

Considering Uncertainty in Optimal Robot Control Through High-Order Cost Statistics

José R. Medina and Sandra Hirche

Abstract—As the application of probabilistic models in robotic applications increases, a systematic robot control approach considering the effects of uncertainty becomes indispensable. Inspired by human sensorimotor findings, in this work we study the stochastic optimal control problem with high-order cost statistics in order to synthesize uncertainty-dependent actions in robotic scenarios with multiple uncertainty sources. We present locally optimal risk-sensitive and cost-cumulant solutions for settings with non-linear dynamics, multiple additive uncertainty sources and non-quadratic costs. The influence of each uncertainty source on the cost can be individually parameterized offering additional flexibility in the control design. We further analyze the case in which static uncertain parameters are involved. Simulations of several linear and non-linear settings with non-quadratic costs and an experiment on a real robotic platform validate our approach and illustrate its peculiarities.

Index Terms—Uncertainty in robot control, stochastic optimal control, risk-sensitive control.

I. INTRODUCTION

ONE of the core capabilities of intelligent systems is the appropriate reaction to uncertainties in its environment as recognized in AI since the late 80's [1]. While modern reasoning methods embrace stochastic uncertainties in a sophisticated way, low level robot control still largely lacks suitable approaches. State-of-the art stochastic control methods consider the expected value and ignore higher order statistics. The objective of this article is to fill the gap between probabilistic modeling approaches and robot control by providing a systematic approach to robot control in the presence of stochastic uncertainties beyond first-order statistics. As an example, consider the task of grasping an object with uncertain pose. Its expected value is only a rough estimation of its probabilistic representation. Its variance, though, may reveal more certainty in some degrees of freedom; control design targeting those more aggressively while allowing more variability in others may be key for task success. Similarly, if we consider a navigation task in a cluttered environment where obstacles have different levels of uncertainty, the possibility of collision may vary significantly depending on obstacle pose variances. These simple prototypical examples illustrate the necessity of a systematic and flexible control approach that considers uncertainty in robotic settings.

Due to the ability of humans succeeding on a wide variety of tasks, a reliable source of inspiration for robot control design is human behavior. In recent years neuroscientists have studied

human sensorimotor control, highlighting their capability to overcome and even benefit from the effects of noise and uncertainty [2]. In particular, results in this area show that human motor control can be modeled as an optimal control problem with consideration of noise in the dynamics [3]. These findings motivate many stochastic optimal control approaches for robot control which minimize the *expected value* of a stochastic cost [4], [5]. However, recent results interpret human actions by means of *risk-sensitive* optimal control [6]. Intuitively, a risk-sensitive decision-maker deviates from the expected optimum in the face of uncertainty by considering not only the expected value of the stochastic cost but also its variance and further higher-order statistics (moments or cumulants). Interestingly, human uncertainty-dependent behavior is not limited to the effects of sensorimotor noise. Recent studies demonstrate that human actions are similarly affected by the uncertainty of task-related variables, such as a goal to reach [7] or the internal model of a cooperating partner [8]. These insights motivate us to explore risk-sensitive decision-makers for prototypical robot control problems where not only the robot dynamics but also task- and environment-dependent variables are uncertain. This problem setting naturally arises when robots are deployed in unstructured environments and must adapt to unknown surroundings, thereby estimating uncertain dynamics, goal and obstacle configurations. To the best of our knowledge, optimal control design considering high-order statistics under multiple uncertainty sources is still an open issue with potential application in many robotic scenarios.

In this work we study the stochastic optimal control problem considering high-order cost statistics in order to synthesize uncertainty-dependent actions in robotic scenarios with multiple uncertainty sources. We present locally optimal risk-sensitive and cost-cumulant solutions for problems with non-linear dynamics, multiple additive uncertainty sources, uncertain parameters and non-quadratic costs. The influence of each uncertainty source on the cost can be individually parameterized offering additional flexibility in the control design. Locally optimal solutions are found by iteratively performing a linear quadratic (LQ) approximation around a nominal trajectory, solving the local problem in closed form and updating the trajectory until convergence. Simulations of several linear and non-linear settings with non-quadratic costs and an experiment on a real robotic platform validate our approach and illustrate its peculiarities. Preliminary results of this work are published in the conference paper [9]; this article contains a modified problem setting, additional results, simulations, and experiments.

José R. Medina is with FRANKA EMIKA GmbH, D-80797 Munich, Germany. jose.medina@franka.de

Sandra Hirche is with the Chair of Information-oriented Control, Department of Electrical Engineering and Information Technology, Technische Universität München, D-80290 Munich, Germany. hirche@tum.de

A. Related Work

Previous works explore the synthesis of uncertainty-dependent decisions in robotic settings from an optimality perspective by modifying the performance measure of deterministic settings to account for the uncertainty's variance in an application specific way. For instance, enhanced collision avoidance is achieved including an additive cost term in representing the probability of collision [10]. Similarly, variance-dependent stiffness is synthesized adding prediction variance to the performance measure [11]. The inverse of the variance is also used as a weighting term for measuring distance to desired states [12], [13]. The heterogeneity of these solutions reveals the lack of a systematic approach to synthesize uncertainty-dependent actions. Control theory offers many alternatives for decision-making in the face of uncertainty, such as uncertainty tubes [14] or set-uncertainty to increase robustness [15], [16]. Among those, risk-sensitive control and cost-cumulant control provide the additional possibility of assessing the effect of uncertainty in both a positive (risk-seeking) and a negative (risk-averse) way. In this work, we follow this approach, which relies on a statistical interpretation of the stochastic performance measure, a flexible criterion independent of the application scenario and therefore valid as a systematic approach.

Stochastic optimal control in non-linear and non-quadratic setting is easily trapped under the curse of dimensionality which prevents from realtime implementations. In this context, local linear-quadratic approximations are an effective solution. The iterative Linear Quadratic Gaussian (iLQG) method presented in [17] which is a simplified version of Stochastic Differential Dynamic Programming (SDDP) [18], [19] is a recurrent example. The resulting locally optimal feedback policies have been applied in many robot navigation problems combined with belief roadmaps [20] or adding the state variance as a part of an augmented state [4], [21]. All these approaches consider the expected value of the random cost as a performance measure, neglecting cost variance. Here, we also consider higher order cumulants of the cost in order to capture the influence of uncertainty into robot actions.

The statistics of random costs have been primarily studied in the context of mathematical finance, initially exploring mean and variance solutions [22]. Further than the first two cumulants, the risk-sensitive performance criterion initially proposed in [23] and widely applied in modern portfolio theory, considers a weighted sum of all infinite cumulants of the cost [24] and has closed form solutions in LQ settings. The discrete-time solution was analyzed in depth in [25]. High-order cost statistics have also been studied in the context of other control approaches, such as game theory, entropy control [16] or polynomial chaos expansion [26], [27]. Roboticians have also recently gained interest in risk-sensitive control, applying it to stochastic cost functions learned with Gaussian Processes [28] as well as in approximate inference control [29], [30]. Risk-sensitive control assumes a specific weighting in the infinite summation of the cumulants. When more flexibility in the weightings is needed, cost-cumulant control is a suitable generalization of the risk-sensitive crite-

ron. Main results for continuous-time LQ settings are found in [31] or for the discrete-time case in [32]. Although [32] presents results for quadratic costs, in this work we extend this result to a more general family of LQ systems including a linear term in the cost that enables the application of dynamic programming for locally optimal solutions in non-linear non-quadratic settings. We also consider multiple uncertainty sources and propose a cost functional that enables an individual assessment of the influence of each uncertainty source.

The remainder of this article is organized as follows: Section II formally defines the problem and exposes the proposed locally optimal approach. Section III and Section IV present risk-sensitive and cost-cumulant solutions respectively. The iteration towards the local optimum is described in Section V. Examples with static uncertainty sources are presented in VI. Simulations and an experiment in a real robotic platform are presented in Section VII. Section VIII presents concluding remarks.

Notation: $N \succ 0$ or $N \succeq 0$ denote that matrix N is positive-definite or positive-semidefinite respectively. $f(\theta)^{[r]}$ denotes the r -th derivative of $f(\theta)$ w.r.t θ at point $\theta = 0$, i.e. $f(\theta)^{[r]} = \left. \frac{\partial^r f(\theta)}{\partial \theta^r} \right|_{\theta=0}$. An $r = 0$ index corresponds to the function itself at $\theta = 0$. C_r^j denotes the binomial coefficient $\binom{j}{r}$.

II. PROBLEM SETTING AND APPROACH

Consider a robot with dynamics given by the stochastic differential equation

$$d\mathbf{x}^r(t) = \mathbf{f}^r(\mathbf{x}^r(t), \mathbf{u}'(t))dt + \mathcal{G}^r(t)dB^r(t), \quad (1)$$

where $\mathbf{x}^r \in \mathbb{R}^v$, $\mathbf{u}' \in \mathbb{R}^m$ represent the robot's state and control input respectively, B^r is a v -dimensional standard Brownian motion noise defined in the complete probability space $(\Omega_r, \mathcal{F}_r, \mathbb{P}_r)$, $\mathcal{G}^r \in \mathbb{R}^{v \times v}$ is its diffusion coefficient matrix and \mathbf{x}_0^r is the initial state. Additionally, let \mathbf{x}^g and \mathbf{x}^o be a desired trajectory to follow and the state of an obstacle respectively with stochastic dynamics

$$\begin{aligned} d\mathbf{x}^g(t) &= \mathbf{f}^g(\mathbf{x}^g(t))dt + \mathcal{G}^g(t)dB^g(t) \\ d\mathbf{x}^o(t) &= \mathbf{f}^o(\mathbf{x}^o(t))dt + \mathcal{G}^o(t)dB^o(t), \end{aligned} \quad (2)$$

where B^g and B^o are v -dimensional standard Brownian motion noises defined in complete probability spaces $(\Omega_g, \mathcal{F}_g, \mathbb{P}_g)$ and $(\Omega_o, \mathcal{F}_o, \mathbb{P}_o)$ respectively, \mathcal{G}_g and \mathcal{G}_o their diffusion coefficient matrices and \mathbf{x}_0^g and \mathbf{x}_0^o their respective initial states¹.

Remark 1. Throughout this article we will consider this general form of a non-linear dynamics for the robot, goal, and obstacle in order to keep the derivations general and applicable to the different robotic tasks. Needless to say that any continuous robot dynamics can be expressed in terms

¹The emergence of probabilistic models as an effective tool for acquiring motion models of a desired trajectory [33], [34], obstacles [35] or any other dynamic elements in unstructured environment motivates stochastic dynamics (1) and (2).

of (1). Consider as example an n -link rigid manipulator represented in joint space with $\mathbf{x}^r = [\boldsymbol{\theta}^T \dot{\boldsymbol{\theta}}^T]^T$, $\mathbf{u}' = \boldsymbol{\tau}$ and

$$\mathbf{f}^r(\mathbf{x}^r, \mathbf{u}') = \begin{bmatrix} \dot{\boldsymbol{\theta}} \\ M(\boldsymbol{\theta})^{-1}(\boldsymbol{\tau} - C(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}) - g(\boldsymbol{\theta})) \end{bmatrix},$$

where $\boldsymbol{\theta} \in \mathbb{R}^n$ are the joint angles, $M(\boldsymbol{\theta})$ is the positive definite inertia matrix, $C(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}) \in \mathbb{R}^n$ is the vector of centripetal and Coriolis forces, and $\boldsymbol{\tau} \in \mathbb{R}^n$ are the joint torques. The additive stochastic uncertainties $\mathcal{G}^i(t)dB^i$ with $i \in \{r, g, o\}$ in (1) and (2) can represent unmodelled dynamics or unknown external wrenches. Analogously, this holds for the dynamics represented in task space by applying the appropriate coordinate transformation. An example on a 2-link manipulator and a further example in terms of a mobile robot with car-like dynamics is given in Section VII.

In order to keep a compact and general formulation, let $\boldsymbol{\xi}' \in \mathbb{R}^n$ be a joint state comprising all dynamic elements in the problem with dynamics

$$d\boldsymbol{\xi}'(t) = \mathbf{f}(\boldsymbol{\xi}'(t), \mathbf{u}'(t))dt + \sum_{s=1}^S \mathcal{G}^s(t)dB^s(t), \quad (3)$$

where S is the number of independent Brownian motions, B^s is the corresponding n -dimensional standard Brownian motion noise defined in the s -th probability space $(\Omega_s, \mathcal{F}_s, \mathbb{P}_s)$, \mathcal{G}^s its diffusion coefficient matrix and $\boldsymbol{\xi}'_0$ is the initial state. In our particular scenario $S = 3$ and for later convenience we define the joint state in terms of differences such that

$$\begin{aligned} \boldsymbol{\xi}' &= \begin{bmatrix} \mathbf{x}^g - \mathbf{x}^r \\ \mathbf{x}^o - \mathbf{x}^r \end{bmatrix} & \boldsymbol{\xi}'_0 &= \begin{bmatrix} \mathbf{x}_0^g - \mathbf{x}_0^r \\ \mathbf{x}_0^o - \mathbf{x}_0^r \end{bmatrix} \\ \mathbf{f}(\boldsymbol{\xi}'(t), \mathbf{u}'(t)) &= \begin{bmatrix} \mathbf{f}^g(\mathbf{x}^g(t)) - \mathbf{f}^r(\mathbf{x}^r(t), \mathbf{u}'(t)) \\ \mathbf{f}^o(\mathbf{x}^o(t)) - \mathbf{f}^r(\mathbf{x}^r(t), \mathbf{u}'(t)) \end{bmatrix} \\ \mathcal{G}^1 &= \begin{bmatrix} \mathcal{G}^r & \mathbf{0} \\ \mathbf{0} & \mathcal{G}^r \end{bmatrix} & \mathcal{G}^2 &= \begin{bmatrix} \mathcal{G}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} & \mathcal{G}^3 &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{G}^o \end{bmatrix}. \end{aligned} \quad (4)$$

The control goal is to find the input policy that minimizes cost function

$$J(\boldsymbol{\xi}'(\cdot), \mathbf{u}'(\cdot)) = h_{T_c}(\boldsymbol{\xi}'(T_c)) + \int_{t=0}^{T_c} h(\boldsymbol{\xi}'(t), \mathbf{u}'(t))dt, \quad (5)$$

where T_c is the time horizon, h is the cost rate and h_{T_c} the end term. This performance index is commonly designed penalizing both the distance of the robot to the desired trajectory \mathbf{x}^g and the necessary control efforts \mathbf{u}' while favoring configurations distant to obstacles \mathbf{x}^o .

The optimal control solution is given by the control law $\mathbf{u}'(\cdot)$ that minimizes (5) constrained to dynamics (3). Note that cost (5) is a random variable as it is a function of the random variable $\boldsymbol{\xi}'$ which is affected by random variables dB^s with $s \in \{1, \dots, S\}$. To enable the optimization, an assessment of (5) in terms of a deterministic performance measure is necessary. A valid approach consists of evaluating a statistical measure of (5) commonly limited to the expected value, i.e. $\mathbb{E}_{\mathbb{P}}[J]$, and where the expectation is defined in the product probability space of all uncertainty sources, i.e. the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ given by

$$\Omega = \Omega_1 \times \Omega_2 \times \dots \times \Omega_S, \quad \mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \dots \times \mathcal{F}_S, \quad (6)$$

where \times denotes the Cartesian product and \mathbb{P} is the joint measure defined on measurable space (Ω, \mathcal{F}) . However, this formulation has two drawbacks. First, all statistics except the expected value are neglected. The influence of uncertainty on performance representing for instance the required precision while following a desired trajectory or the probability of collision are typically captured only in high-order statistics of (5). Considering only its expectation ignores this relevant information in the optimization. Second, the evaluation of the statistics in probability space (6) considers all random variables *jointly*. This limits the way cost variability influences decisions. It might be desirable that the desired trajectory uncertainty *decreases* the overall cost to reflect less tracking precision or that the robot model uncertainty decreases the cost to foster exploratory behaviors while, at the same time, obstacle uncertainty *increases* the cost as the probability of collision rises. Considering both contradictory assessments of cost variability (decreasing and increasing) at the same time in a systematic way is not possible in probability space (6). Concerning the first issue, cost-cumulant control [36] generalizes optimal state feedback solutions to performance measures given by an arbitrary finite sum of cost cumulants

$$\kappa_{\mathbb{P}}^{(1)}(J) + \sum_{r=2}^K \gamma_r \kappa_{\mathbb{P}}^{(r)}(J), \quad (7)$$

where $\kappa_{\mathbb{P}}^{(r)}$ denotes the r -th cumulant calculated in probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\gamma_r \in \mathbb{R} \forall r \in 2, \dots, K$. The first moment and cumulant are equivalent and defined by the *expected cost*. The second cumulant is the cost variance and the third and fourth cumulants are related to the skewness and kurtosis of the distribution respectively. We will informally refer to these high-order terms as *cost variability* indistinctly. Note that high-order statistical terms are also denoted *risk measures* in modern portfolio theory [31]. Cumulants are derived by means of the cumulant generating function

$$\Psi_{\mathbb{P}}(\theta J) = \log \mathbb{E}_{\mathbb{P}}[\exp\{\theta J\}], \quad (8)$$

with $\theta \in \mathbb{R}$. Its power series expansion is given by

$$\Psi_{\mathbb{P}}(\theta J) = \sum_{r=1}^{\infty} \frac{\theta^r}{r!} \kappa_{\mathbb{P}}^{(r)}(J), \quad (9)$$

where

$$\kappa_{\mathbb{P}}^{(r)}(J) = \left. \frac{\partial^r \Psi_{\mathbb{P}}(\theta)}{\partial \theta^r} \right|_{\theta=0}, \quad (10)$$

providing a compact way to calculate the desired cumulants. Regarding the second issue, instead of considering (7), in this work we evaluate the cost variability produced marginally by the s -th uncertainty source by solving

$$\kappa_{\mathbb{P}_s}^{(1)}(\mathbb{E}_{\mathbb{P}_s}[J]) + \sum_{r=2}^K \gamma_{r,s} \kappa_{\mathbb{P}_s}^{(r)}(\mathbb{E}_{\mathbb{P}_s}[J]),$$

where $\mathbb{E}_{\mathbb{P}_s}[J]$ is the expectation defined in the product probability space of all uncertainty sources except for the s -th one. This operator enables the evaluation of cost variability produced marginally by the s -th uncertainty source by performing a neutral assessment of all other uncertainties. We

then consider the marginal influence of all uncertainty sources by solving problem

$$\begin{aligned} \min_{\mathbf{u}'(\cdot)} \bar{\Xi} = \min_{\mathbf{u}'(\cdot)} \frac{1}{S} \sum_{s=1}^S \left(\kappa_{P_s}^{(1)}(E_{P_s}[J]) + \sum_{r=2}^K \gamma_{r,s} \kappa_{P_s}^{(r)}(E_{P_s}[J]) \right) \\ \text{s. t.} \quad (3), (5) \end{aligned} \quad (11)$$

The intuition behind Problem (11) is simple: we aim for the controls that minimize not just the expected cost but also a *weighted average* of marginal cost variabilities. In contrast to (7), a solution to this problem may adopt different or even opposite decisions depending on which uncertainty source is responsible for cost variability. The assessment of the influence of each uncertainty source is determined by their respective weighting factors. If $\gamma_{r,s} > 0$, cost variability increases the overall cost and the optimizer adopts a *risk-averse* attitude; the respective risk measures imply additional costs and more control efforts. In contrast, for $\gamma_{r,s} < 0$ the optimizer adopts a *risk-seeking* attitude; risk measures are interpreted as a cost discounting quantity taking part of the necessary control effort. In case that all $\gamma_{r,s} = 0$, the expected value is recovered, i.e. it neglects any risk measure adopting a *risk-neutral* attitude. In our particular robotic scenario, by means of problem (11), the assessment of obstacle uncertainty becomes independent of the evaluation of the desired trajectory uncertainty. This way the optimizer can adopt at the same time a risk-seeking attitude towards the desired trajectory uncertainty and a risk-averse attitude towards obstacle uncertainty by selecting appropriate weighting factors. As a result, both tracking precision and collision probability are properly considered into robot actions.

Risk-sensitive control [23] is a special case of cost-cumulant control, which has been successfully employed for the synthesis of uncertainty-dependent actions in robotic scenarios [37]. It is a particular case of cost-cumulant minimization [16] where weighting factors are given by the McLaurin coefficients of the power series (9). In fact, we could similarly formulate problem (11) for $K \rightarrow \infty$ as

$$\begin{aligned} \min_{\mathbf{u}'(\cdot)} \bar{\Psi} = \min_{\mathbf{u}'(\cdot)} \frac{1}{S} \sum_{s=1}^S \theta_s^{-1} \Psi_{P_s}(\theta_s E_{P_s}[J]) \\ = \min_{\mathbf{u}'(\cdot)} \frac{1}{S} \sum_{s=1}^S \theta_s^{-1} \log E_{P_s}[\exp\{\theta_s E_{P_s}[J]\}] \\ = \min_{\mathbf{u}'(\cdot)} \frac{1}{S} \sum_{s=1}^S \left(\kappa_{P_s}^{(1)}(E_{P_s}[J]) + \sum_{r=2}^{\infty} \frac{\theta_s^{r-1}}{r!} \kappa_{P_s}^{(r)}(E_{P_s}[J]) \right), \\ \text{s. t.} \quad (3), (5) \end{aligned} \quad (12)$$

by fixing the the high-order statistics weighting factors to the above-mentioned McLaurin coefficients divided by θ_s . The cost functional to minimize reduces therefore to an average of standard risk-sensitive functionals where θ_s , denoted in the literature *risk-sensitivity*, determines the assessment of the high-order statistics of the s -th uncertainty source. Due to its relevance we will explore both the risk-sensitive (12) and the K -cost-cumulant control (11) problems.

A. Locally optimal control for non-linear dynamics and non-quadratic cost

A solution to the non-linear stochastic optimal control problems (12) or (11) is in general not attainable. As an alternative, we aim for a local optimum by means of an iterative algorithm. By linearizing the dynamics and quadratically approximating the cost around a discretized nominal trajectory $(\bar{\xi}'_{0\dots T}, \bar{\mathbf{u}}'_{0\dots T-1})$, a discrete-time LQ approximation of state and control *deviations*, i.e. $\xi_k = \xi'_k - \bar{\xi}'_k$ and $\mathbf{u}_k = \mathbf{u}'_k - \bar{\mathbf{u}}'_k$ is obtained. Its solution is a gradient towards the local optimum, found by iteratively updating the nominal trajectory and repeating the whole process until convergence.

The local deviations LQ problem is defined as follows. Time is discretized in T steps with sample time $\Delta = T_c/T$. Dynamics (3) linearized at time step k around $(\bar{\xi}'_k, \bar{\mathbf{u}}'_k)$ is given by

$$\xi_{k+1} = A_k \xi_k + B_k \mathbf{u}_k + \sum_{s=1}^S \epsilon_k^s, \quad (13)$$

where $A_k \in \mathbb{R}^{n \times n}$, $B_k \in \mathbb{R}^{n \times m}$ are real matrices, $\epsilon_k^s \in \mathbb{R}^n$ is an independent identically distributed Gaussian random variable such that $\epsilon_k^s \sim \mathcal{N}(\mathbf{0}, \Sigma_k^s)$, $\xi_0 = \mathbf{0}$ and

$$A_k = I_n + \Delta \frac{\partial \mathbf{f}}{\partial \xi_k}, \quad B_k = \Delta \frac{\partial \mathbf{f}}{\partial \mathbf{u}'_k}, \quad \Sigma_k^s = \Delta \mathcal{G}^s(k\Delta) \mathcal{G}^s(k\Delta)^\top.$$

Cost functional (5) results in the quadratic approximation

$$\begin{aligned} J(\xi_0, \mathbf{u}_{0\dots T-1}) = \frac{1}{2} \xi_T^\top Q_T \xi_T + \xi_T^\top \mathbf{q}_T + \\ \sum_{k=0}^{T-1} \left[\frac{1}{2} \xi_k^\top Q_k \xi_k + \xi_k^\top \mathbf{q}_k + q_k + \frac{1}{2} \mathbf{u}_k^\top R_k \mathbf{u}_k + \mathbf{u}_k^\top \mathbf{r}_k \right], \end{aligned} \quad (14)$$

where $Q_k \in \mathbb{R}^{n \times n}$, $R_k \in \mathbb{R}^{m \times m}$, $\mathbf{q}_k \in \mathbb{R}^n$ and $\mathbf{r}_k \in \mathbb{R}^m$ and $q_k \in \mathbb{R}$ are computed as

$$\begin{aligned} q_k = \Delta h \quad \mathbf{q}_k = \Delta \frac{\partial h}{\partial \xi_k} \quad Q_k = \Delta \frac{\partial^2 h}{\partial (\xi_k')^2} \\ \mathbf{q}_T = h_{T_c}(\xi'_T) \quad \mathbf{q}_T = \frac{\partial h_{T_c}}{\partial \xi'_T} \quad Q_T = \frac{\partial^2 h_{T_c}}{\partial (\xi'_T)^2} \\ \mathbf{r}_k = \Delta \frac{\partial h}{\partial \mathbf{u}'_k} \quad R_k = \Delta \frac{\partial^2 h}{\partial (\mathbf{u}'_k)^2}, \end{aligned}$$

where $Q_k \succeq 0$ and $R_k \succ 0$.

Note that, from (13), ξ_k is normally distributed and hence (14) is a generalized non-central chi-squared distribution due to terms $\frac{1}{2} \xi_k^\top Q_k \xi_k$ and $\xi_k^\top \mathbf{q}_k$. Thus, it remains an expressive approximation to the original cost where cumulants higher than the expected cost are relevant.

For this LQG setting we restrict our solutions to linear policies with feedforward and feedback terms,

$$\mathbf{u}_k = \mathbf{l}_k + L_k \mathbf{x}_k, \quad (15)$$

where \mathbf{l}_k is an affine input and $L_k \in \mathbb{R}^{m \times n}$ is the feedback matrix.

The local deviations problems for the risk-sensitive (12) and cost-cumulant problems (11) are

$$\begin{aligned} \text{(a)} \quad & \min_{\mathbf{u}_{1 \dots T-1}} \overline{\Psi} & \text{(b)} \quad & \min_{\mathbf{u}_{1 \dots T-1}} \Xi \\ \text{s. t. (13), (14)} & & \text{s. t. (13), (14), (15)} & \end{aligned} \quad (16)$$

respectively. Section III and Section IV solve problems (16a) and (16b) respectively. The iteration, which drives the nominal trajectory towards the local optimum, is detailed out in Section V.

Remark 2. For the risk-sensitive problem (16a) the linear control (15) corresponds also to the optimal solution over all possible control policies. This linear solution is also optimal for the risk-neutral and mean-variance case, but does not hold for any arbitrary weighted sum of cost-cumulants. In the cost-cumulant case, for the sake of computational complexity, we restrict our solutions to linear policies, adding equality constraints (15) to problem (16b).

III. RISK-SENSITIVE SOLUTION

In this section we study the solution to the discrete-time marginal risk-sensitive LQ problems (16a) by means of dynamic programming applying Bellman's optimality principle. Without loss of generality, we first consider a single uncertainty source. Dynamics simplify to

$$\boldsymbol{\xi}_{k+1} = A_k \boldsymbol{\xi}_k + B_k \mathbf{u}_k + \boldsymbol{\epsilon}_k, \quad (17)$$

where $\boldsymbol{\epsilon}_k \in \mathbb{R}^n$ is an independent Gaussian random variable such that $\boldsymbol{\epsilon}_k \sim \mathcal{N}(\mathbf{0}, \Sigma_k)$ and the cost functional to minimize reduces to $\Psi_P(\theta J(\boldsymbol{\xi}_{0 \dots T}, \mathbf{u}_{0 \dots T-1}))$. Given a control law $\mathbf{u}_{k \dots T-1}$ in the form (15) and having observed $\boldsymbol{\xi}_k$, the *cost-to-go* at time step k is computed by means of the backwards recursion

$$\begin{aligned} \theta^{-1} \Psi_P(\theta J(\boldsymbol{\xi}_k, \mathbf{u}_{k \dots T-1})) &= J(\boldsymbol{\xi}_k, \mathbf{u}_k) \\ &+ \theta^{-1} \Psi_P(\theta J(\boldsymbol{\xi}_{k+1}, \mathbf{u}_{k+1 \dots T-1})). \end{aligned} \quad (18)$$

Note that the overall cost given control policy $\mathbf{u}_{0 \dots T-1}$ corresponds to the cost-to-go at $k = 0$, i.e. $\Psi(J(\boldsymbol{\xi}_0, \mathbf{u}_{0 \dots T-1}))$.

The following lemma computes the risk-sensitive cost-to-go recursively yielding a quadratic form in $\boldsymbol{\xi}_k$.

Lemma 1. *If $(\Sigma_i^{-1} - \theta W_{i+1}) \succ 0$ for $i = k \dots T - 1$, the analytic solution to (18) with cost (14) and dynamics (17) is given by the backwards recursion*

$$\theta^{-1} \Psi_P(\theta J(\boldsymbol{\xi}_k, \mathbf{u}_{k \dots T-1})) = \frac{1}{2} \boldsymbol{\xi}_k^\top W_k \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\top \mathbf{w}_k + w_k, \quad (19)$$

where

$$W_k = Q_k + A_k^\top \widetilde{W}_{k+1} A_k + L_k^\top H_k L_k + G_k^\top L_k + L_k^\top G_k \quad (20)$$

$$\mathbf{w}_k = \mathbf{q}_k + A_k^\top \widetilde{\mathbf{w}}_{k+1} + L_k^\top H_k \mathbf{l}_k + L_k^\top \mathbf{g}_k + G_k^\top \mathbf{l}_k \quad (21)$$

$$w_k = q_k + \widetilde{w}_{k+1} + \frac{1}{2} \mathbf{l}_k^\top H \mathbf{l}_k + \mathbf{l}_k^\top \mathbf{g}_k \quad (22)$$

$$H_k = R_k + B_k^\top \widetilde{W}_{k+1} B_k \quad G_k = B_k^\top \widetilde{W}_{k+1} A_k$$

$$\mathbf{g}_k = \mathbf{r}_k + B_k^\top \widetilde{\mathbf{w}}_{k+1}$$

with $W_T = Q_T$, $\mathbf{w}_T = \mathbf{q}_T$ and $w_T = q_T$ and

$$\widetilde{W}_{k+1} = (I - \theta W_{k+1} \Sigma_k)^{-1} W_{k+1} \quad (23)$$

$$\widetilde{\mathbf{w}}_{k+1} = (I - \theta W_{k+1} \Sigma_k)^{-1} \mathbf{w}_{k+1} \quad (24)$$

$$\widetilde{w}_{k+1} = w_{k+1} + \theta \mathbf{w}_{k+1}^\top \Sigma_k \widetilde{\mathbf{w}}_{k+1} - \frac{1}{2} \theta^{-1} \log F_k \quad (25)$$

with $F_k = |I - \theta W_{k+1} \Sigma_k|$ and $F_T = 1$.

Proof. See Appendix. ■

The influence of θ on the cost becomes apparent observing (23), (24) and (25). A positive θ results in higher quadratic cost coefficients as it diminishes the eigenvalues of the inverted factor $(I - \theta W_{k+1} \Sigma_k)^{-1}$, thereby adopting a risk-averse attitude. In contrast, a negative θ increases the inverted factor and therefore decreases the overall cost, adopting a risk-seeking attitude. In fact, extreme risk-seekingness, i.e. $\theta \ll 0$, yields $\widetilde{W}_{k+1} \rightarrow 0$ and $\widetilde{\mathbf{w}}_{k+1} \rightarrow \mathbf{0}$. This degenerate case is described as *euphoria* in [25], as cost variability nullifies the overall cost.

The cost-to-go for problem (16a), i.e. the remaining average cost of all marginal problems at time step k is given by the following corollary.

Corollary 1. *The cost-to-go for problem (16a) is given by*

$$\overline{\Psi}(\boldsymbol{\xi}_k, \mathbf{u}_{k \dots T-1}) = \frac{1}{2} \boldsymbol{\xi}_k^\top \overline{W}_k \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^\top \overline{\mathbf{w}}_k + \overline{w}_k, \quad (26)$$

where

$$\begin{aligned} \overline{W}_k &= \frac{1}{S} \sum_{s=1}^S W_k^s & \overline{\mathbf{w}}_k &= \frac{1}{S} \sum_{s=1}^S \mathbf{w}_k^s \\ \overline{w}_k &= \frac{1}{S} \sum_{s=1}^S w_k^s + \sum_{t=k}^{T-1} \sum_{g=1, g \neq s}^S \text{Tr}\{\Sigma_t^g W_{t+1}^s\}, \end{aligned} \quad (27)$$

and where W_k^s , \mathbf{w}_k^s and w_k^s are given by (23), (24) and (25) resulting from applying the backwards recursion from Lemma 1 in the s -th probability space, i.e. with $\Sigma_k = \Sigma_k^s$ and $\theta = \theta_s$.

Note that the trace term in (27) follows from the evaluation of operator $\mathbb{E}_P^s[J]$. The solution to problem (16a) is computed applying Bellman's optimal equation, which minimizes (26) w.r.t L_k and \mathbf{l}_k at each step of the backwards recursion from Corollary 1, yielding

$$\begin{aligned} L_k &= -(R_k + B_k^\top \widetilde{W}_{k+1} B_k)^{-1} (B_k^\top \widetilde{W}_{k+1} A_k) \\ \mathbf{l}_k &= -(R_k + B_k^\top \widetilde{W}_{k+1} B_k)^{-1} (B_k^\top \widetilde{\mathbf{w}}_{k+1} + \mathbf{r}_k), \end{aligned} \quad (28)$$

where

$$\widetilde{W}_k = \frac{1}{S} \sum_{s=1}^S \widetilde{W}_k^s \quad \widetilde{\mathbf{w}}_k = \frac{1}{S} \sum_{s=1}^S \widetilde{\mathbf{w}}_k^s,$$

and where \widetilde{W}_k^s and $\widetilde{\mathbf{w}}_k^s$ are given by (23) and (24) computed in the s -th probability space.

Remark 3. Note that the computational effort for solving the risk-sensitive control problem - even for the case of multiple uncertainties - is comparably low as i) the control law is linear, and ii) for finding the optimal feedforward and feedback terms only basic algebraic operations are involved. The approach

is therefore also well-suited for robotic applications with high sampling rates and on low-cost computational platforms (assuming that a linear approximation to the system dynamics is sufficient for the purpose of control design, e.g. if feedback linearization is employed.)

IV. COST-CUMULANT SOLUTION

In this section we address the discrete-time cost-cumulant control problem (16b). In contrast to risk-sensitive control, cost-cumulant control allows more flexible decision-makers as the number of cumulants and their respective weighting factors are design parameters.

We consider first a single probability space and dynamics (17). From Lemma 1, the cumulant generating function of the cost-to-go at time step k is given by

$$\Psi_{\mathbb{P}}(\theta J(\boldsymbol{\xi}_k, \mathbf{u}_{k \dots T-1})) = \theta \left(\frac{1}{2} \boldsymbol{\xi}_k^{\top} W_k \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^{\top} \mathbf{w}_k + w_k \right). \quad (29)$$

The r -th cumulant is calculated by means of expression (10). As an illustrative example, the first cumulant is given by

$$\begin{aligned} \frac{\partial \Psi_{\mathbb{P}}(\theta J(\boldsymbol{\xi}_k, \mathbf{u}_{k \dots T-1}))}{\partial \theta} \Big|_{\theta=0} &= \frac{1}{2} \boldsymbol{\xi}_k^{\top} W_k^{[0]} \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^{\top} \mathbf{w}_k^{[0]} + w_k^{[0]} \\ &+ \theta \left(\frac{1}{2} \boldsymbol{\xi}_k^{\top} W_k^{[1]} \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^{\top} \mathbf{w}_k^{[1]} + w_k^{[1]} \right) \Big|_{\theta=0}, \end{aligned}$$

that, after evaluating $\theta = 0$ yields $\mathbb{E}_{\mathbb{P}}[J]$, the cost-to-go of a standard LQ Gaussian problem. Similarly, the second cumulant, i.e. the cost variance $\text{Var}_{\mathbb{P}}[J]$, is given by

$$\begin{aligned} \frac{\partial^2 \Psi_{\mathbb{P}}(\theta J(\boldsymbol{\xi}_k, \mathbf{u}_{k \dots T-1}))}{\partial \theta^2} \Big|_{\theta=0} &= 2 \left(\frac{1}{2} \boldsymbol{\xi}_k^{\top} W_k^{[1]} \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^{\top} \mathbf{w}_k^{[1]} + w_k^{[1]} \right) \\ &+ \theta \left(\frac{1}{2} \boldsymbol{\xi}_k^{\top} W_k^{[2]} \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^{\top} \mathbf{w}_k^{[2]} + w_k^{[2]} \right) \Big|_{\theta=0}. \end{aligned} \quad \Xi(\boldsymbol{\xi}_k) = \frac{1}{2} \boldsymbol{\xi}_k^{\top} \widehat{W}_k \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^{\top} \widehat{\mathbf{w}}_k + \widehat{w}_k, \quad (33)$$

The expressions for the d -th cumulant are calculated recursively, as a function of lower order cumulants, yielding the following lemma [32].

Lemma 2. *The d -th cost cumulant of the cost-to-go at sample time k of random cost (14) constrained to dynamics (17) is given by*

$$\begin{aligned} \kappa_{\mathbb{P}}^{(d)}(J(\boldsymbol{\xi}_k, \mathbf{u}_{k \dots T-1})) &= \\ &d \left(\frac{1}{2} \boldsymbol{\xi}_k^{\top} W_k^{[d-1]} \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^{\top} \mathbf{w}_k^{[d-1]} + w_k^{[d-1]} \right) \end{aligned}$$

where for $d = 1$

$$\begin{aligned} W_k^{[0]} &= Q_k + A_k^{\top} \widehat{W}_{k+1}^{[0]} A_k + L_k^{\top} H_k^{[0]} L_k + G_k^{[0]\top} L_k + L_k^{\top} G_k^{[0]} \\ \mathbf{w}_k^{[0]} &= \mathbf{q}_k + A_k^{\top} \widehat{\mathbf{w}}_{k+1}^{[0]} + L_k^{\top} H_k^{[0]} \mathbf{l}_k + L_k^{\top} \mathbf{g}_k^{[0]} + G_k^{[0]\top} \mathbf{l}_k \\ w_k^{[0]} &= q_k + \widehat{w}_{k+1}^{[0]} + \frac{1}{2} \mathbf{l}_k^{\top} H_k^{[0]} \mathbf{l}_k + \mathbf{l}_k^{\top} \mathbf{g}_k^{[0]} \\ H_k^{[0]} &= R_k + B_k^{\top} \widehat{W}_{k+1}^{[0]} B_k \quad G_k^{[0]} = B_k^{\top} \widehat{W}_{k+1}^{[0]} A_k \\ \mathbf{g}_k^{[0]} &= \mathbf{r}_k + B_k^{\top} \widehat{\mathbf{w}}_{k+1}^{[0]} \end{aligned}$$

with $W_T^{[0]} = Q_T$, $\mathbf{w}_T^{[0]} = \mathbf{q}_T$ and $w_T^{[0]} = q_T$ and for $d > 1$

$$\begin{aligned} W_k^{[r]} &= A_k^{\top} \widehat{W}_{k+1}^{[r]} A_k + L_k^{\top} H_k^{[r]} L_k + G_k^{[r]\top} L_k + L_k^{\top} G_k^{[r]} \\ \mathbf{w}_k^{[r]} &= A_k^{\top} \widehat{\mathbf{w}}_{k+1}^{[r]} + L_k^{\top} H_k^{[r]} \mathbf{l}_k + L_k^{\top} \mathbf{g}_k^{[r]} + G_k^{[r]\top} \mathbf{l}_k \\ w_k^{[r]} &= \widehat{w}_{k+1}^{[r]} + \frac{1}{2} \mathbf{l}_k^{\top} H_k^{[r]} \mathbf{l}_k + \mathbf{l}_k^{\top} \mathbf{g}_k^{[r]} \\ H_k^{[r]} &= B_k^{\top} \widehat{W}_{k+1}^{[r]} B_k \quad G_k^{[r]} = B_k^{\top} \widehat{W}_{k+1}^{[r]} A_k \\ \mathbf{g}_k^{[r]} &= B_k^{\top} \widehat{\mathbf{w}}_{k+1}^{[r]} \end{aligned}$$

with $W_N^{[r]} = 0^{n \times n}$, $\mathbf{w}_N^{[r]} = \mathbf{0}$ and $w_N^{[r]} = 0$ and for both cases

$$\widehat{W}_{k+1}^{[r]} = W_{k+1}^{[r]} + r \sum_{j=0}^{r-1} C_{r-1}^j W_{k+1}^{[j]} \Sigma_k \widehat{W}_{k+1}^{[r-1-j]} \quad (30)$$

$$\widehat{\mathbf{w}}_{k+1}^{[r]} = \mathbf{w}_{k+1}^{[r]} + r \sum_{j=0}^{r-1} C_{r-1}^j W_{k+1}^{[j]} \Sigma_k \widehat{\mathbf{w}}_{k+1}^{[r-1-j]} \quad (31)$$

$$\begin{aligned} \widehat{w}_{k+1}^{[r]} &= w_{k+1}^{[r]} + r \sum_{j=0}^{r-1} C_{r-1}^j \mathbf{w}_{k+1}^{[j]} \Sigma_k \widehat{w}_{k+1}^{[r-1-j]} \\ &+ \sum_{t=k}^{T-1} \text{Tr}(P_t^{[r-1]}) \end{aligned} \quad (32)$$

with

$$P_t^{[r]} = (r+1) W_{t+1}^{[r]} \Sigma_t + r \sum_{j=1}^{r-1} C_{r-1}^j W_{t+1}^{[j]} \Sigma_t P_t^{[r-1-j]}.$$

Proof. See Appendix. ■

The cost-to-go for problem (16b) at time step k follows immediately and is given in the following corollary.

Corollary 2. *The cost-to-go of the cost-cumulant control problem (16b) is given by*

$$\Xi(\boldsymbol{\xi}_k) = \frac{1}{2} \boldsymbol{\xi}_k^{\top} \widehat{W}_k \boldsymbol{\xi}_k + \boldsymbol{\xi}_k^{\top} \widehat{\mathbf{w}}_k + \widehat{w}_k, \quad (33)$$

where

$$\begin{aligned} \widehat{W}_k &= \frac{1}{S} \sum_{s=1}^S \sum_{r=1}^K r \gamma_{r,s} W_k^{s[r-1]} \quad \widehat{\mathbf{w}}_k = \frac{1}{S} \sum_{s=1}^S \sum_{r=1}^K r \gamma_{r,s} \mathbf{w}_k^{s[r-1]} \\ \widehat{w}_k &= \frac{1}{S} \sum_{s=1}^S \sum_{r=1}^K r \gamma_{r,s} \left(w_k^{s[r-1]} + \sum_{t=k}^{T-1} \sum_{g=1, g \neq s}^S \text{Tr}\{\Sigma_t^g W_{t+1}^s\} \right) \end{aligned}$$

where $W_k^{s[r]}$ and $\mathbf{w}_k^{s[r]}$ and $w_k^{s[r]}$ are the quadratic coefficients of the r -th cumulant resulting from applying the backwards recursion from Lemma 2 in the s -th probability space constrained to the s -th marginal dynamics and $\gamma_{1,s} = 1$.

The solution to problem (16b) follows from the minimization of (33) w.r.t L_k and \mathbf{l}_k at each step of the backwards recursion of Corollary 1 yielding

$$\begin{aligned} L_k &= - (R_k + B_k^{\top} \widehat{W}_{k+1} B_k)^{-1} (B_k^{\top} \widehat{W}_{k+1} A_k) \\ \mathbf{l}_k &= - (R_k + B_k^{\top} \widehat{W}_{k+1} B_k)^{-1} (B_k^{\top} \widehat{\mathbf{w}}_{k+1} + \mathbf{r}_k), \end{aligned} \quad (34)$$

where

$$\widehat{W}_k = \frac{1}{S} \sum_{s=1}^S \sum_{r=1}^K \gamma_{r,s} \widehat{W}_k^{s[r-1]} \quad \widehat{\mathbf{w}}_k = \frac{1}{S} \sum_{s=1}^S \sum_{r=1}^K \gamma_{r,s} \widehat{\mathbf{w}}_k^{s[r-1]}$$

where $\widetilde{W}_{k+1}^{s[r-1]}$ and $\widetilde{w}_{k+1}^{s[r-1]}$ result from the computation of the r -th cumulant in the s -th probability space in (33).

As cost cumulants in LQ settings are always positive [31], the effect of weightings $\gamma_{r,s}$ on the resulting cost (33) is evident: risk-aversion is achieved by selecting $\gamma_{r,s} > 0$ and thereby increasing the resulting quadratic coefficients. Selecting $\gamma_{r,s} < 0$ has the opposite effect yielding a risk-seeking evaluation, i.e. 'less' cost than the expected value. Note that the existence of solutions (34) is only warranted if the Hessian $(R_k + B_k^T \widetilde{W}_{k+1} B_k) \succ 0$. Although this constraint imposes no limits in terms of risk aversion, a negative $\gamma_{r,s}$ must fulfill this condition.

Remark 4. As for the risk-sensitive case, also the computational effort for solving the cost-cumulant control problem even for the case of multiple uncertainties is low. Also this approach is well-suited for robotic applications with high sampling rates and on low-cost computational platforms.

V. ITERATION TO A LOCALLY OPTIMAL SOLUTION

A numerical approximation of problem (12) or (11) that computes locally optimal solutions in non-linear and non-quadratic problems requires a procedure that iteratively approximates and updates a nominal trajectory. The main iteration of our approach follows the iLQG algorithm [38]. In this section, we summarize it pointing out the subtle changes that arise due to the different problem setting. The resulting algorithm is either an iterative Linear Exponential Quadratic Regulator (iLEQR) for the risk-sensitive case or an iterative K -Cost Cumulant Regulator (iKCCR) for the cost-cumulant optimization.

Given an initial state ξ'_0 , the algorithm iterates around the nominal control trajectory $\bar{\mathbf{u}}'_{0\dots T-1}$ by calculating the optimal control deviations that improve performance. This initial trajectory can be first obtained from a planning algorithm, for instance [39] or, if no suitable initialization is available, initially set to $\mathbf{0}$. At the i -th iteration, the locally optimal solution is denoted $\mathbf{u}'_{0\dots T-1}^{(i)} = \bar{\mathbf{u}}'_{0\dots T-1} + L_{0\dots T-1}^{(i)} \xi'_{0\dots T-1}$ and is updated to obtain the next $\mathbf{u}'_{0\dots T-1}^{(i+1)}$ following the next steps:

- 1) The corresponding state trajectory $\bar{\xi}'_{0\dots T}$ is computed simulating the discretized dynamics, for instance by Euler integration, i.e. $\bar{\xi}'_{k+1} = \bar{\xi}'_k + \Delta f(\bar{\xi}'_k, \bar{\mathbf{u}}'_{k+1})$.
- 2) The dynamics are linearized and the costs quadratically approximated around $(\bar{\xi}'_{0\dots T}, \bar{\mathbf{u}}'_{0\dots T-1})$, obtaining an LQ approximation of state and control deviations, $\xi_k = \xi'_k - \bar{\xi}'_k$ and $\mathbf{u}_k = \mathbf{u}'_k - \bar{\mathbf{u}}'_k$ as explained in Section II-A.
- 3) Depending on the desired optimization criteria, the optimal deviations law $\mathbf{u}_k = \mathbf{l}_k + L_k \xi_k$ is computed by means of either the risk-sensitive (28) or cost-cumulant recursions (34).
- 4) As feedforward deviations $\mathbf{l}_{0\dots T-1}$ only hold in the close vicinity of the current nominal trajectory, a line search algorithm aims for an adapted step that yields a policy

improvement. Starting with $\alpha = 1$, a potential policy update is given by

$$\mathbf{u}'_k^{(i+1)}(\alpha) = \bar{\mathbf{u}}'_k^{(i)} + \alpha \mathbf{l}_k + L_k \xi_k \quad (35)$$

where ξ_k follows from simulating the linearized system $\xi_{k+1} = A_k \xi_k + B_k(\alpha \mathbf{l}_k + L_k \xi_k)$ with initially $\xi_k = \mathbf{0}$. If the performance of (35) improves, the new policy is accepted, i.e. $\bar{\mathbf{u}}'_{0\dots T-1}^{(i+1)} = \bar{\mathbf{u}}'_{0\dots T-1}^{(i)} + \alpha \mathbf{l}_k$ and $L_k^{(i+1)} = L_k$. If not, the line search parameter is halved, i.e. $\alpha = \frac{1}{2}\alpha$ and the corresponding performance for $\mathbf{u}'_{0\dots T-1}^{(i+1)}(\alpha)$ is evaluated again until improvement. If all $\mathbf{l}_{0\dots T-1}$ are close to $\mathbf{0}$, the algorithm ends. Otherwise, the next iteration starts going back to step 1) with $\mathbf{u}'_{0\dots T-1}^{(i+1)} = \mathbf{u}'_{0\dots T-1}^{(i+1)}(\alpha)$. This algorithm ensures convergence to a locally optimal control policy [40].

The performance evaluation of (35) in step 4) is not straightforward. In fact, an analytical expression of (11) or (12) for an arbitrary J is rarely available. As an alternative, an LQ approximation of the cost around (35) is obtained as in step 2) and its expected performance is computed by means of lemma 1 or 2 respectively. This approximation is effective when the nominal trajectory is close to the local optimum. However, if this is not the case, the feedback matrix L_k may lead to inaccurate results as it is based on local second order information. To avoid this issue, a two-step optimization is applied. First, a solution close to the local optimum is found considering only the feedforward term, i.e. $L_k = 0$. Once the feedforward trajectory converges, the quadratic feedback problem is solved. This second solution converges in few iterations, as the local optimum is already in the vicinity of the nominal trajectory.

Note that the application of this algorithm can be also recomputed after every time step in a non-linear Model Predictive Control manner, yielding a closed-loop control. With the appropriate optimizations and exploiting the problem structure, the solution to this main iteration can be computed in close to real-time computation times [41].

VI. UNCERTAIN STATIC GOAL AND OBSTACLE

Although the problem setting from Section II considers only uncertainty in dynamic settings, many robotics scenarios comprise uncertain parameters in its performance measure representing for example a *static* obstacle/goal or piecewise constant desired trajectory. These problems arise due to sensor noise and also in many programming by demonstration settings [42] when task models are obtained as time series, for example [12] or [43]. The analysis of cost statistics in this case reduces to a static problem that can be solved preceding the optimization of the dynamic program considered in previous sections. To illustrate this, in this section we study two specific examples considering a static uncertain goal and obstacle.

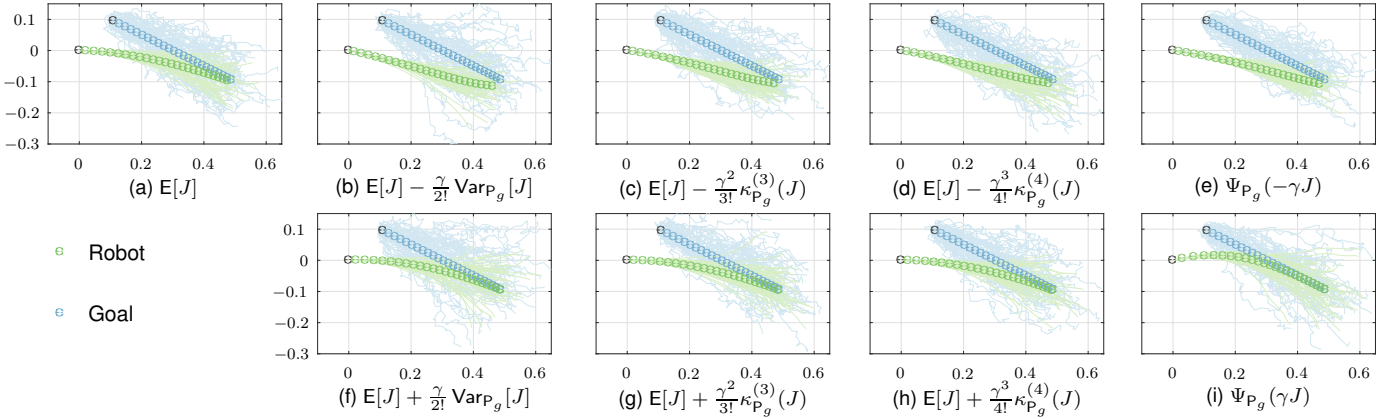


Fig. 1: Simulated optimal trajectories for a 2D point mass damper robot \mathbf{x}^r tracking goal \mathbf{x}^g with uncertain mass-damper system dynamics and where $\gamma = 9$ for a horizon of $T_c = 0.5s$. Initial states of robot and goal are $\mathbf{x}_0^r = [0 \ 0]$, $\dot{\mathbf{x}}_0^r = [0 \ 0]$, $\mathbf{x}_0^g = [0.1 \ 0.1]$ and $\dot{\mathbf{x}}_0^g = [1 \ -0.5]$. Results are shown every $0.02s$.

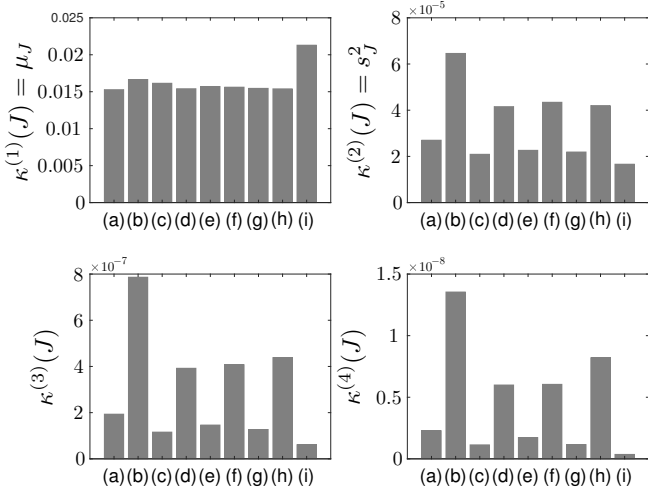


Fig. 2: Sample statistics of the cost for 10^4 simulations of each of the policies (a)-(i) from Fig. 1 showing the first 4 cumulants.

Consider a cost functional in the form

$$J(\mathbf{u}'(\cdot), \mathbf{x}'(\cdot)) = h_g(\mathbf{x}^r(T_c), \mathbf{x}^g(T_c)) + \int_{t=0}^{T_c} h_g(\mathbf{x}^r(t), \mathbf{x}^g(t)) + h_o(\mathbf{x}^r(t), \mathbf{x}^o(t)) + h_{so}(\mathbf{x}^r(t), \mathbf{x}^{so}(t)) + h_u(\mathbf{u}(t)) dt, \quad (36)$$

where $h_g(\mathbf{x}^r(t), \mathbf{x}^g(t))$ penalizes the distance to desired configurations, $h_u(\mathbf{u}^r(t))$ penalizes control efforts and $h_o(\mathbf{x}^r(t), \mathbf{x}^o(t))$ favors configurations distant to dynamic obstacles and $h_{so}(\mathbf{x}^r(t), \mathbf{x}^{so}(t))$ to static obstacles.

A. Static uncertain goal

Let \mathbf{x}^g be normally distributed and static goal such that $\mathbf{x}^g(t) \sim \mathcal{N}(\boldsymbol{\mu}^g, \Sigma^g)$. A typical convex functional used to reach a desired goal is the quadratic functional

$$h_g(\mathbf{x}^r(t), \mathbf{x}^g) = \frac{1}{2}(\mathbf{x}^r(t) - \mathbf{x}^g)^\top Q(\mathbf{x}^r(t) - \mathbf{x}^g),$$

where $Q \succeq 0$. In this case, a marginal analysis of the cumulant generating function of (36) in probability space $(\Omega_g, \mathcal{F}_g, P_g)$ is only relevant concerning $h_g(\mathbf{x}^r(t), \mathbf{x}^g)$ (any other terms are only present in the expected value). Its cumulant generating function is given by

$$\log E_{P_g}[\exp\{\theta h_g(\mathbf{x}^r(t), \mathbf{x}^g)\}] = -\frac{1}{2} \log |I_{q_g} - \theta Q \Sigma^g| + \frac{1}{2}(\mathbf{x}^r(t) - \boldsymbol{\mu}^g)^\top (I - \theta Q \Sigma^g)^{-1} Q(\mathbf{x}^r(t) - \boldsymbol{\mu}^g). \quad (37)$$

Applying (10) and ignoring the constant term which does not depend on $\mathbf{x}^r(t)$, the computation of any cumulant is straightforward. For example, the mean and variance of h_g are

$$E_{P_g}[h_g] = \frac{1}{2}(\mathbf{x}^r(t) - \boldsymbol{\mu}^g)^\top Q(\mathbf{x}^r(t) - \boldsymbol{\mu}^g) + const$$

$$\text{Var}_{P_g}[h_g] = \frac{1}{2}(\mathbf{x}^r(t) - \boldsymbol{\mu}^g)^\top Q \Sigma^g Q(\mathbf{x}^r(t) - \boldsymbol{\mu}^g) + const$$

Remark 5. Previous work rely on the Mahalanobis distance in order to approach this problem [13], [44], i.e.

$$\frac{1}{2}(\mathbf{x}^r(t) - \boldsymbol{\mu}^g)^\top (\Sigma^g)^{-1} Q(\mathbf{x}^r(t) - \boldsymbol{\mu}^g).$$

This expression becomes pathological and numerically unstable when the goal tends to the deterministic case, i.e. $\Sigma^g \rightarrow 0_{q_g}$. Adding a regularization quantity given by the identity matrix to avoid this issue yields

$$\frac{1}{2}(\mathbf{x}^r(t) - \boldsymbol{\mu}^g)^\top (I + \Sigma^g)^{-1} Q(\mathbf{x}^r(t) - \boldsymbol{\mu}^g). \quad (38)$$

Observing (37), it becomes apparent that (38) is a specific instance of a risk-seeking evaluation of h_g , where $\theta = -1$ and the goal variance is $\Sigma^{g'} = Q^{-1} \Sigma^g$.

B. Static uncertain obstacle

Let \mathbf{x}^{so} be normally distributed and static such that $\mathbf{x}^{so}(t) \sim \mathcal{N}(\boldsymbol{\mu}^{so}, \Sigma^{so})$. An effective convex and

continuously differentiable cost functional typically used in potential fields for collision avoidance is given by

$$h_{so}(\mathbf{x}^r(t), \mathbf{x}^{so}) = \beta \exp\left\{-\frac{1}{2}(\mathbf{x}^r(t) - \mathbf{x}^{so})^\top Q(\mathbf{x}^r(t) - \mathbf{x}^{so})\right\},$$

where $\beta \in \mathbb{R}$ is a weighting factor. In this case, its cumulant generating function has no closed form. The expectation, though, is given by

$$\begin{aligned} \mathbb{E}_{p_{so}}[h_{so}] &= \text{const} \cdot \beta \\ &\cdot \exp\left\{-\frac{1}{2}(\mathbf{x}^r(t) - \boldsymbol{\mu}^{so})^\top (I + Q\Sigma^{so})^{-1} Q(\mathbf{x}^r(t) - \boldsymbol{\mu}^{so})\right\} \end{aligned} \quad (39)$$

and already considers goal variance Σ^{so} . In fact, we can interpret this expression as the moment generating function of the negative quadratic cost. This result is equivalent to the generalized binary saturating cost considered in [45] as an attractor instead of a repelling obstacle.

VII. EVALUATION

In order to illustrate the peculiarities of the proposed approach we performed simulations on a point-mass robot following an uncertain goal trajectory in a scene with uncertain dynamic and static obstacles. A similar implementation on a two-link manipulator and a car-like robot validates the proposed approach for non-linear dynamics. We also implemented our approach in a robot manipulator to validate its performance in a real setting. All experiments rely on the iterative algorithm explained in the previous section.

A. 2D Point-Mass robot

Consider a two-dimensional point robot with state $\mathbf{x}^r = [\mathbf{p}^r \ \dot{\mathbf{p}}^r]$ where $\mathbf{p}^r \in \mathbb{R}^2$ is its position and with dynamics

$$M\ddot{\mathbf{p}}^r + D\dot{\mathbf{p}}^r = \mathbf{u}^r. \quad (40)$$

1) *Uncertain goal trajectory*:: we first validate the proposed iterative algorithm for high-order cost statistics. The control task consists of following desired trajectory $\mathbf{x}^g \in \mathbb{R}^2$ with dynamics given by a mass-damper system with noise. The joint dynamics of the problem are given by (4) but considering only robot and goal difference, i.e. $\boldsymbol{\xi}^r = [\mathbf{x}^g - \mathbf{x}^r]$, and a single uncertainty source related to the desired trajectory. We consider performance measure (36) with

$$h_g = \frac{1}{2}\boldsymbol{\xi}^r{}^\top Q_g \boldsymbol{\xi}^r \quad h_u = \frac{1}{2}\mathbf{u}^r{}^\top R \mathbf{u}^r \quad h_o = 0 \quad h_{so} = 0.$$

The optimal policy takes the form $\mathbf{u}^r = \bar{\mathbf{u}}^r + [L_{\mathbf{x}^g} \ L_{\dot{\mathbf{x}}^g}]\boldsymbol{\xi}$. In the following results, parameters were fixed to $M = I_2$ kg, $D = I_2$ Ns/m, $R = 10^{-2}I_2$, $Q_g = \text{diag}\{1 \ 1 \ 0.1 \ 0.1\}$, $\Delta = 10^{-2}$ and $T_c = 0.5$ s. Goal dynamics are assumed to have diffusion coefficient matrix $\mathcal{G}^g(t) = I_4$ and identical mass and damping to the robot's.

Simulated trajectories for several cost-cumulant and risk-sensitive controls are depicted in Fig. 1. The solution corresponding to the expected cost depicted in Fig. 1(a) serves as

the risk-neutral reference. Fig. 1(e) and Fig. 1(i) show risk-sensitive solutions in their seeking and averse variants respectively. The risk-seeking policy adapts its trajectory tolerating higher distances to the goal. This policy is desirable when goal uncertainty suggests more flexibility or unconstrained regions, for instance in PbD settings [44]. The risk-averse solution has the opposite effect, tracking goal dynamics more aggressively as well as increasing feedback gains thereby becoming stiffer. This behavior is more in accordance with navigation scenarios where uncertainty may hinder performance. Cost cumulant solutions are shown in Fig. 1(f), Fig. 1(g), Fig. 1(h) for the second, third and fourth cumulant in their risk-averse variants. All three cases are similar: the feedforward trajectory is marginally adapted diminishing the distance to the goal with higher feedback gains. Their risk-seeking counterparts depicted in Fig. 1(b), Fig. 1(c), Fig. 1(d) accept less tracking precision w.r.t the risk-neutral policy Fig. 1(a). To validate the implementation of our iterative algorithm, the first sample four cost cumulants of all control policies depicted in Fig. 1 are shown in Fig. 2. Although the first cumulant (sample mean) is similar for all policies except for (i), higher order sample cumulants show that all risk-seeking variants produce higher cost variability w.r.t the expected cost policy (a) while risk-averse policies increase robustness by diminishing cost variability.

2) *Uncertain goal trajectory and dynamic+static obstacles*: we now include an uncertain dynamic obstacle \mathbf{x}^o in the scene with mass-damper dynamics. This is a similar setting to the one considered in [35]. The joint error dynamics are given by (4). The cost remains identical to the previous setting except for the dynamic obstacle-related term, which is set to $h_o = \beta \exp\{-0.5(\mathbf{x}^r - \mathbf{x}^o)^\top Q_o(\mathbf{x}^r - \mathbf{x}^o)\}$, with $Q_o = 200I_2$ and $\beta = 0.1$. The optimal policy is now

$$\mathbf{u}^r = \bar{\mathbf{u}}^r + \begin{bmatrix} L_{\mathbf{x}^g} & L_{\dot{\mathbf{x}}^g} \\ L_{\mathbf{x}^o} & L_{\dot{\mathbf{x}}^o} \end{bmatrix} \boldsymbol{\xi}.$$

Obstacle dynamics are assumed to have diffusion coefficient matrix $\mathcal{G}^o(t) = 0.3I_4$ and identical mass and damping to the robot's. We also consider static normally distributed obstacles in the scene by setting the static obstacle term h_{so} from (36) to the expected value of the weighted exponential decay (39) as explained in Section VI, with $Q = I_2$ and $\beta = 0.1$. To compute collisions we assume that the robot has a radius of 0.05m. In addition, the goal and obstacle mass-damper systems are driven by external forces $\mathbf{u}_g = [\cos(2x_1^g) \ \sin(2x_2^g)]^\top$ and $\mathbf{u}_o = [\cos(10x_1^o) \ \sin(10x_2^o)]^\top$. Note that the uncertainty of an obstacle is an indicator of the probability of collision. Therefore, we consider its marginal cost variability in a risk-averse manner, as it is instrumental to diminish the risk of collisions, while the goal variability is assessed with a risk-seeking attitude assuming more flexibility.

The resulting trajectories are shown in Fig. 3(b) for the cost cumulant case in comparison with the risk-neutral policy from Fig. 3(a). On one side, risk-aversion towards obstacle variability yields a pronounced deviation at the beginning of the simulation due to the proximity of the obstacle. On the other side risk-seekingness w.r.t the goal uncertainty tolerates higher distance to the desired trajectory enabling

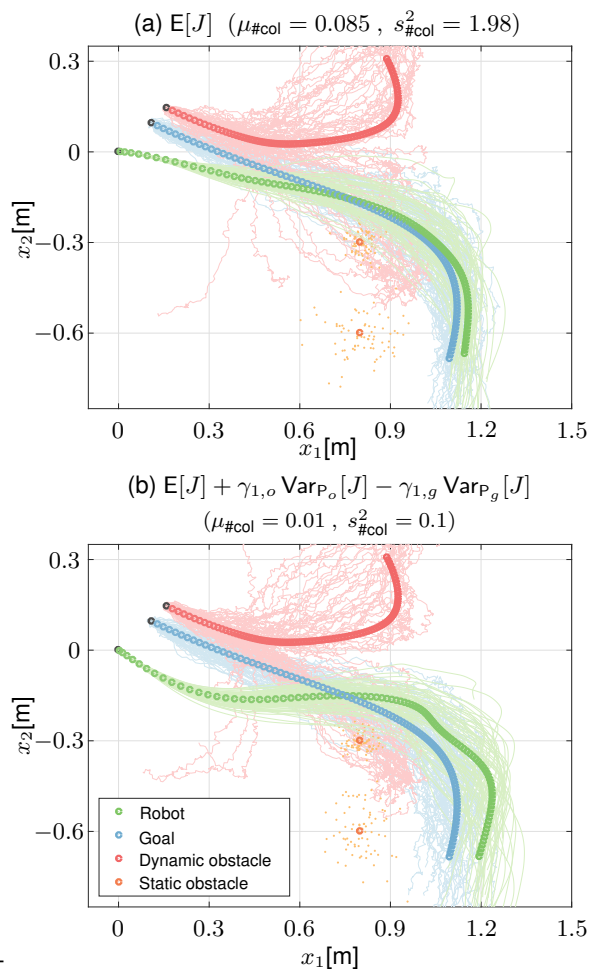


Fig. 3: Simulated trajectories for a 2D point mass damper robot x^r tracking goal x^g with mass-damper system dynamics and avoiding obstacle x^o with noisy mass-damper system dynamics including static obstacles and where $\gamma_{1,g} = 4$ and $\gamma_{1,o} = 16$ for a horizon of $T_c = 2s$. The two static obstacles considered are normally distributed, centered in $(0.8, -0.3)$ and $(0.8, -0.6)$ with covariance matrices $0.1I_2$ and $0.5I_2$ respectively. Initial states of robot, goal and obstacle are $x_0^r = [0 \ 0]$, $\dot{x}_0^r = [0 \ 0]$, $x_0^g = [0.1 \ 0.1]$, $\dot{x}_0^g = [1 \ -0.5]$, $x_0^o = [0.15 \ 0.15]$ and $\dot{x}_0^o = [1 \ -0.5]$. Results are shown every 0.02s.

better obstacle avoidance. The influence of the static obstacles' uncertainty is also evident; the one with low variance centered at $(0.8, -0.3)$ produces almost no deviations from the expected goal trajectory in comparison with the obstacle with high variance centered at $(0.8, -0.6)$. The benefits of the cost cumulant policy are also shown in terms of the number of collisions: both the sample average $\mu_{\#col}$ and the sample variance $s_{\#col}^2$ are significantly reduced w.r.t. the risk-neutral policy. In order to illustrate the influence of the uncertainty level in the resulting policy, Fig. 4 shows solutions for the same setting but with different diffusion coefficient matrices. From Fig. 4(a) to Fig. 4(d), decreasing uncertainty for both obstacle and goal dynamics are considered. As uncertainty diminishes, results approach the risk-neutral solution as shown

by the almost deterministic setting from Fig. 4(d). This validates our proposed approach as a suitable systematic method of considering uncertainty the seamlessly generalizes to a deterministic problem.

B. Two-link manipulator

To evaluate the validity of the algorithm in settings with non-linear dynamics, we consider a torque-controlled arm with two joints moving in the horizontal plane with inverse dynamics

$$M(\theta)\ddot{\theta} + C(\theta, \dot{\theta}) + B\dot{\theta} = \tau,$$

where $\theta \in \mathbb{R}^2$ are the joint angles, $M(\theta)$ is the inertia matrix, $C(\theta, \dot{\theta})$ is the vector of centripetal and Coriolis forces, B is the joint friction matrix and $\tau \in \mathbb{R}^2$ are the joint torques. We set the mass of each link to $m_1 = 1.4\text{kg}$ and $m_2 = 1.1\text{kg}$, the length of each link to $l_1 = 0.3\text{m}$ and $l_2 = 0.33$, the moments of inertia to $I_1 = 0.025 \text{ kg} \cdot \text{m}^2$ and $I_2 = 0.045 \text{ kg} \cdot \text{m}^2$ and we assume the center of mass of each link is placed at the link's center. The joint friction matrix is set to

$$B = \begin{bmatrix} 0.05 & 0.025 \\ 0.025 & 0.05 \end{bmatrix}.$$

In line with the previous subsection, the control task consists of tracking a goal $x^g \in \mathbb{R}^2$ with uncertain mass-damper dynamics while avoiding a dynamic obstacle $x^o \in \mathbb{R}^2$ with similar dynamics. The joint state is given by (4), where x^r represents now the workspace coordinates of the manipulator and the control input u^r are the joint torques in workspace coordinates. The cost function is identical to the previous subsection with the same parameters.

Optimal trajectories are shown in Fig. 5 for the expected cost policy in comparison with the mean variance solution. While the expected cost solution of Fig. 5(a) opts for a faster trajectory at the beginning to track more precisely the goal dynamics, the risk-averse evaluation of obstacle variability in Fig. 5(b) waits for the obstacle at the beginning and significantly reduces its tracking performance. This effect is also boosted by the risk-seeking evaluation of goal variability. As a result, the average number of collisions is also significantly reduced.

C. Car-like Robot

Consider a simplified car-like robot model with state space $x^r = [x_1 \ x_2 \ \theta \ v \ c]^T$ and dynamics

$$\dot{x}_1 = v \cos \theta \quad \dot{x}_2 = v \sin \theta \quad \dot{\theta} = vc \quad \dot{v} = u_1 \quad \dot{c} = u_2 \quad (41)$$

where x_1 and x_2 denote the 2D position, θ the orientation, v the velocity, c the curvature and the control input is given by $u = [u_1 \ u_2]^T$.

The control task consists of tracking goal x^g with dynamics (41) with additive noise while avoiding an obstacle x^o with uncertain mass-damper dynamics. Their diffusion coefficient matrices are given by $\mathcal{G}^g = \text{diag}\{1 \ 1 \ 1 \ 0 \ 0\}$

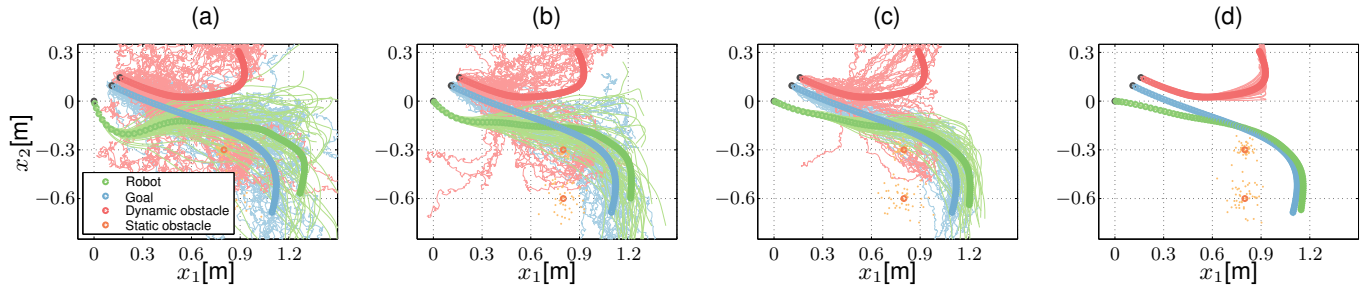


Fig. 4: Optimal trajectories and feedback gains for the same setting from Fig. 3(b) but considering different diffusion coefficient matrices for the obstacle and goal diffusions. (a) shows results for $\mathcal{G}^g = \mathcal{G}^o = 1.6I_4$, (b) for $\mathcal{G}^g = \mathcal{G}^o = 0.8I_4$, (c) for $\mathcal{G}^g = \mathcal{G}^o = 0.3I_4$ and (d) for $\mathcal{G}^g = \mathcal{G}^o = 0.001I_4$.

and $\mathcal{G}^o = I_4$ respectively. The state of the system is given by

$$\xi' = \left[(\mathbf{x}^g - \mathbf{x}^r)^\top (\mathbf{x}^o - [x_1 \ x_2]^\top)^\top (\dot{\mathbf{x}}^o - [\dot{x}_1 \ \dot{x}_2]^\top)^\top \right]^\top.$$

The cost function considered is identical to the previous setting, except for the goal's quadratic weighting matrix and the control input's weighting matrix, which are set to $Q_g = \text{diag}\{1 \ 1 \ 0 \ 0 \ 0 \ 0\}$ and $R = 10^{-5}$ respectively.

The resulting optimal trajectories are depicted in Fig. 6 for the risk neutral and the mean variance solution with aversion towards obstacle-related variability and risk-seekingness towards goal-related cost variance. The goal tracking flexibility provided by the risk-preferring evaluation of goal variability enables deviations from the desired trajectory, providing improved obstacle avoidance. As a result the number of collisions is also diminished.

D. Real manipulator

To validate our approach in a real system we implemented our algorithm on a 7 DOF KUKA LWR 4+ manipulator equipped with a JR3 force/torque sensor implementing dynamics (40) in \mathbb{R}^3 , with $M = 10I_3$ kg and $D = 30I_3$ Ns/m. We consider a similar setting to the one from Section VII-B, where the control task consists of tracking a goal $\mathbf{x}^g \in \mathbb{R}^3$ with uncertain linear attractor dynamics while avoiding a dynamic obstacle $\mathbf{x}^o \in \mathbb{R}^3$ with similar dynamics. The obstacle position is retrieved online by means of a Qualisys tracking system. Both goal and obstacle have goals and initial positions that could potentially lead to collisions, see Fig. 7. The joint state is given by (4), where \mathbf{x}^r represents now the position of the endeffector and the control input \mathbf{u}' are the forces applied at the endeffector. Goal and obstacle dynamics are assumed to have diffusion coefficient matrix $\mathcal{G}^g(t) = I_4$ and identical mass and damping to the robot's. The cost function is identical to the previous subsection with parameters $R = 10^2 I_3$, $Q_g = \text{diag}\{10^4 \ 10^4 \ 10^4 \ 0 \ 0 \ 0\}$, $\Delta = 10^{-2}$, $Q_o = 100I_3$, $\beta = 10^6$ and $T_c = 5.0s$.

The trajectories of 7 runs of a risk-neutral control are depicted in Fig. 8(a) against a control with a risk-averse attitude towards obstacle-related variability and risk-seekingness towards goal-related cost variance, shown in Fig. 8(b). Similar to the other examples, the resulting trajectories for the risk-sensitive solution deviate from the desired goal trajectory

keeping a safer distance to the obstacle, while the expected solution ignores variability and closely follows the desired trajectory.

VIII. CONCLUSION AND FUTURE WORK

Motivated by the application of probabilistic methods as a tool to estimate dynamic models of a desired trajectory, an obstacle or the robot, in this work we presented a systematic approach to uncertainty-dependent optimal robot control for non-linear dynamics, non-quadratic costs and multiple uncertainty sources. Our approach iteratively computes locally optimal feedback solutions considering high-order cost cumulants. As a result, robot actions react according to and specifically to each uncertainty source. Results on both linear and non-linear plants with non-quadratic costs validate the approach and highlight its flexibility. Interesting areas of future research are the analysis of predictive control variants as well as input- and state-dependent noise extensions.

ACKNOWLEDGMENTS

This research is partly supported by the ERC Starting Grant ‘‘Control based on Human Models (con-humo)’’ under grant agreement 337654. The authors would like to thank Pietro Blandizzi for his help with the robotic experiments and the reviewers for their insightful and helpful comments.

APPENDIX

Proof of Lemma 1. The solution to recursive equation (18) entails only complication in the term involving the expectation, i.e. $\theta^{-1} \Psi_P(\theta J(\xi_{k+1}, \mathbf{u}_{k+1 \dots T-1}))$. Assuming (19) holds yields

$$\begin{aligned} & \theta^{-1} \Psi_P(\theta J(\xi_{k+1}, \mathbf{u}_{k+1 \dots T-1})) = \\ & \theta^{-1} \log \text{E}_P[\exp\{\theta(\frac{1}{2} \xi_{k+1}^\top W_{k+1} \xi_{k+1} + \xi_{k+1}^\top \mathbf{w}_{k+1} + w_{k+1})\}]. \end{aligned}$$

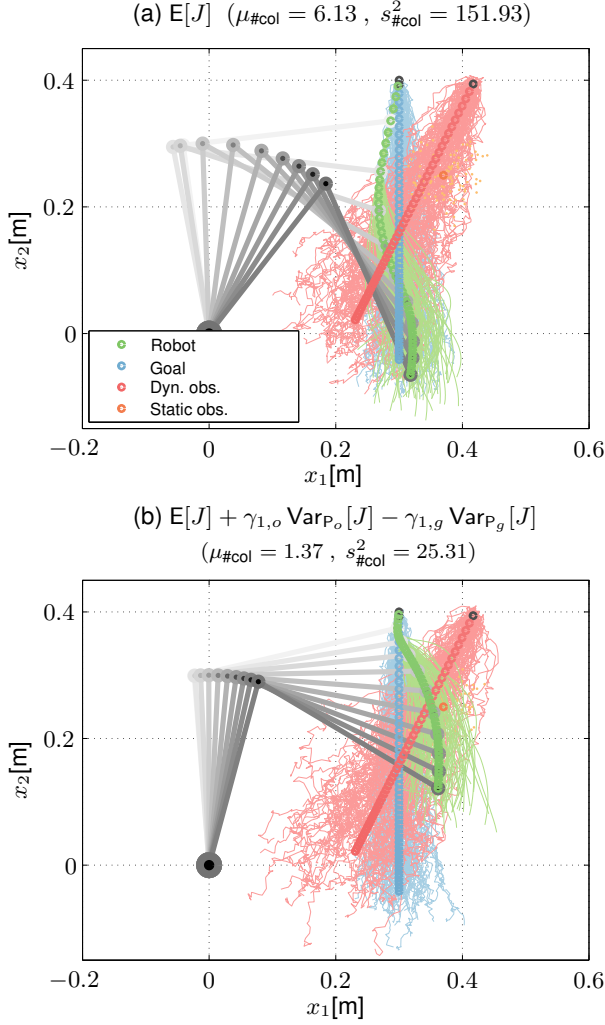


Fig. 5: Simulated trajectories for a two-link manipulator \mathbf{x}^r tracking goal \mathbf{x}^g with uncertain mass-damper dynamics and avoiding obstacle \mathbf{x}^o with uncertain mass-damper dynamics including static obstacles and where $\gamma_{1,g} = 100$ and $\gamma_{1,o} = 100$. The static obstacle considered is normally distributed, centered at $(0.37, 0.25)$ with covariance matrix $0.3I_2$. Initial states of robot, goal and obstacle are $\mathbf{x}_0^r = [0.3 \ 0.4]$, $\dot{\mathbf{x}}_0^r = [0 \ 0]$, $\mathbf{x}_0^g = [0.3 \ 0.4]$, $\dot{\mathbf{x}}_0^g = [0 \ -0.7]$, $\mathbf{x}_0^o = [0.42 \ 0.4]$ and $\dot{\mathbf{x}}_0^o = [0.3 \ -0.6]$. Results are shown every 0.02s.

Let now $\mathbf{m}_k = A_k \boldsymbol{\xi}_k + B_k \mathbf{u}_k$. Considering dynamics (17), the expression for the expectation is given by the Gaussian integral

$$\begin{aligned}
 & \int \sqrt{(2\pi)^n |\Sigma_k|} \exp \left\{ -\frac{1}{2} \boldsymbol{\epsilon}_k^T \Sigma_k^{-1} \boldsymbol{\epsilon}_k \right\} \\
 & \cdot \exp \left\{ \theta \left(\frac{1}{2} (\mathbf{m}_k + \boldsymbol{\epsilon}_k)^T W_{k+1} (\mathbf{m}_k + \boldsymbol{\epsilon}_k) \right. \right. \\
 & \quad \left. \left. + (\mathbf{m}_k + \boldsymbol{\epsilon}_k)^T \mathbf{w}_{k+1} + w_{k+1} \right) \right\} d\boldsymbol{\epsilon}_k \\
 & = \sqrt{(2\pi)^n |\Sigma_k|} \\
 & \cdot \exp \left\{ \theta \left(\frac{1}{2} \mathbf{m}_k^T W_{k+1} \mathbf{m}_k + \mathbf{m}_k^T \mathbf{w}_{k+1} + w_{k+1} \right) \right\} \\
 & \cdot \int \exp \left\{ -\frac{1}{2} \boldsymbol{\epsilon}_k^T (\Sigma_k^{-1} - \theta W_{k+1}) \boldsymbol{\epsilon}_k \right. \\
 & \quad \left. + \boldsymbol{\epsilon}_k^T (\theta W_{k+1} \mathbf{m}_k + \theta \mathbf{w}_{k+1}) \right\} d\boldsymbol{\epsilon}_k . \quad (42)
 \end{aligned}$$

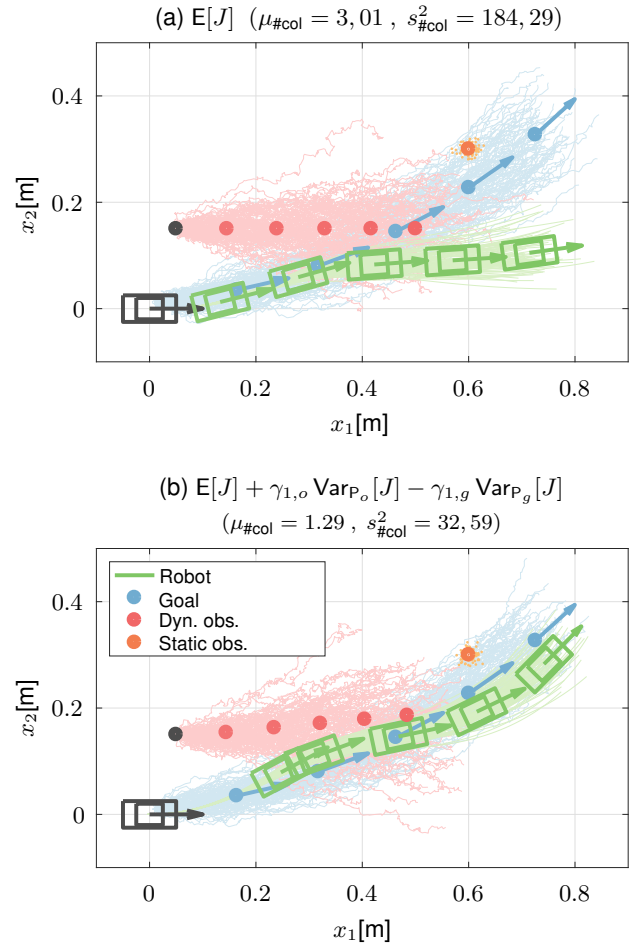


Fig. 6: Simulated trajectories for a car-like robot \mathbf{x}^r tracking a goal \mathbf{x}^g with uncertain mass-damper dynamics and avoiding obstacle \mathbf{x}^o with uncertain mass-damper dynamics for a horizon of $T_c = 2s$, where $\gamma_{1,o} = \gamma_{1,g} = 0.1$. The static obstacle considered is normally distributed, centered at $(0.6, 0.3)$ with covariance matrix $0.1I_2$. Initial states of robot, goal, and obstacle are $\mathbf{x}^r = [0 \ 0 \ 0 \ 0.01 \ 0.01]$, $\mathbf{x}^g = [0.01 \ 0.01 \ 0.1 \ 0.4 \ 0.8]$, $\mathbf{x}_0^o = [0.05 \ 0.15]$ and $\dot{\mathbf{x}}_0^o = [0.2 \ 0.3]$. Results are shown every 0.2s.

Let $N_k = (\Sigma_k^{-1} - \theta W_{k+1})$ and $\mathbf{n}_k = 2\theta W_{k+1} \mathbf{m}_k + \theta \mathbf{w}_{k+1}$. If $N_k \succ 0$ the integral has analytical solution

$$\int \exp \left\{ -\frac{1}{2} \boldsymbol{\epsilon}^T N_k \boldsymbol{\epsilon} + \boldsymbol{\epsilon}^T \mathbf{n}_k \right\} d\boldsymbol{\epsilon} = \sqrt{(2\pi)^n |N_k|^{-1}} \exp \left\{ \mathbf{n}_k^T N_k^{-1} \mathbf{n}_k \right\} .$$

Note that condition $(\Sigma_k^{-1} - \theta W_{k+1}) \succ 0$ motivates the initial condition from Lemma 1 and is satisfied if $\theta < (\lambda_{max}(\Sigma_k) \lambda_{max}(W_{k+1}))^{-1}$, where λ_{max} denotes the largest eigenvalue [32].

Considering that $|(\Sigma_k^{-1} - \theta W_{k+1})^{-1} (I - \theta W_{k+1} \Sigma_k)| = |\Sigma_k|$

$$\begin{aligned}
 & W_{k+1} + \theta W_{k+1} (\Sigma_k^{-1} - \theta W_{k+1})^{-1} W_{k+1} = \\
 & (I - \theta W_{k+1} \Sigma_k)^{-1} W_{k+1} ,
 \end{aligned}$$

with similar formulations for \mathbf{w}_{k+1} and w_{k+1} , the expectation term

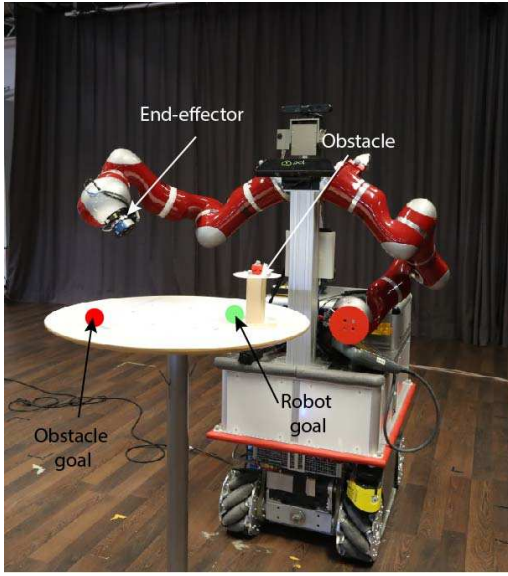


Fig. 7: Robotic setup consisting of a robot manipulator equipped with a force-torque sensor at its end-effector and an obstacle equipped with markers and tracked with a Qualisys system.

has closed form solution

$$\sqrt{|I - \theta W_{k+1} \Sigma_k|} \cdot \exp \left\{ \theta_s \frac{1}{2} \mathbf{m}_k^T \widetilde{W}_{k+1} \mathbf{m}_k + \theta_s \mathbf{m}_k^T \widetilde{\mathbf{w}}_{k+1} + \theta_s \widetilde{w}_{k+1} \right\}$$

where \widetilde{W}_{k+1} , $\widetilde{\mathbf{w}}_{k+1}$, and \widetilde{w}_{k+1} are defined as in (23), (24) and (25). Inserting this result and expressions (14) and (15) into (18) completes the proof. ■

Proof of Lemma 2. The results for $W_k^{[r]}$, $\mathbf{w}_k^{[r]}$ and $w_k^{[r]}$ are straightforward from their definitions in (20), (21) and (22). For the proof of expression $\widetilde{W}_{k+1}^{[r]}$ see [32]. The solution to $\widetilde{\mathbf{w}}_{k+1}^{[r]}$ and $\widetilde{w}_{k+1}^{[r]}$ is derived with the same procedure. We reformulate (24) as

$$\widetilde{\mathbf{w}}_{k+1} = \mathbf{w}_{k+1} + \theta W_{k+1} \Sigma_k \widetilde{\mathbf{w}}_{k+1}. \quad (43)$$

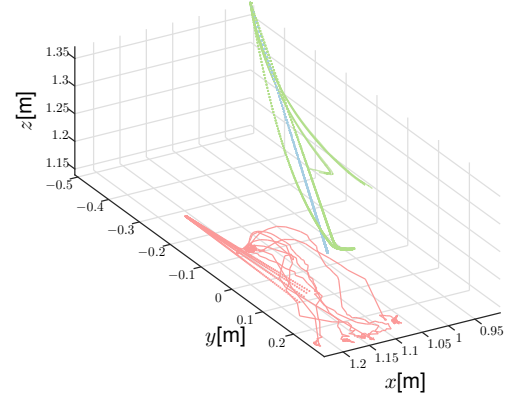
Applying Leibniz's Theorem twice we get

$$\begin{aligned} \widetilde{\mathbf{w}}_{k+1}^{[r]} &= \mathbf{w}_{k+1}^{[r]} + \theta (W_{k+1} \Sigma_k \widetilde{\mathbf{w}}_{k+1}^{[r]})^{[r]} \\ &\quad + r (W_{k+1} \Sigma_k \widetilde{\mathbf{w}}_{k+1}^{[r]})^{[r-1]} \\ &= \mathbf{w}_{k+1}^{[r]} + \theta (W_{k+1} \Sigma_k \widetilde{\mathbf{w}}_{k+1}^{[r]})^{[r]} \\ &\quad + r \sum_{j=0}^{r-1} \mathcal{C}_{r-1}^j W_{k+1}^{[j]} \Sigma_k \widetilde{\mathbf{w}}_{k+1}^{[r-1-j]}, \end{aligned}$$

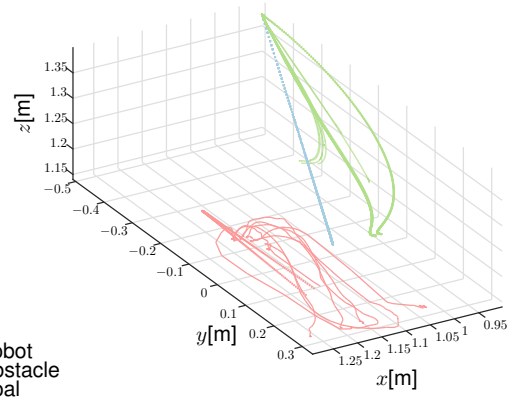
which after evaluating $\theta = 0$ yields (31). The procedure for $\widetilde{w}_{k+1}^{[r]}$ is analogous. Note that the term $-\frac{1}{2} \log F_k$ from (25) gives rise to expression $H_t^{[r]}$ as already shown in [32]. ■

REFERENCES

- [1] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.
- [2] A. Faisal, L. Selen, and D. Wolpert, "Noise in the nervous system," *Nature Reviews Neuroscience*, vol. 9, no. 4, pp. 292–303, 2008.
- [3] E. Todorov, "Optimality principles in sensorimotor control." *Nature Neuroscience*, vol. 7, pp. 907–15, 2004.



(a) $E[J]$



(b) $\Psi_{P_g}(-10J) + \Psi_{P_o}(10^{-4}J)$

Fig. 8: Resulting trajectories of the robot, obstacle and goal of a risk-neutral control (a) and a risk-sensitive control (b) for the robotic setup depicted in Fig. 7.

- [4] J. Berg, S. Patil, and R. Alterovitz, "Motion planning under uncertainty using iterative local optimization in belief space," *The International Journal of Robotics Research*, vol. 31, no. 11, pp. 1263–1278, 2012.
- [5] D. Mitrovic, S. Klanke, and S. Vijayakumar, "Adaptive optimal feedback control with learned internal dynamics models," in *From Motor Learning to Interaction Learning in Robots*, ser. Studies in Computational Intelligence. Springer Berlin Heidelberg, 2010, vol. 264, pp. 65–84.
- [6] D. Braun, A. Nagengast, and D. Wolpert, "Risk-sensitivity in sensorimotor control," *Frontiers in Human Neuroscience*, vol. 5, no. 1, 2011.
- [7] J. Grau-Moya, P. Ortega, and D. A. Braun, "Risk-sensitivity in bayesian sensorimotor integration," *PLoS Comput Biol*, vol. 8, no. 9, p. e1002698, 09 2012.
- [8] J. Grau-Moya, E. Hez, G. Pezzulo, and D. Braun, "The Effect of Model Uncertainty on Cooperation in Sensorimotor Interactions," *Journal of the Royal Society Interface*, vol. 10, no. 87, 2013.
- [9] J. Medina and S. Hirche, "Uncertainty-dependent Optimal Control for Robot Control Considering High-order Cost Statistics," in *Proc. IEEE/RSJ IROS*, 2015, pp. 3995–4002.
- [10] J. Müller and G. Sukhatme, "Risk-aware trajectory generation with application to safe quadrotor landing," in *Proc. IEEE/RSJ IROS*, 2014, pp. 3642–3648.
- [11] D. Mitrovic, S. Klanke, and S. Vijayakumar, "Learning impedance control of antagonistic systems based on stochastic optimization principles," *IJRR*, no. 5, pp. 556–573, 2010.
- [12] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," *IEEE Trans. Syst., Man, Cybern. B*, vol. 37, no. 2, pp. 286–298, 2007.
- [13] J. Medina, D. Lee, and S. Hirche, "Risk Sensitive Optimal Feedback Control for Haptic Assistance," in *Proc. IEEE ICRA*, 2012, pp. 1025–

- 1031.
- [14] W. Langson, I. Chrysochoos, S. Raković, and D. Mayne, "Robust model predictive control using tubes," *Automatica*, vol. 40, no. 1, pp. 125–133, 2004.
- [15] A. Bemporad and M. Morari, "Robust model predictive control: A survey," pp. 207–226, 1999.
- [16] C. Won, "Cost distribution shaping: the relation between bode integral, entropy, risk-sensitivity, and cost cumulant control," in *Proc. ACC*, vol. 3, 2004, pp. 2160–2165 vol.3.
- [17] E. Todorov and W. Li, "A generalized iterative lqg method for locally-optimal feedback control of constrained nonlinear stochastic systems," in *Proc. ACC*, 2005, pp. 300–306.
- [18] D. Jacobsen and D. Mayne, *Differential Dynamic Programming*. Elsevier, 1970.
- [19] E. Theodorou, Y. Tassa, and E. Todorov, "Stochastic differential dynamic programming," in *American Control Conference (ACC), 2010*. IEEE, 2010, pp. 1125–1132.
- [20] V. Huynh and N. Roy, "icLQG: combining local and global optimization for control in information space," in *Proc. IEEE ICRA*, 2009, pp. 2851–2858.
- [21] T. Erez and W. Smart, "A scalable method for solving high-dimensional continuous pomdps using local approximation," pp. 160–167, 2010.
- [22] L. Cosenza, "On the minimum variance control of discrete-time systems." Ph.D. dissertation, University of Notre Dame, Department of Electrical Engineering, 1969.
- [23] D. Jacobson, "Optimal stochastic linear systems with exponential performance criteria and their relation to deterministic differential games," *IEEE Trans. Autom. Control*, vol. 18, no. 2, pp. 124 – 131, apr 1973.
- [24] M. Sain, C. Won, J. Spencer, and S. Liberty, "Cumulants and risk-sensitive control: a cost mean and variance theory with application to seismic protection of structures," in *Advances in Dynamic Games and Applications*. Springer, 2000, pp. 427–459.
- [25] P. Whittle, *Risk-Sensitive Optimal Control*. Wiley, 1990.
- [26] F. Hover, S. Franz, and M. Triantafyllou, "Application of polynomial chaos in stability and control," *Automatica*, vol. 42, no. 5, pp. 789–795, 2006.
- [27] A. Mesbah, S. Streif, R. Findeisen, and R. Braatz, "Stochastic nonlinear model predictive control with probabilistic constraints," in *American Control Conference (ACC), 2014*. IEEE, 2014, pp. 2413–2419.
- [28] S. Kuindersma, R. Grunpen, and A. Barto, "Variational bayesian optimization for runtime risk-sensitive control," *Robotics: Science and Systems VIII*, pp. 201–208, 2012.
- [29] K. Rawlik, M. Toussaint, and S. Vijayakumar, "On stochastic optimal control and reinforcement learning by approximate inference," in *Proc. IJCAI*. AAAI Press, 2013, pp. 3052–3056.
- [30] K. Rawlik, "On probabilistic inference approaches to stochastic optimal control," The University of Edinburgh, 2013.
- [31] K. Pham, *Linear-Quadratic Controls in Risk-Averse Decision Making: Performance-Measure Statistics and Control Decision Optimization*. Springer, 2012.
- [32] F. Qian, J. Gao, and D. Li, "Complete statistical characterization of discrete-time lqg and cumulant control," *IEEE Trans. Autom. Control*, vol. 57, no. 8, pp. 2110–2115, 2012.
- [33] A. Paraschos, C. Daniel, J. Peters, and G. Neumann, "Probabilistic movement primitives," in *Advances in neural information processing systems*, 2013, pp. 2616–2624.
- [34] S. Calinon, F. D'Halluin, E. Sauser, D. Caldwell, and A. Billard, "Learning and reproduction of gestures by imitation," *IEEE Robot. Autom. Mag.*, vol. 17, no. 2, pp. 44–54, 2010.
- [35] G. Aoude, B. Luders, J. Joseph, N. Roy, and J. How, "Probabilistically safe motion planning to avoid dynamic obstacles with uncertain motion patterns," *Autonomous Robots*, vol. 35, no. 1, pp. 51–76, 2013.
- [36] M. Sain, C. Won, and B. Spencer, "Cumulant minimization and robust control," in *Stochastic Theory and Adaptive Control*. Springer, 1992, pp. 411–425.
- [37] J. Medina, T. Lorenz, and S. Hirche, "Synthesizing Anticipatory Haptic Assistance Considering Human Behavior Uncertainty," *IEEE Trans. Robot.*, vol. 31, no. 1, pp. 180–190, 2015.
- [38] W. Li and E. Todorov, "Iterative linearization methods for approximately optimal control and estimation of non-linear stochastic system," *International Journal of Control*, vol. 80, no. 9, pp. 1439–1453, 2007.
- [39] S. LaValle and J. Kuffner, "Randomized kinodynamic planning," *Int. J. Robot. Res.*, vol. 20, no. 5, pp. 378–400, 2001.
- [40] S. Yakowitz, "Algorithms and computational techniques in differential dynamic programming," *Control and Dynamic Systems*, vol. 31, pp. 75–91, 1989.

- [41] T. Erez, Y. Tassa, and E. Todorov, "Infinite-horizon model predictive control for periodic tasks with contacts," *Robotics: Science and Systems VII*, p. 73, 2012.
- [42] A. Billard, S. Calinon, R. Dillmann, and S. Schaal, "Robot Programming by Demonstration," in *Handbook of Robotics*, B. Siciliano and O. Khatib, Eds. Springer, 2008, pp. 1371–1394.
- [43] D. Lee and C. Ott, "Incremental kinesthetic teaching of motion primitives using the motion refinement tube," *Auton. Robot.*, pp. 1–17, 2011.
- [44] S. Calinon and A. Billard, "Statistical learning by imitation of competing constraints in joint space and task space," *Advanced Robotics*, vol. 23, pp. 2059–2076, 2009.
- [45] M. Deisenroth, D. Fox, and C. Rasmussen, "Gaussian Processes for Data-Efficient Learning in Robotics and Control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PP, no. 99, pp. 1–1, 2013.



José R. Medina received the Ingeniero Superior en Informática degree from the University of Seville, Spain, in 2008 and the Doktor-Ingenieur degree in electrical engineering, machine learning and control in 2015 from the Technische Universität München (TUM), Germany. From 2016 to 2017 he was a post-doctoral fellow at the Learning Algorithms and Systems Laboratory (LASA) of the École Polytechnique Fédérale de Lausanne (EPFL) and since September 2017 he is a robotics engineer at FRANKA EMIKA GmbH. His research interests include machine learn-

ing and robot control with applications to robot learning and human-robot interaction.



Sandra Hirche (M'03-SM'11) received the Diplom-Ingenieur degree in aeronautical engineering from Technical University Berlin, Germany, in 2002 and the Doktor-Ingenieur degree in electrical engineering from Technical University Munich, Germany, in 2005. From 2005 to 2007 she was awarded a Postdoc scholarship from the Japanese Society for the Promotion of Science at the Fujita Laboratory, Tokyo Institute of Technology, Tokyo, Japan. From 2008 to 2012 she has been an associate professor at Technical University Munich. Since 2013 she

is TUM Liesel Beckmann Distinguished Professor and heads the Chair of Information-oriented Control in the Department of Electrical and Computer Engineering at Technical University Munich. Her main research interests include cooperative, distributed and networked control with applications in human-robot interaction, multi-robot systems, and general robotics. She has published more than 150 papers in international journals, books and refereed conferences. Dr. Hirche has served on the Editorial Boards of the IEEE Transactions on Control of Network Systems, IEEE Transactions on Control Systems Technology, and the IEEE Transactions on Haptics.