|_2 = \left(1 + \|\mathbf{C}\|_2\right)\left\|\hat{\mathbf{e}}\right\|_2 + \left(\left\|\hat{\boldsymbol{\Theta}}_x\right\|_2 + \left\|\boldsymbol{\Theta}_x\right\|_2\right)\left\|\boldsymbol{\varphi}(\zeta,\boldsymbol{\eta})\right\|_2 + \left|b(\zeta,\boldsymbol{\eta})\right|\left\|\mathbf{B}_L\right\|_2\left(\left\|\hat{\boldsymbol{\Lambda}}_L\right\|_2 + \left\|\boldsymbol{\Lambda}_L\right\|_2\right)\left\|\mathbf{u}\right\|_2 + \left\|\boldsymbol{\delta}(\zeta,\boldsymbol{\eta},\mathbf{u}_N,\mathbf{d})\right\|_2.$$

With bounds (F.45), (F.46), (F.50), (F.51), assumptions D, I, J, and bound on $\mathbf{u}$ in (F.54), we get

$$\left\|\dot{\hat{\mathbf{e}}}\right\|_2 \le E_{\hat{e}}\left\|\hat{\mathbf{e}}\right\|_2 + E_0 \tag{F.89}$$

where

$$\begin{aligned}E_{\hat{e}} &= 1 + \|\mathbf{C}\|_2 + \sqrt{\bar{\lambda}_{\Gamma_L}}\left(\lambda_{L,\max} + \lambda_L\right)\left(\bar{b}\|\mathbf{B}_L\|_2\right)^2 k_4\left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2\right) \\ E_0 &= \sqrt{\bar{\lambda}_{\Gamma_x}}\left(\theta_{x,\max} + \theta_x\right)F + D \\ &\quad + \sqrt{\bar{\lambda}_{\Gamma_L}}\left(\lambda_{L,\max} + \lambda_L\right)\left(\bar{b}\|\mathbf{B}_L\|_2\right)^2 k_4\left(\|\mathbf{K}\|_2 \bar{\zeta} + \|\mathbf{A}_0\|_2 \bar{y} + \left(\theta_{x,\max}\sqrt{\bar{\lambda}_{\Gamma_x}} + \left\|\hat{\mathbf{K}}_x\right\|_2\right)F\right)\end{aligned} \tag{F.90}$$

Insertion of results (F.55), (F.58), (F.88), (F.89) into (F.83) yields

$$\left\|\dot{\mathbf{v}}\right\|_2 \le A_{\hat{e}}\left\|\hat{\mathbf{e}}\right\|_2 + A_u\left\|\tilde{\mathbf{u}}_N\right\|_2 + A_0 \tag{F.91}$$

where

$$
\begin{aligned}
A_{\hat{e}} &= \left(\|\mathbf{K}\|_2 + \sqrt{\bar{\lambda}_{\Gamma_x}}\theta_{x,\max}F_\zeta\right)Z_{\hat{e}} + \left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2\right)E_{\hat{e}} + \sqrt{\frac{\bar{\lambda}_{\Gamma_x}^3}{\underline{\lambda}_{\Gamma_x}}}\bar{\lambda}_{P_E}F^2 + \sqrt{\bar{\lambda}_{\Gamma_x}}\theta_{x,\max}F_\eta Y_{\hat{e}} \\
A_u &= \left(\|\mathbf{K}\|_2 + \sqrt{\bar{\lambda}_{\Gamma_x}}\theta_{x,\max}F_\zeta\right)k_2 \\
A_0 &= \left(\|\mathbf{K}\|_2 + \sqrt{\bar{\lambda}_{\Gamma_x}}\theta_{x,\max}F_\zeta\right)Z_0 + \|\mathbf{A}_0\|_2\,\bar{y}_d + \left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2\right)E_0 + \sigma_x\sqrt{\bar{\lambda}_{\Gamma_x}}\theta_{x,\max}F \\
&\quad + \kappa\sqrt{\frac{\bar{\lambda}_{\Gamma_x}}{\underline{\lambda}_{\Gamma_x}}}\left(\bar{w} + \frac{mLD\sqrt{N}}{2}\right)\|\mathbf{\Gamma}_x\|_F F + \sqrt{\bar{\lambda}_{\Gamma_x}}\theta_{x,\max}F_\eta Y_0
\end{aligned}
\tag{F.92}
$$

Using results (F.47), (F.49), (F.55), (F.58), (F.79), (F.91), assumption H in (F.62) yields

$$
\left\|\dot{\mathbf{w}}^*\right\|_2 \le C_{\hat{e}}\|\hat{\mathbf{e}}\|_2 + C_{\hat{e}^2}\|\hat{\mathbf{e}}\|_2^2 + C_{\hat{e}^3}\|\hat{\mathbf{e}}\|_2^3 + C_{\hat{e}u}\|\hat{\mathbf{e}}\|_2\|\tilde{\mathbf{u}}_N\|_2 + C_u\|\tilde{\mathbf{u}}_N\|_2 + C_0 \tag{F.93}
$$

where

$$
\begin{aligned}
C_{\hat{e}} &= B_0\left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2\right) + A_{\hat{e}} + \left\|\hat{\mathbf{K}}_x\right\|_2\left(F_\zeta Z_{\hat{e}} + F_\eta Y_{\hat{e}}\right) \\
&\quad + B_{\hat{e}}\left(\|\mathbf{K}\|_2\bar{\zeta} + \|\mathbf{A}_0\|_2\bar{y} + \theta_{x,\max}\sqrt{\bar{\lambda}_{\Gamma_x}}F + \left\|\hat{\mathbf{K}}_x\right\|_2 F\right) \\
C_{\hat{e}^2} &= B_{\hat{e}}\left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2\right) + B_{\hat{e}^2}\left(\|\mathbf{K}\|_2\bar{\zeta} + \|\mathbf{A}_0\|_2\bar{y} + \theta_{x,\max}\sqrt{\bar{\lambda}_{\Gamma_x}}F + \left\|\hat{\mathbf{K}}_x\right\|_2 F\right) \\
C_{\hat{e}^3} &= B_{\hat{e}^2}\left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2\right) \\
C_{\hat{e}u} &= B_u\left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2\right) \\
C_u &= B_u\left(\|\mathbf{K}\|_2\bar{\zeta} + \|\mathbf{A}_0\|_2\bar{y} + \theta_{x,\max}\sqrt{\bar{\lambda}_{\Gamma_x}}F + \left\|\hat{\mathbf{K}}_x\right\|_2 F\right) + A_u + \left\|\hat{\mathbf{K}}_x\right\|_2 F_\zeta k_2 \\
C_0 &= B_0\left(\|\mathbf{K}\|_2\bar{\zeta} + \|\mathbf{A}_0\|_2\bar{y} + \theta_{x,\max}\sqrt{\bar{\lambda}_{\Gamma_x}}F + \left\|\hat{\mathbf{K}}_x\right\|_2 F\right) + A_0 + \left\|\hat{\mathbf{K}}_x\right\|_2 F_\zeta Z_0 + \left\|\hat{\mathbf{K}}_x\right\|_2 F_\eta Y_0.
\end{aligned}
\tag{F.94}
$$

*Bound on w**

From linearizing state feedback (4.323), we get

$$
\left\|\mathbf{w}^*\right\|_2 = \left\|\left[\hat{\mathbf{B}}(\boldsymbol{\zeta},\boldsymbol{\eta})\hat{\mathbf{B}}^T(\boldsymbol{\zeta},\boldsymbol{\eta}) + \mathbf{I}\right]^{-1}\right\|_2\left[\|\mathbf{v}\|_2 + \left\|\hat{\mathbf{K}}_x\right\|_2\|\boldsymbol{\varphi}(\boldsymbol{\zeta},\boldsymbol{\eta})\|_2\right]. \tag{F.95}
$$

Using results (F.47), (F.49) and assumption H yields

$$
\left\|\mathbf{w}^*\right\|_2 = \left(\|\mathbf{K}\|_2 + \|\mathbf{C}\|_2\right)\|\hat{\mathbf{e}}\|_2 + \|\mathbf{K}\|_2\bar{\zeta} + \|\mathbf{A}_0\|_2\bar{y} + \left(\theta_{x,\max}\sqrt{\bar{\lambda}_{\Gamma_x}} + \left\|\hat{\mathbf{K}}_x\right\|_2\right)F\ . \tag{F.96}
$$