Technische Universität München
Zentrum Mathematik
Lehrstuhl für Mathematische Optimierung

Numerical Methods and Second Order Theory for Nonsmooth Problems

Andre Manfred Milzarek

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Martin Brokate
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Michael Ulbrich
	2.	Univ.-Prof. Dr. Christian Kanzow Julius-Maximilians-Universität Würzburg
	3.	Prof. Dr. Defeng Sun National University of Singapore, Singapur

Die Dissertation wurde am 02.02.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 08.06.2016 angenommen.

to Renate and Manfred Müller

Abstract

In this thesis, we develop and investigate numerical methods for solving nonsmooth optimization problems and generalized variational inequalities. A proximal-type fixed point equation representing the optimality or stationarity conditions forms the basis of the different approaches. The algorithmic framework we focus on uses semismooth Newton steps for the fixed point equation to enhance an underlying globally convergent descent method. We present both global and local convergence results and derive an abstract second order theory that can be used to characterize and to verify the conditions for local convergence. We conclude with numerical examples demonstrating the efficiency of the proposed methods.

Zusammenfassung

Diese Arbeit befasst sich mit der Entwicklung und Untersuchung numerischer Verfahren zur Lösung nichtglatter Probleme und verallgemeinerter Variationsungleichungen. Die verschiedenen Verfahrensansätze basieren auf einer Reformulierung der Optimalitäts- oder Stationaritätsbedingungen als proximale Fixpunktgleichung. Im Fokus steht die Verwendung eines semiglatten Newton-Verfahrens, welches ein zugrunde liegendes, global konvergentes Abstiegsverfahren erweitern und beschleunigen soll. Globale und lokale Konvergenzresultate werden präsentiert und eine abstrakte Optimalitätstheorie zweiter Ordnung wird hergeleitet, die zur Sicherstellung und Überprüfung schneller, lokaler Konvergenz angewendet werden kann. Numerische Experimente belegen die Effektivität der vorgestellten Verfahren.

Acknowledgements

The completion of this thesis would not have been possible without the help of several people.

First of all, I want to thank my advisor Prof. Dr. Michael Ulbrich for his guidance and constant support and for introducing me to the broad and interesting field of nonsmooth optimization. His valuable comments and constructive suggestions contributed significantly to the success of this work.

I gratefully acknowledge the support of the TopMath program, a graduate program of the Elite Network of Bavaria, which partially supported the first phase of my PhD project.

I am grateful to many colleagues at the TUM. Especially, I want to thank Karin Tichmann and Konstantin Pieper for many fruitful discussions on the proximity operator, on image processing problems, and on abstract second order conditions. I am also grateful to Simon Plazotta for discussions and suggestions related to this work. Furthermore, I want to thank all my former and current colleagues at M1 and M17 for providing such a great working atmosphere. In particular, I would like to thank Christian, Dennis, Florian, Florian, Ira, Johannes, Martin, Moritz, Sebastian, and Sebastian for their support throughout the last years.

Finally, I want to thank my friends and my family for their patience, their constant encouragement and never-ending support. Thank you so much!

Notations

Sets and operations on sets:

$\emptyset$	the empty set
$\mathbb{N}$	set of natural numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$, $\mathbb{R}_{++}$	the sets of nonnegative and positive real numbers
$(-\infty, +\infty], [-\infty, +\infty]$	the sets of real extended numbers
$B_\varepsilon(x)$	open ball with radius $\varepsilon > 0$ around x w.r.t. the Euclidean norm
$B_{\vert\vert\vert\cdot\vert\vert\vert}(x, \varepsilon)$	open ball with radius $\varepsilon > 0$ around x w.r.t. the norm $\vert\vert\vert\cdot\vert\vert\vert$, $B_{\vert\vert\vert\cdot\vert\vert\vert}(x, \varepsilon) := \{y : \vert\vert\vert y - x\vert\vert\vert < \varepsilon\}$
aff S	affine hull of the set $S \subset \mathbb{R}^n$
conv S	convex hull of S
lin S	lineality space of S
sp S, sp$\{x\}$	linear span of S and $\{x\}$
S°	polar cone of S
$S^\perp$, $\{x\}^\perp$	orthogonal complement of S and $\{x\}$
cl S, $\bar{S}$	closure of S
int S	interior of S
ri S	relative interior of S, i.e., ri $S := \{x \in S : \exists\ \varepsilon > 0$ such that $B_\varepsilon(x) \cap \text{aff } S \subset S\}$
$\mathcal{R}_S(x)$	radial cone of S at $x \in \mathbb{R}^n$
$T_S(x)$	contingent tangent cone of S at x
$T_S^i(x)$	inner tangent cone of S at x
$N_S(x)$	normal cone of the set S

Matrices:

$\mathbb{S}^n$	set of symmetric, real $n \times n$ matrices
$\mathbb{S}^n_+$	set of symmetric, real, positive semidefinite $n \times n$ matrices
$\mathbb{S}^n_{++}$	set of symmetric, real, and positive definite $n \times n$ matrices
$\lambda_{\min}(M)$, $\lambda_{\max}(M)$	smallest and largest eigenvalue of a symmetric matrix M
$\kappa(M)$	condition number of a matrix M

$M_{[\mathcal{I}\mathcal{J}]}$	submatrix of a matrix $M \in \mathbb{R}^{m\times n}$ w.r.t. the index sets $\mathcal{I} \subset \{1, ..., m\}$, $\mathcal{J} \subset \{1, ..., m\}$
$M_{[i\cdot]}$, $M_{[\cdot j]}$	i-th row and j-th column of M
$\mathrm{diag}(x)$	diagonal matrix with entries $\mathrm{diag}(x)_{[ii]} = x_i$, $i = 1, ..., n$
$\mathrm{sym}(M)$	symmetric part of M, i.e., $\mathrm{sym}(M) = \frac{1}{2}(M + M^\top)$

Operations on vectors and matrices:

$\langle\cdot,\cdot\rangle$	Euclidean inner product, i.e., $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$
$\|\cdot\|$	Euclidean norm, i.e., $\|x\|^2 = \langle x, x\rangle$
$\mathrm{dist}(x, S)$	distance between the point x and the set S, i.e., $\mathrm{dist}(x, S) = \inf_{y\in S}\ \|x - y\|$
$\langle\cdot,\cdot\rangle_\Lambda$, $\|\cdot\|_\Lambda$,	induced Λ-scalar product and Λ-norm; for $\Lambda \in \mathbb{S}^n_{++}$ it holds $\langle x, y\rangle_\Lambda := \langle x, \Lambda y\rangle, \quad \|x\|_\Lambda := \sqrt{\langle x, x\rangle_\Lambda}$
$\mathrm{tr}(\cdot)$	trace of a square matrix
$\langle\cdot,\cdot\rangle_F$, $\|\cdot\|_F$,	Frobenius inner product and Frobenius-norm of a matrix (we use $\langle\cdot,\cdot\rangle \equiv \langle\cdot,\cdot\rangle_F$, $\|\cdot\| \equiv \|\cdot\|_F$ if the context is clear)
$\succ$, $\succeq$	partial ordering on the space $\mathbb{S}^n$, i.e., $B \succ A \ :\Leftrightarrow\ B - A \in \mathbb{S}^n_{++}, \quad B \succeq A \ :\Leftrightarrow\ B - A \in \mathbb{S}^n_{+}$
$\odot$	Hadamard product, i.e., $(x \odot y)_i = x_i \cdot y_i$
$\oslash$	component-wise division, i.e., $(x \oslash y)_i = x_i / y_i$

In the following, let $\varphi : \mathbb{R}^n \to [-\infty, +\infty]$, $f : \mathbb{R}^n \to \mathbb{R}$, and $F : \mathbb{R}^n \to \mathbb{R}^m$ be arbitrary functions.

Functions and operations on functions:

$\mathrm{dom}\ \varphi$	effective domain of the function φ
$\mathrm{epi}\ \varphi$	epigraph of φ
$\mathrm{lev}_\alpha\ \varphi$	lower level set of φ at level $\alpha \in \mathbb{R}$
$\mathrm{lin}\ \varphi$	lineality space of φ
$\mathrm{gra}\ \Phi$	graph of a multifunction $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, i.e., $\mathrm{gra}\ \Phi := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : y \in \Phi(x)\}$
$F^{-1}(\cdot)$	the inverse multifunction $F^{-1} : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, $F^{-1}(y) := \{x \in \mathbb{R}^n : F(x) = y\}$
$\iota_S(\cdot)$	indicator function of the set S
$\sigma_S(\cdot)$	support function of S
$\mathrm{prox}^\Lambda_\varphi(\cdot)$	proximity operator of φ with parameter matrix $\Lambda \in \mathbb{S}^n_{++}$
$\mathrm{env}^\Lambda_\varphi(\cdot)$	Moreau envelope of φ with parameter matrix $\Lambda \in \mathbb{S}^n_{++}$
$\mathcal{P}^\Lambda_S(x)$	projection of x onto the set S w.r.t. the Λ-norm

Derivatives and subdifferentials:

$\varphi'_-(x;h)$, $\varphi'_+(x;h)$	lower and upper directional derivatives of φ at x in the direction h
$\varphi'(x;h)$, $F'(x;h)$	directional derivate of φ, F at x in the direction h
$\varphi^{\downarrow}_-(x;h)$, $\varphi^{\downarrow}_+(x;h)$	lower and upper directional epiderivatives of φ at x in the direction h
$\varphi^{\downarrow}(x;h)$	directional epiderivative of φ at x in the direction h
$\varphi''_-(x;h,w)$, $\varphi''_+(x;h,w)$	lower and upper (parabolic) second order directional derivatives of φ at x in the directions h, w
$\varphi''(x;h,w)$	(parabolic) second order directional derivative of φ at x in the directions h, w
$\varphi^{\downarrow\downarrow}_-(x;h,w)$, $\varphi^{\downarrow\downarrow}_+(x;h,w)$	lower and upper (parabolic) second order directional epiderivatives of φ at x in the directions h, w
$\varphi^{\downarrow\downarrow}(x;h,w)$	(parabolic) second order directional epiderivative of φ at x in the directions h, w
$\mathrm{d}^2\varphi(x\|h)(w)$	second order subderivative of φ at x relative to h in the direction w
$\nabla f(x)$, $\nabla^2 f(x)$	the gradient and the Hessian of f at x
$DF(x)$, $D^2F(x)$	first and second order Fréchet derivative of F at x; for $h \in \mathbb{R}^n$ it holds $D^2F(x)[h,h] \in \mathbb{R}^m$ and $D^2F(x)[h,h] = (h^\top \nabla^2 F_1(x)h, ..., h^\top \nabla^2 F_m(x)h)^\top$
$\partial_B F(x)$	Bouligand subdifferential of F at x
$\partial F(x)$	Clarke subdifferential of F at x
$\partial_C F(x)$	C-subdifferential of F at x

Contents

1. Introduction

In this thesis, we consider and investigate efficient numerical algorithms for solving general nonsmooth optimization problems of the form

$$(\mathcal{P}) \qquad \min_{x \in \mathbb{R}^n} \; f(x) + \varphi(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable, possibly nonconvex function and $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous mapping.

The algorithmic framework we focus on is primarily based on the idea to use semismooth Newton steps for a proximal-type reformulation of the corresponding first order optimality conditions of the problem $(\mathcal{P})$,

$$(\mathcal{E}) \qquad F^{\Lambda}(x) = x - \mathrm{prox}_{\varphi}^{\Lambda}(x - \Lambda^{-1}\nabla f(x)) = 0, \quad \Lambda \in \mathbb{S}^n_{++},$$

to augment and accelerate an underlying globally convergent descent method. Unlike other or more common globalization strategies, we utilize a multidimensional filter mechanism to monitor the acceptance of the semismooth Newton steps and to achieve both global and local fast convergence. The presented approach can be naturally extended to solve generalized variational inequalities of the type

find $x \in G^{-1}(\mathrm{dom}\ \varphi)$ *such that*

$$(\mathcal{P}_{\mathrm{vip}}) \qquad \langle F(x), y - G(x)\rangle + \varphi(y) - \varphi(G(x)) \geq 0, \quad \forall\, y \in \mathbb{R}^n,$$

where the functions $F : \mathbb{R}^n \to \mathbb{R}^n$ and $G : \mathbb{R}^n \to \mathbb{R}^n$ are typically supposed to be continuously differentiable on an open set that contains the domain $G^{-1}(\mathrm{dom}\ \varphi)$. In one of its simplest forms,

$$F(x) := \nabla f(x), \quad G(x) := x,$$

the generalized variational inequality $(\mathcal{P}_{\mathrm{vip}})$ coincides with an alternative, variational-based representation of the optimality conditions of problem $(\mathcal{P})$ illustrating the deep connection between the two problems $(\mathcal{P})$ and $(\mathcal{P}_{\mathrm{vip}})$.

In the following work, we provide a detailed convergence theory for the different described methods and for both the convex composite problem $(\mathcal{P})$ and the generalized variational inequality problem $(\mathcal{P}_{\mathrm{vip}})$. In contrast to many other convergence analyses, our theory will cover both convex and nonconvex situations. Moreover, to the best of our knowledge, convergence results for semismooth Newton-type methods and for generalized variational inequalites $(\mathcal{P}_{\mathrm{vip}})$ seem to be only available in the context of classical variational inequalities, where the function G again corresponds to the identity mapping and the nonsmooth mapping φ is chosen as a specific indicator function of a convex, nonempty, and closed set [121, 120, 227, 76].

Besides, a strong emphasis is also put on the investigation and derivation of second order conditions that allow a rather elegant verification and characterization of the conditions for local fast convergence and that can be used to analyze the behavior and local structure of stationary points and of the minimization problem ($\mathcal{P}$). More specifically, by focusing on the class of so-called decomposable problems, we derive a new representation of the curvature that is induced by the nonsmooth function φ and show that it can be intrinsically described by the Fréchet derivative of the proximity operator $\mathrm{prox}_\varphi^\Lambda$. We show that this abstract second order framework can be applied to a large variety of nonsmooth optimization problems and that it can be systematically extended to more general classes of composite problems.

In recent years, composite-type problems of the form ($\mathcal{P}$) that consist of the sum of a differentiable and a nonsmooth, but often simpler function have become a quite popular and ubiquitous tool to model various practical problems and applications, such as, e.g., signal and image processing problems, matrix completion problems, or feature selection and classification problems in machine learning. Typically, the smooth function f is utilized as a loss function or a data-fitting term to express the difference between estimated values of data and given measurements. On the other hand, the nonsmooth function φ is often chosen as a highly specialized regularization term to induce a certain desired structure on the iterates of a method and on the solutions of the problem ($\mathcal{P}$). In particular, this is the case for, e.g., ℓ_1-, group sparse, or nuclear norm regularizations where a particularly *sparse* or parsimonious representation of the data is sought. In the following, we provide several practical examples demonstrating the broad conceptual applicability and importance of the problem ($\mathcal{P}$).

Compressive Sensing and ℓ_1-minimization. In the last decade, there has been a considerable interest in ℓ_1-regularized minimization problems of the form

$$\min_{x\in\mathbb{R}^n} \; f(x) + \mu\,\|x\|_1,$$

which can be primarily traced back to the sparsity promoting properties of the ℓ_1-norm. The remarkably universal role and the computational attractiveness of sparse solutions stems from the fact that in many applications, such as signal or image processing, there exist canonical sparse representations of the relevant data. In the convex quadratic case $f(x) := \frac{1}{2}\|Ax-b\|_2^2$, the ℓ_1-problem reduces to an ℓ_1-regularized least squares problem that is closely related to the so-called *basis pursuit problem*:

$$\min_{x\in\mathbb{R}^n} \; \|x\|_1 \quad \text{s.t.} \quad Ax = b.$$

It is known that the basis pursuit problem can be interpreted as a convex relaxation of the NP-hard problem of finding the sparsest solution of the (in general underdetermined) system $Ax = b$:

$$\min_{x\in\mathbb{R}^n} \; \|x\|_0 \quad \text{s.t.} \quad Ax = b.$$

Here, the ℓ_0-quasi norm counts the number of nonzeros in x, i.e., $\|x\|_0 := |\{i : x_i \neq 0\}|$. Under appropriate assumptions on the matrix A and on the sparsity of a solution $\bar{x}$, the solutions of the latter problems coincide and the computationally more tractable ℓ_1-basis pursuit problem can be used to reconstruct the signal or solution $\bar{x}$ from far less than n measurements. This fundamental principle is also known as *compressed sensing* or *compressive sensing.* Details and further information can be found in Candès, Romberg, Tao [35, 36, 37] and Donoho [66]. Compressive sensing significantly extended the class of existing data and signal acquisition methods and has been used in a broad variety of fields, such as compressive imaging [69, 248], magnetic resonance and computed tomography imaging [141, 142], seismics [137], or communication [9]. Other recent applications comprise logistic regression [170, 123, 221] or Laplacian interpolation-based image compression [111, 47].

Group and joint sparsity. In contrast to the ℓ_1-regularization, *group* or *joint sparse* penalty terms allow to incorporate and utilize more specific information about the structure of the sparsity pattern of a solution. In particular, not only single components of a solution $\bar{x}$ are required to be zero, but whole groups and clusters of components can be modeled to be zero. The basic ℓ_2-ℓ_1 group sparse problem can be written in the form

$$\min_{x \in \mathbb{R}^n} \; f(x) + \mu \sum_{i=1}^{s} \|x_{g_i}\|_2,$$

where the groups $g_i \subset \{1, ..., n\}$ are usually chosen as a disjoint partitioning of the set $\{1, ..., n\}$ but are also allowed to overlap in certain situations. Additionally, if the smooth loss function f is a quadratic mapping, then the latter problem is referred to as the group Lasso [262]. This specific type of problem has been used extensively in statistics and machine learning [262, 7, 115], or biomedical applications [144, 166] throughout the last years. Furthermore, if the optimization variable is a matrix and if the groups g_i, $i = 1, ..., s$, correspond to the different rows of the matrix variable, then the group sparse problem is called *multiple measurement vector* (MMV) *problem* [59, 143, 45, 230, 243]. Let us note that for MMV problems other norm constellations or more general group penalizations have been considered in the literature, see [85, 230, 116]. In infinite dimensions, a similar problem was analyzed by Herzog et al. [106, 39] in an optimal control setting. Here, a *directional sparsity* term was applied to obtain controls with a striped sparsity pattern.

Other convex composite problems arise in the context of image deconvolution and total variation minimization [209, 171, 41, 249, 259], semidefinite programming, or recommender systems and low rank matrix completion [34, 125, 33, 38, 201]. Of course, our brief discussion is rather incomplete; for a more detailed overview of convex composite problems and different nonsmooth penalty terms, we refer to Bach et al. [6]. At this point, let us also highlight that most of the mentioned problems are large scale.

Related work

The steadily growing interest in composite objective functions and nonsmooth optimization problems of the form $(\mathcal{P})$ has initiated the development and investigation of many different, numerical methods. Their applicability ranges from very general settings to highly specific problem formulations, where the functions f and φ have a fixed form. Moreover and in contrast to our proposed globalized semismooth Newton method, a large class of these algorithms is centered on the usage of first order gradient-based information and a strong focus so far has been on the case where the mapping f is convex. In the following paragraph, we give a brief overview of several approaches that can be used to solve the general nonsmooth optimization problem ($\mathcal{P}$).

First, the optimality condition ($\mathcal{E}$) immediately leads to the simple and basic fixed point iteration scheme

$$(\mathcal{F}_p) \qquad x^{k+1} = \mathrm{prox}_{\varphi}^{\Lambda}(x^k - \Lambda^{-1}\nabla f(x^k)), \quad k = 0, 1, ...$$

If the mapping f is convex and if the parameter matrix $\Lambda \in \mathbb{S}^n_{++}$ is chosen via $\Lambda = \tau^{-1}I$ for some $\tau > 0$, then this iterative procedure is usually referred to as the classical, basic *proximal gradient* or *forward-backward splitting method* and convergence to a fixed point can be achieved under suitable assumptions on the step size τ and on the gradient ∇f. The proximal gradient method forms the basis of a broad variety of approaches. In particular, different variants and extensions of this basic method were analyzed by Combettes, Pesquet, and Wajs in [54, 52]. Fukushima and Mine [88] studied another variant for nonconvex f and used an additional line search technique to establish global convergence. Tseng and Yun [236] refined and extended the theory presented in [88] and developed a general block coordinate descent method for nonsmooth problems of the type ($\mathcal{P}$) with block-separable structure. In [200, 265], Qin et al. and Yun et al. proposed several related coordinate descent schemes for ℓ_1- and group sparse problems. The method SpaRSA [256] uses a nonmonotone line search technique and (adaptive) Barzilai-Borwein step sizes to accelerate the underlying proximal gradient method. Recently, Ochs et al. [178] presented and analyzed the algorithm “iPiano” that is an extension of the basic proximal gradient scheme for nonconvex problems. Here, motivated by Polyak’s Heavy ball technique, a so-called *inertial term* is added to account for the nonconcexity of f and to enhance the performance of the approach.

A slightly different class of algorithms utilizes two- or multi-stage schemes and extrapolation steps to improve the performance of the underlying proximal gradient method. This class includes Nesterov’s well-known accelerated gradient methods [168, 169], TwIST [19], or the fast iterative thresholding algorithm FISTA [12]. Further methodologies comprise variants of the alternating direction method of multipliers (ADMM) [91, 71, 93] and the primal-dual algorithms [42, 192, 105]. While the latter algorithms can be usually applied to more general problems where both f and φ may be nondifferentiable, convergence is often only achievable under certain convexity assumptions.

Becker, Fadili, and Lee et al. [14, 129] proposed an inexact proximal Newton-type method to solve *convex* problems of the form ($\mathcal{P}$). The method uses the Hessian of the differentiable function f or a suitable approximation as a parameter matrix,

$$\Lambda \approx \nabla^2 f(x^k),$$

to accelerate the base algorithm ($\mathcal{F}_p$). In [232], a combination of proximal Newton-type steps and an interior point framework for constrained convex problems is considered. Here, the parameter matrix Λ models and approximates the Hessian of an appropriately chosen (self-concordant) barrier function. In [222] a proximal Newton-type method was used within a stochastic framework for machine learning problems.

For a more detailed discussion of proximal-based methods we refer to the surveys [6, 53, 184] and the references therein.

Patrinos, Stella, and Bempora [187] investigate a semismooth Newton method to solve a *forward-backward*-based reformulation of the nonsmooth equation ($\mathcal{E}$). The key idea is to multiply the function F^Λ with the matrix $\Lambda(I - \Lambda^{-1}\nabla^2 f(x))$. The resulting function then corresponds to the gradient of the so-called *forward-backward envelope* which can be used as a merit function and to globalize the semismooth Newton method if the mapping f is strongly convex. Although this approach is certainly based on similar algorithmic ideas, it also significantly differs from our proposed semismooth Newton method in that its convergence theory is limited to strongly convex problems. Several other and more specialized semismooth Newton methods have been proposed in the context of ℓ_1- and matrix minimization problems. In particular, Byrd et al. [32] present and discuss a family of semismooth Newton methods for convex ℓ_1-regularized problems that represents different realizations of the semismooth Newton method and incorporates block active set and orthant-based methods, respectively. In [268, 138, 117, 119, 260], exploiting the local fast convergence of the semismooth Newton method, different inexact Newton-type approaches are investigated to approximately solve inner subproblems of an augmented Lagrangian and a proximal point method for semidefinite programming and nuclear norm problems.

In infinite dimensions, semismooth Newton methods have been successfully applied to a variety of nonsmooth composite problems. For instance, Griesse and Lorenz [97] considered a semismooth Newton method for ℓ_1-minimization in the Hilbert space ℓ_2. Stadler [225] analyzed a local semismooth Newton method for elliptic optimal control problems with an L^1-cost functional. Moreover, Hans and Raasch [99] developed a globally convergent, damped B-semismooth Newton method for ℓ_1-Tikhonov regularized quadratic problems. Based on the so-called normal map, Pieper [191] investigated a globalized semismooth Newton framework for elliptic and parabolic optimal control problems with an abstract and general regularization term φ. Further applications of the local semismooth Newton method can be found, e.g., in [65, 106].

For more details on generalized variational inequalities and on their applications, we refer to chapter 6.

Organization and contribution

This thesis is structured as follows. In chapter 2, we provide various definitions, properties, and concepts from convex and nonsmooth analysis that will form the mathematical basis of our investigations. Since the nonsmooth function φ can be real extended valued in general (for instance, it can be chosen as an indicator function to model convex constraints), we will require an appropriate notion of (directional) differentiability to cope with this situation. In particular, we will see that the classical directional derivative $\varphi'(x; h)$ may no longer define a lower semicontinuous mapping with respect to the considered direction $h \in \mathbb{R}^n$ and that several important properties will not hold in such general case. Thus, a focus is set on the introduction and investigation of so-called (directional) epiderivatives which are based on Painlevé-Kuratowski- or Γ-convergence processes and appropriately generalize and extend the directional derivative $\varphi'(x; h)$. Moreover, since the mapping φ is usually assumed to be convex, these epiderivatives often enjoy a rich calculus and can be connected to the common convex subdifferential. Details can be found in section 2.4 and 2.5. Finally, we also present basic properties of the Clarke subdifferential for vector valued functions and discuss the concept of semismoothness.

Chapter 3 is concerned with a detailed analysis of the proximity operator $\mathrm{prox}_\varphi^\Lambda$ and summarizes its most relevant properties. While many other works concentrate on the classical definition of the proximity operator with a fixed parameter matrix of the form

$$\Lambda = \tau^{-1} I, \quad \tau > 0,$$

we will consider the more general case and investigate the dependence of the proximity operator on the matrix Λ in some more detail. Since the proximity operator cannot be expected to be semismooth in general, we also present and discuss a specific class of functions φ for which semismoothness of the proximity operator $\mathrm{prox}_\varphi^\Lambda$ can be guaranteed. The corresponding results were first derived by Bolte et al. [21] and are provided in section 3.3. Furthermore, as preparation for chapter 5, we derive several second order properties which are mainly immediate consequences of the differentiability and convexity of the Moreau envelope.

In chapter 4, we propose and analyze the globalized semismooth Newton method in detail. At first, we introduce different equivalent first order optimality conditions and derive the nonsmooth mapping F^Λ. Then, based on [88, 236], we discuss the proximal gradient descent method that will be used as an underlying base algorithm and analyze its global convergence properties. Furthermore, the filter mechanism and the filter acceptance test are presented in detail. Our main contribution in this chapter is the development of a general global and local convergence theory for the proposed approach. In particular, we will verify that the filter globalization guarantees global convergence in the sense that every accumulation point of a generated sequence of iterates is a stationary point of the problem $(\mathcal{P})$. Moreover, transition to fast local convergence is shown under rather mild and standard assumptions. Here, the main requirements are:

- Semismoothness of the proximity operator $\mathrm{prox}_\varphi^\Lambda$.
- Uniformly bounded invertibility of the generalized derivatives of the nonsmooth mapping F^Λ in a neighborhood of an accumulation point.
- Existence of an accumulation point that is a strict local minimum and an isolated stationary point of the problem $(\mathcal{P})$.

Additionally, if the function f is convex, we will also present and investigate a simpler globalization strategy and prove its efficiency. Let us note that chapter 4 is essentially based on Milzarek and Ulbrich [157], where a similar algorithmic framework has been analyzed for ℓ_1-regularized problems. In this thesis, we generalize and extend the results of [157], which were established for ℓ_1-optimization problems, to the convex composite setting.

Chapter 5 is dedicated to the second order analysis of problem $(\mathcal{P})$. Our overall goal in this chapter is to rephrase the conditions for local convergence of the globalized semismooth Newton method as suitable second order conditions. Based on an abstract and profound second order framework developed by Bonnans, Cominetti, and Shapiro [23, 24, 27] and concentrating on the class of so-called decomposable problems, we derive a pair of no gap second order conditions that ensure isolated stationarity and local optimality of a stationary point of $(\mathcal{P})$. The concept of decomposability was first proposed by Shapiro [217] and is closely related to the concept of cone-reducible sets in constrained optimization. The considered class of decomposable problems comprises a large variety of interesting and important examples, such as polyhedral problems, group sparse problems, nonlinear programming, total variation imaging, or nuclear norm-regularized problems. For more details we refer to section 5.3. Inspired by [212], we show that the strict complementarity condition can be used to characterize differentiability of the proximity operator $\mathrm{prox}_{\varphi}^{\Lambda}$ if the underlying function φ is decomposable. Furthermore, as already mentioned and as one of our main results, we derive a new representation of the curvature induced by the nonsmooth function φ in terms of the Moore-Penrose inverse of the Fréchet derivative of the proximity operator $\mathrm{prox}_{\varphi}^{\Lambda}$. This formulation is then used to prove that the strict complementarity condition and the second order sufficient conditions imply invertibility of all generalized derivatives of the nonsmooth mapping F^{Λ} at a stationary point. Finally, in section 5.5, we extend these different results and a second order framework for more general composite problems is presented.

In chapter 6, we show that the described algorithmic prototype can also be utilized to solve generalized variational inequalities problems of the form $(\mathcal{P}_{\mathrm{vip}})$. In this case, the underlying proximal gradient method is substituted by a D-gap function-based descent method that solves an optimization-based reformulation of the problem $(\mathcal{P}_{\mathrm{vip}})$. In section 6.1 and 6.2, we briefly discuss conditions that ensure existence of a solution of problem $(\mathcal{P}_{\mathrm{vip}})$ and summarize the main properties of the D-gap function and of the so-called regularized gap function. Specifically, we derive several new stationarity results that guarantee global optimality of a stationary point of the regularized gap or the D-gap function. Similar to chapter 4, a strong focus is set on the development of a detailed and general convergence theory for the proposed Newton-type method.

Finally, in chapter 7 we present extensive numerical results for the globalized semismooth Newton method that was introduced and analyzed in chapter 4. In particular, the performance of the method is investigated on convex and nonconvex ℓ_1-regularized least squares problems and on group sparse optimization problems. We will focus on large scale experiments where the application of the Hessian of f is only available as a matrix-free operation and compare our algorithm with different state-of-the-art methods.

2. Convex and nonsmooth analysis

In this chapter, we state and collect basic definitions, properties, and various concepts from convex and nonsmooth analysis that will be used repeatedly throughout this thesis.

The next sections are organized as follows. At first, we recall some elementary definitions, specify notational aspects and discuss helpful properties of tangent cones and sublinear functions. Thereafter, we present several important, theoretical frameworks and results from convex and variational analysis, such as, e.g., convex conjugation, multifunctions, the convex subdifferential, and subdifferential calculus. Moreover, in section 2.4 and 2.5.1, we briefly introduce the concepts of epi-convergence and directional epidifferentiability. Directional epiderivatives are an extension of the classical directional derivatives and turn out to be the right tool to study real extended valued functions. Here, since we want to consider classes of optimization problems that generally allow extended valued objective functionals (to model, e.g., convex constraints), we will need these (epi-)concepts at different parts of this thesis. For instance, the epi-calculus presented in section 2.5 will be utilized in section 5 to derive and discuss general first and second order optimality conditions. Finally, in section 2.5.3 and 2.6, we present Clarke's generalized subdifferential and the concept of semismoothness for possibly nonconvex and nonsmooth functions.

Most of the material that is provided here can be found in the monographs [208, 27, 11]. Furthermore, the overall structure of this introductory chapter is essentially based on [27, Chapter 2]. The work of Bonnans and Shapiro [27] also includes a broader and deeper introduction and discussion of the different topics that will be presented in the following passages. For a more detailed introduction to convex analysis let us refer to [109, 11]. A quite advanced, systematic study of various subjects and developments in nonsmooth and variational analysis can be found in the book of Rockafellar and Wets [208]. For more information on Clarke's subdifferential see also [50].

2.1. Preliminary definitions and tangent cones

2.1.1. Basics and semicontinuity

Let us start with some elementary definitions.

Definition 2.1.1. *Let $\varphi : \mathbb{R}^n \to [-\infty, +\infty]$ be a functional. The* (effective) domain *of φ is defined by*

$$\text{dom } \varphi := \{x \in \mathbb{R}^n : \varphi(x) < +\infty\}.$$

The epigraph *of φ is*

$$\text{epi } \varphi := \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : \varphi(x) \le t\} \subset \mathbb{R}^n \times \mathbb{R}.$$

The lower level set *of φ at level $\alpha \in \mathbb{R}$ is given by*

$$\operatorname{lev}_\alpha \varphi := \{x \in \mathbb{R}^n : \varphi(x) \le \alpha\}.$$

The function φ is called proper *if $\varphi(x) \neq -\infty$ for all $x \in \mathbb{R}^n$ and* dom $\varphi \neq \emptyset$.

Definition 2.1.2. *The function $\varphi : \mathbb{R}^n \to [-\infty, +\infty]$ is said to be* lower semicontinuous *at a point $x \in \mathbb{R}^n$ if*

$$\liminf_{\tilde{x} \to x} \varphi(\tilde{x}) \ge \varphi(x), \quad \textit{or equivalently} \quad \liminf_{\tilde{x} \to x} \varphi(\tilde{x}) = \varphi(x).$$

We say that φ is lower semicontinuous, *if φ is lower semicontinuous at every $x \in \mathbb{R}^n$.*

The following Lemma shows that lower semicontinuity of a function φ can be completely characterized by closedness of its corresponding epigraph epi φ.

Lemma 2.1.3 (cf. [11, Lemma 1.24]). *Let $\varphi : \mathbb{R}^n \to [-\infty, +\infty]$ be given. Then, the following statements are equivalent:*

(i) *The function φ is lower semicontinuous.*

(ii) *The epigraph* epi φ *is closed in $\mathbb{R}^n \times \mathbb{R}$.*

(iii) *For every $\alpha \in \mathbb{R}$, the level set* $\operatorname{lev}_\alpha \varphi$ *is closed in $\mathbb{R}^n$.*

2.1.2. Tangent cones

A nonempty set $S \subset \mathbb{R}^n$ is called *cone* if we have $tx \in S$ for all $t \ge 0$ and any $x \in S$. The *polar cone* of a set $S \subset \mathbb{R}^n$ is defined via

$$S^\circ := \{x \in \mathbb{R}^n : \langle x, y \rangle \le 0, \ \forall \ y \in S\}.$$

Clearly, the polar cone is always a closed, convex cone. In this thesis, we will work with the following, different tangent cones.

Definition 2.1.4 (Tangent cones, cf. [27, Definition 2.54]). *Let $S \subset \mathbb{R}^n$ and $x \in S$ be a given set and a given point. The* radial cone *of S at x is defined by*

$$\mathcal{R}_S(x) := \{d \in \mathbb{R}^n : \exists \ t_* > 0, \ \forall \ t \in [0, t_*], \ x + td \in S\}.$$

Furthermore, the sets

$$\begin{aligned} T_S(x) &:= \{d \in \mathbb{R}^n : \exists \ t_k \downarrow 0, \ \operatorname{dist}(x + t_k d, S) = o(t_k)\}, \\ T_S^i(x) &:= \{d \in \mathbb{R}^n : \operatorname{dist}(x + td, S) = o(t), \ t \ge 0\} \end{aligned}$$

are called the contingent (Bouligand) cone *and the* inner tangent cone, *respectively.*

Of course, it immediately follows from the latter definition that the sets $\mathcal{R}_S(x)$, $T_S(x)$, and $T_S^i(x)$ are cones. Moreover, if S is a convex, closed set and $x \in S$, then it holds

$$\mathcal{R}_S(x) = \bigcup_{t>0} \{t^{-1}(S - x)\} = \mathbb{R}_+(S - x) \quad \text{and} \quad T_S(x) = T_S^i(x) = \operatorname{cl} \mathcal{R}_S(x). \tag{2.1.1}$$

Thus, in this situation, the contingent and the inner tangent cone coincide (see [27, Proposition 2.55] for details). Since the closure of a convex set is convex, (2.1.1) also implies that the cones $T_S(x)$ and $T_S^i(x)$ are convex sets in this case.

The polar cone of the contingent cone $T_S(x)$ is the so-called *normal cone* to S at x and we will write $N_S(x) := T_S(x)^\circ$. If the set S is convex, then we obtain $N_S(x) = \mathcal{R}_S(x)^\circ$. Moreover, in this situation, using the representation of the radial cone, $N_S(x)$ can be expressed as follows

$$N_S(x) := \{v \in \mathbb{R}^n : \langle v, y - x\rangle \leq 0, \; \forall \; y \in S\}.$$

If $x \notin S$, we will use the convention $N_S(x) := \emptyset$. Let us conclude with an important example.

Example 2.1.5 (cf. [27, Example 2.62]). Let $S \subset \mathbb{R}^n$ be a convex, closed cone and let $x \in S$ be arbitrary. Then, it follows $\mathcal{R}_S(x) = S + \mathrm{sp}\{x\}$ and, by applying [11, Proposition 6.26], we obtain

$$N_S(x) = \mathcal{R}_S(x)^\circ = [S + \mathrm{sp}\{x\}]^\circ = S^\circ \cap [\mathrm{sp}\{x\}]^\circ = S^\circ \cap \{x\}^\perp.$$

In particular, if $K \subset \mathbb{R}^n$ is a convex, closed set and $x \in K$, $y \in N_K(x)$ are arbitrary points, then this implies

$$N_{N_K(x)}(y) = N_K(x)^\circ \cap \{y\}^\perp = T_K(x) \cap \{y\}^\perp. \tag{2.1.2}$$

2.1.3. Sublinear functions and support functions

In the following, we list several relevant definitions and properties of sublinear and support functions. The results, which will be presented here, are essentially taken from [109]. Rigorous proofs and a more detailed overview of sublinear or support functions can be found in [109, Chapter C].

Definition 2.1.6. *A convex and positively homogeneous mapping $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is called a* sublinear function. *If φ is sublinear, then the* lineality space *of φ is defined as the linear subspace*

$$\mathrm{lin}\;\varphi := \{x \in \mathbb{R}^n : \varphi(x) + \varphi(-x) = 0\}.$$

Now, suppose that $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is proper. Then, it can be easily shown that φ is sublinear if and only if it is positively homogeneous and *subadditive*, i.e.,

$$\varphi(x + y) \leq \varphi(x) + \varphi(y), \quad \forall \; x, y \in \mathbb{R}^n.$$

Next, we provide a definition of the support function of a set and collect some basic properties of support functions.

Definition 2.1.7. *Let $S \subseteq \mathbb{R}^n$ be a nonempty set. The function $\sigma_S : \mathbb{R}^n \to (-\infty, +\infty]$ defined by*

$$x \mapsto \sigma_S(x) := \sup_{s \in S} \; \langle s, x\rangle$$

is called the support function *of the set S.*

Lemma 2.1.8. *The support function has the following properties:*

(i) *Let $S \subset \mathbb{R}^n$ be nonempty. Then, the support function σ_S is a convex, proper, lower semicontinuous, and positively homogeneous function.*

(ii) *Let $S_1, S_2 \subset \mathbb{R}^n$ be two convex, closed sets. The set S_1 is a subset of S_2 if and only if we have $\sigma_{S_1}(x) \leq \sigma_{S_2}(x)$ for all $x \in \mathbb{R}^n$. Moreover, in this case, it follows* $\text{dom } \sigma_{S_2} \subset \text{dom } \sigma_{S_1}$.

(iii) *Let $S \subset \mathbb{R}^n$ be a nonempty cone, then it holds* $\text{dom } \sigma_S \subset S^\circ$.

(iv) *The support function of a set $S \subset \mathbb{R}^n$ is finite everywhere if and only if S is bounded.*

Proof. The first two parts are elementary and will be omitted. (See, e.g., [109, Proposition C-2.1.2 and Theorem C-3.3.1]). We continue with a proof of statement (iii). Therefore, let $\bar{x} \in \text{dom } \sigma_S$ be arbitrary and assume that $\bar{x} \notin S^\circ$. Then, there exist $y \in S$ and $\varepsilon > 0$ such that $\langle \bar{x}, y \rangle > \varepsilon$. Since S is a cone, it follows $ty \in S$ for all $t \geq 0$. This shows

$$\sigma_S(\bar{x}) = \sup_{x \in S} \langle \bar{x}, x \rangle \geq \sup_{t \geq 0} \langle \bar{x}, ty \rangle \geq \sup_{t \geq 0} t\varepsilon = +\infty,$$

which contradicts $\bar{x} \in \text{dom } \sigma_S$ and finishes the proof. A proof of part (iv) can be found in [109, Proposition C-2.1.3]. □

The next lemma shows that the support function of a set $S \subset \mathbb{R}^n$ and its corresponding lineality space can be used to characterize the affine hull of the set S. This result is presented in [109, Theorem C-2.2.3] and will turn out to be very useful when working with the strict complementarity condition.

Lemma 2.1.9. *Let $S \subseteq \mathbb{R}^n$ be a nonempty, closed, convex set. Then, $s \in \text{aff } S$ if and only if it holds $\langle s, d \rangle = \sigma_S(d)$ for all $d \in \text{lin } \sigma_S$.*

2.1.4. Robinson's constraint qualification

Let us consider the optimization problem

$$\min_x \; f(x) + \varphi(F(x)),$$

where $f : U \to \mathbb{R}$, $F : U \to \mathbb{R}^m$ are continuously differentiable functions on a certain open set $U \subset \mathbb{R}^n$ and $\varphi : \mathbb{R}^m \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous mapping. Throughout this thesis, we will work with the following constraint qualification.

Definition 2.1.10. *We say that* Robinson's constraint qualification *holds at a point $\bar{x} \in \mathbb{R}^n$, $F(\bar{x}) \in \text{dom } \varphi$, if the following condition is satisfied*

$$0 \in \text{int}\{F(\bar{x}) + DF(\bar{x})\mathbb{R}^n - \text{dom } \varphi\}. \tag{2.1.3}$$

Robinson's constraint qualification is stable under small perturbations. In particular, if condition (2.1.3) holds at $\bar{x}$ (and if f and F are continuously differentiable in a neighborhood of $\bar{x}$), then it also follows

$$0 \in \text{int}\{F(x) + DF(x)\mathbb{R}^n - \text{dom } \varphi\}$$

for all $x \in \mathbb{R}^n$ in a sufficiently small neighborhood of $\bar{x}$. A proof and discussion of this result can be found in [27, Section 2.3.4 and Remark 2.88]. We will refer to this property as the *stability property* of Robinson's constraint qualification.

2.2. Convexity and the convex conjugate

At first, we give an equivalent characterization of the continuity of a convex function. The following theorem combines Theorem 8.29 and Corollary 8.30 in [11].

Theorem 2.2.1. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex and proper function. Then, φ is continuous at $x \in \operatorname{dom} \varphi$ if and only if $x \in \operatorname{int} \operatorname{dom} \varphi$. Furthermore, in that case, φ is also locally Lipschitz continuous near x and on the whole set $\operatorname{int} \operatorname{dom} \varphi$.*

Definition 2.2.2. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be given. The* convex conjugate *or* Fenchel conjugate *of φ is defined as*

$$\varphi^* : \mathbb{R}^n \to [-\infty, +\infty], \quad \varphi^*(x) := \sup_{y \in \mathbb{R}^n} \, \langle y, x \rangle - \varphi(y),$$

and the biconjugate φ^{**} of φ is defined as $\varphi^{**} := (\varphi^*)^*$.

Next, we state the classical Moreau-Fenchel Theorem that establishes a duality-like relation between a function φ and its conjugate φ^* and biconjugate φ^{**}.

Theorem 2.2.3 (Moreau-Fenchel, cf. [11, Theorem 13.31]). *The convex conjugate φ^* of a mapping $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, lower semicontinuous function. Moreover, if φ itself is a convex, proper, and lower semicontinuous function, then φ^* is proper and it holds $\varphi^{**} = \varphi$.*

We proceed with two basic examples.

Example 2.2.4 (Indicator function). Let $S \subset \mathbb{R}^n$ be a convex, nonempty, and closed set and consider the *indicator function*

$$\iota_S : \mathbb{R}^n \to (-\infty, +\infty], \quad \iota_S(x) := \begin{cases} 0 & \text{if } x \in S, \\ +\infty & \text{if } x \neq S. \end{cases}$$

Then, by Lemma 2.1.3, it easily follows that the indicator function ι_S is convex, proper, and lower semicontinuous. The convex conjugate of ι_S is the support function of S, i.e.,

$$\iota_S^*(x) = \sup_{s \in S} \, \langle s, x \rangle = \sigma_S(x).$$

Example 2.2.5 (Dual norms). Let $|||\cdot||| : \mathbb{R}^n \to \mathbb{R}$ be a norm on $\mathbb{R}^n$. The *dual norm* $|||\cdot|||_\circ$ of $|||\cdot|||$ is defined as

$$|||x|||_\circ := \sup_{|||y||| \le 1} \, \langle y, x \rangle, \tag{2.2.1}$$

i.e., the dual norm is the support function of the ball $B_{|||\cdot|||}(0,1) := \{x \in \mathbb{R}^n : |||x||| \leq 1\}$. The convex conjugate of the general norm $|||\cdot|||$ can be represented as follows

$$(2.2.2) \qquad |||x|||^* = \iota_{B_{|||\cdot|||_\circ}(0,1)}(x), \quad \forall\, x \in \mathbb{R}^n.$$

In particular, if we consider the ℓ_p-norm $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$ for arbitrary $p \in (1,\infty)$, then, for $q \in (1,\infty)$, $\frac{1}{p} + \frac{1}{q} = 1$, we obtain

$$(\|x\|_p)_\circ = \|x\|_q \quad \text{and} \quad \|x\|_p^* = \iota_{B_{\|\cdot\|_q}(0,1)}(x), \quad \forall\, x \in \mathbb{R}^n.$$

The same relation can also be verified for the ℓ_1-norm and the maximum norm $\|x\|_\infty := \max_{i=1,\ldots,n}\ |x_i|$ or, more generally, for dual matrix norms, such as the spectral and the nuclear norm.

Proof. Here, we will only briefly prove formula (2.2.2). From the definition of the dual norm, it immediately follows

$$(2.2.3) \qquad |||z||| \cdot |||x|||_\circ = \sup_{|||y|||\leq 1}\ |||z||| \cdot \langle y, x\rangle \geq \langle z, x\rangle, \quad \forall\, z \in \mathbb{R}^n.$$

First, let us consider the case $|||x|||_\circ \leq 1$. Then, inequality (2.2.3) implies $\langle z, x\rangle - |||z||| \leq 0$ for all $z \in \mathbb{R}^n$ and for the choice $z = 0$ we obtain $|||x|||^* = \sup_{z\in\mathbb{R}^n}\ \langle z, x\rangle - |||z||| = 0$. On the other hand, if we have $|||x|||_\circ > 1$, then it holds

$$|||x|||^* = \sup_{\mu\geq 0}\ \sup_{|||z|||\leq 1}\ \mu\langle z, x\rangle - |||\mu z||| \geq \sup_{\mu\geq 0}\ \mu \cdot (|||x|||_\circ - 1) = +\infty,$$

as desired. □

2.3. Multifunctions

In the following, we recall several continuity and monotonicity concepts for general, set-valued multifunctions. The definitions of upper semicontinuity and upper Lipschitz continuity are taken from [50, Proposition 2.6.2] and [27, Section 2.3], respectively.

Definition 2.3.1. *A multifunction $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be* upper semicontinuous *at a point $x \in \mathbb{R}^n$, if for any $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$\Phi(y) \subseteq \Phi(x) + B_\varepsilon(0), \quad \forall\, y \in B_\delta(x).$$

Moreover, the multifunction Φ is said to be upper Lipschitzian *at x with modulus $L > 0$, if there exists $\delta > 0$ such that*

$$\Phi(y) \subseteq \Phi(x) + L\|x - y\| \cdot B_1(0), \quad \forall\, y \in B_\delta(x).$$

Definition 2.3.2. *A multifunction $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is called* monotone*, if it holds*

$$\langle u - v, x - y\rangle \geq 0, \quad \forall\, (x,u), (y,v) \in \text{gra}\ \Phi,$$

where gra $\Phi := \{(x,u) \in \mathbb{R}^n \times \mathbb{R}^n : u \in \Phi(x)\}$ *denotes the* graph *of* Φ.

2.4. Epi-convergence

Epi-convergence extends the classical pointwise or uniform convergence of a sequence of real valued functionals $(\varphi_\nu)_\nu$, $\varphi_\nu : \mathbb{R}^n \to \mathbb{R}$, $\nu \in \mathbb{N}$, to the extended real valued setting and is strongly related to the notions of Γ- and *Mosco-convergence*. Simply put, the sequence $(\varphi_\nu)_\nu$ is said to epi-converge if the epigraphs epi φ_ν, $\nu \in \mathbb{N}$, converge to a certain limit set.

In the following subsection, we give a brief overview of the concept of epi-convergence and provide some basic definitions and corresponding tools. For more details on set and epi-convergence we refer to the book of Rockafellar and Wets [208] and the references therein. We also want to note that the material, which is presented in this subsection, is essentially based on the chapters 4 and 7 in [208]. As in [208], let us set

$$\mathcal{N}_\infty^\# := \{N \subset \mathbb{N} : N \text{ is infinite}\} \quad \text{and} \quad \mathcal{N}_\infty := \{N \subset \mathbb{N} : \mathbb{N} \setminus N \text{ is finite}\}.$$

We start with the following definition.

Definition 2.4.1 (cf. [208, Definition 4.1]). *For a sequence $(C_\nu)_\nu$ of subsets of $\mathbb{R}^n$ the* outer limit *is the set*

$$\limsup_{\nu\to\infty} \ C_\nu := \{x \in \mathbb{R}^n : \exists \ N \subset \mathcal{N}_\infty^\#, \ \exists \ x^\nu \in C_\nu \textit{ such that } x^\nu \to x, \ N \ni \nu \to \infty\},$$

while the inner limit *is the set*

$$\liminf_{\nu\to\infty} \ C_\nu := \{x \in \mathbb{R}^n : \exists \ N \subset \mathcal{N}_\infty, \ \exists \ x^\nu \in C_\nu \textit{ such that } x^\nu \to x, \ N \ni \nu \to \infty\}.$$

The limit *of the sequence $(C_\nu)_\nu$ exists if the outer and inner limit sets are equal:*

$$\lim_{\nu\to\infty} \ C_\nu := \limsup_{\nu\to\infty} \ C_\nu = \liminf_{\nu\to\infty} \ C_\nu.$$

In Definition 2.4.1, when the limit $\lim_\nu \ C_\nu$ exists and is equal to a set C, then the sequence $(C_\nu)_\nu$ is said to *converge* to the set C. Set convergence in this sense is known as *Painlevé-Kuratowski convergence*. An exemplary illustration of the set convergence process and of specific inner and outer limits is given in Figure 2.1. Besides Definition 2.4.1, there exist many different, but equivalent characterizations of the inner and outer limit of a sequence of sets. For instance, one has

$$\limsup_{\nu\to\infty} \ C_\nu = \left\{x : \liminf_{\nu\to\infty} \ \mathrm{dist}(x, C_\nu) = 0\right\}, \quad \liminf_{\nu\to\infty} \ C_\nu = \left\{x : \limsup_{\nu\to\infty} \ \mathrm{dist}(x, C_\nu) = 0\right\}.$$

Other alternative definitions and corresponding discussions can be found in [208]. Finally, let us mention the following two basic properties:

- It holds $\liminf_\nu \ C_\nu \subset \limsup_\nu \ C_\nu$.

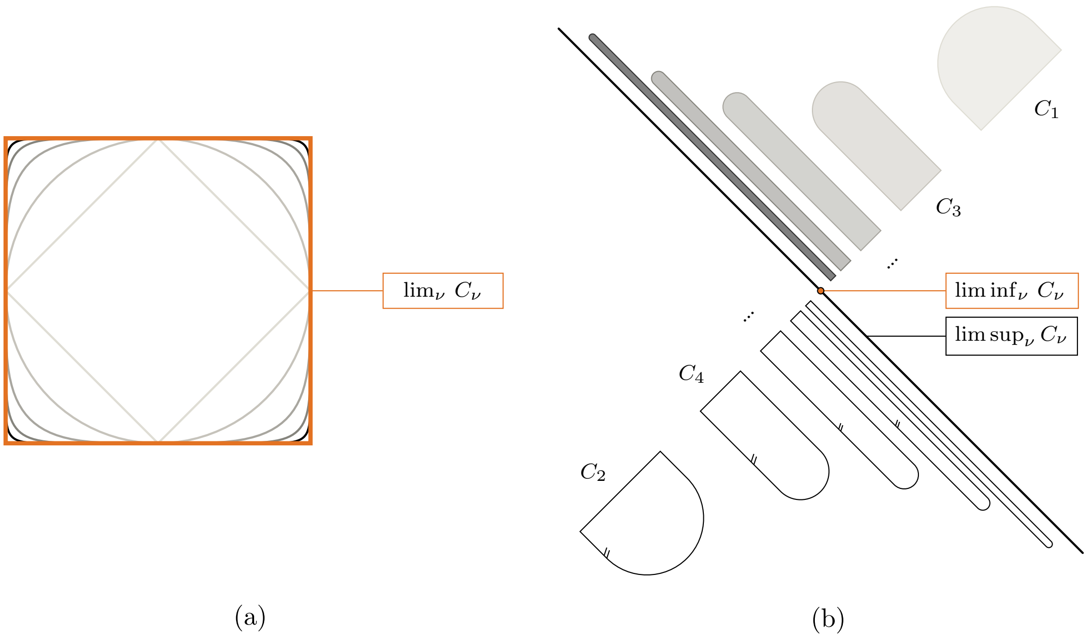


Figure 2.1.: Illustration of the Painlevé-Kuratowski convergence of a sequence of sets. In subfigure (a), the unit disks $C_\nu := \{x \in \mathbb{R}^2 : \|x\|_\nu = 1\}$, $\nu \in \mathbb{N}$, converge to the limit set $C = \{x \in \mathbb{R}^2 : \|x\|_\infty = 1\}$. In subfigure (b), an example for a diverging sequence of sets is given. In particular, in this situation, the outer and inner limit of $(C_\nu)_\nu$ are different sets.

- The inner and outer limit of $(C_\nu)_\nu$ are always closed sets. (See [208, Proposition 4.4]).

We continue with the definition of epi-limits and epi-convergence of a sequence of possibly real extended valued functionals.

Definition 2.4.2 (cf. [208, Definition 7.1]). *Let $(\varphi_\nu)_\nu$, $\varphi_\nu : \mathbb{R}^n \to (-\infty, +\infty]$, be a family of functions. The* lower *and* upper epi-limits *of φ_ν, as $\nu \to \infty$, are defined as the functions whose epigraphs are given by the outer and inner limits of the sets* epi φ_ν*, i.e., it holds*

$$\operatorname{epi}\left[\operatorname{e-}\liminf_{\nu\to\infty}\ \varphi_\nu\right] = \limsup_{\nu\to\infty}\ \operatorname{epi}\ \varphi_\nu, \quad \textit{and} \quad \operatorname{epi}\left[\operatorname{e-}\limsup_{\nu\to\infty}\ \varphi_\nu\right] = \liminf_{\nu\to\infty}\ \operatorname{epi}\ \varphi_\nu.$$

When the two functions $\operatorname{e-}\liminf_\nu\ \varphi_\nu$ *and* $\operatorname{e-}\limsup_\nu\ \varphi_\nu$ *coincide, we say that the* epi-limit function $\varphi := \operatorname{e-}\lim_\nu\ \varphi_\nu$ *exists. Moreover, in this case the sequence $(\varphi_\nu)_\nu$ is said to* epi-converge *to φ.*

Again, there is a large number of alternative expressions and definitions of the lower and upper epi-limits and of epi-convergence. The following representations will turn out to be very useful for our calculations and our subsequent analysis. It holds

$$\operatorname{e-}\liminf_{\nu\to\infty}\ \varphi_\nu(x) = \liminf_{\nu\to\infty,\, \tilde{x}\to x}\ \varphi_\nu(\tilde{x}), \quad \operatorname{e-}\limsup_{\nu\to\infty}\ \varphi_\nu(x) = \sup_{(\nu_k)_k\in\mathcal{N}_\infty^\#}\ \liminf_{k\to\infty,\, \tilde{x}\to x}\ \varphi_{\nu_k}(\tilde{x}),$$

see, e.g., [208, Exercise 7.3] or section 2.2.3 in [27]. Additionally, these alternative formulations also allow for a local and pointwise conception of epi-convergence. In particular, the sequence $(\varphi_\nu)_\nu$ epi-converges at a certain point $x \in \mathbb{R}^n$ if the *upper* and *lower epi-limit values* $\text{e-}\limsup_\nu \varphi_\nu(x)$ and $\text{e-}\liminf_\nu \varphi_\nu(x)$ coincide at x. Next, let us list some important properties of epi-limits.

- The upper and lower epi-limits of $(\varphi_\nu)_\nu$ are lower semicontinuous functions.
- If φ_ν is positively homogeneous for all $\nu \in \mathbb{N}$, then the functions $\text{e-}\liminf_\nu \varphi_\nu$ and $\text{e-}\limsup_\nu \varphi_\nu$ are also positively homogeneous, [208, Proposition 7.4].

Now, we present another, essential characterization of epi-convergence.

Lemma 2.4.3 (cf. [208, Proposition 7.2]). *Let $(\varphi_\nu)_\nu$, $\varphi_\nu : \mathbb{R}^n \to (-\infty, +\infty]$, be a given sequence of functionals. Then, $(\varphi_\nu)_\nu$ epi-converges to φ if and only if at each point $x \in \mathbb{R}^n$ it holds*

$$\begin{cases} \liminf\limits_{\nu\to\infty} \ \varphi_\nu(x^\nu) \geq \varphi(x) & \textit{for every sequence } x^\nu \to x, \\ \limsup\limits_{\nu\to\infty} \ \varphi_\nu(x^\nu) \leq \varphi(x) & \textit{for some sequence } x^\nu \to x. \end{cases} \tag{2.4.1}$$

Clearly, this criterion can also be used for local or pointwise epi-convergence. In this case, (2.4.1) has to be verified only at the points of interest. In the following theorem, we establish an important connection between epi-convergence and uniform convergence of a sequence of convex functions $(\varphi_\nu)_\nu$.

Theorem 2.4.4 (cf. [208, Theorem 7.17]). *Let $(\varphi_\nu)_\nu$, $\varphi_\nu : \mathbb{R}^n \to (-\infty, +\infty]$, be a given sequence of convex functions and suppose that $(\varphi_\nu)_\nu$ epi-converges to a convex, real valued, and lower semicontinuous limit function $\varphi : \mathbb{R}^n \to \mathbb{R}$. Then, $(\varphi_\nu)_\nu$ converges uniformly to φ on every compact set $C \subset \mathbb{R}^n$.*

Remark 2.4.5 (cf. [208, Definition 7.12]). Since the functions φ_ν are generally extended real valued, we have to clarify the meaning of uniform convergence of the sequence $(\varphi_\nu)_\nu$ in Theorem 2.4.4. For a function $\varphi : \mathbb{R}^n \to [-\infty, +\infty]$ and arbitrary $\rho > 0$, the *ρ-truncation* of φ is defined as

$$\varphi|_\rho(x) := \begin{cases} -\rho & \text{if } \varphi(x) < -\rho, \\ \varphi(x) & \text{if } \varphi(x) \in [-\rho, \rho], \\ \rho & \text{if } \varphi(x) > \rho. \end{cases}$$

Then, a sequence $(\varphi_\nu)_\nu$ is said to *converge uniformly* to φ on a set $S \subset \mathbb{R}^n$, if, for every $\rho > 0$, the sequence of ρ-truncations $(\varphi_\nu|_\rho)_\nu$ converges uniformly to $\varphi|_\rho$ on S.

2.5. Directional (epi-)derivatives and subdifferentials

Epiderivatives are one of the numerous extensions of the classical directional derivative that have been developed to study and express differentiability properties of general nonconvex, nonsmooth and extended real valued functions. As the name already indicates, epiderivatives are based on epigraphical convergence processes of certain difference quotients and not on the common notion of convergence that is used, e.g., in the derivation of classical directional derivatives. In the following, we introduce and list important definitions and calculation rules for epiderivatives. Afterwards, we continue with a discussion of the convex subdifferential, present connections to epidifferentiability and introduce the Clarke subdifferential for general vector valued functions.

An extensive literature review on epiderivatives and related topics can be found in the commentaries at the end of the chapters 7 and 8 in [208]. Our terminology for epiderivatives follows the notation in [27]; more details can be found in the sections 2.2 and 2.4 in [27]. Moreover, the interested reader is once more referred to the chapters 7 and 8 in [208], where many additional properties and further concepts are provided.

For more information on subdifferential calculus and Clarke's subdifferential we refer to [27, 11] and [50, 208, 238].

2.5.1. Directional (epi-)differentiability

Definition 2.5.1. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ and $x \in \operatorname{dom} \varphi$ be given. The* lower *and* upper directional derivatives *of φ at x are defined as*

$$\varphi'_-(x; h) := \liminf_{t \downarrow 0} \frac{\varphi(x + th) - \varphi(x)}{t},$$

and

$$\varphi'_+(x; h) := \limsup_{t \downarrow 0} \frac{\varphi(x + th) - \varphi(x)}{t},$$

respectively. We say that φ is directionally differentiable *at x in direction h if $\varphi'_+(x; h) = \varphi'_-(x; h)$. In this case, we will use the term $\varphi'(x; h)$ to denote the coinciding derivative.*

Since φ is an extended real valued function, the directional derivative $\varphi'(x; \cdot)$ is also an extended real valued function that can take the values $-\infty$ and $+\infty$. Obviously, if for some $h \in \mathbb{R}^n$, $\varphi'(x; h)$ is finite, then it coincides with the usual directional derivative. If the directional derivative $\varphi'(x; h)$ exists for all $h \in \mathbb{R}^n$, then the function φ is said to be *directionally differentiable* at x. We want to point out that the latter definitions do also make sense for general vector valued functions $F : \mathbb{R}^n \to \mathbb{R}^m$. Now, let us turn to epidifferentiability.

Definition 2.5.2. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ and $x \in \operatorname{dom} \varphi$ be given. We define the* lower *and* upper directional epiderivatives *of φ at x in the direction $h \in \mathbb{R}^n$ as follows:*

$$\varphi^{\downarrow}_-(x; h) := \liminf_{t \downarrow 0,\, \tilde{h} \to h} \frac{\varphi(x + t\tilde{h}) - \varphi(x)}{t},$$

$$\varphi^{\downarrow}_+(x; h) := \sup_{(t_k)_k \in \mathcal{N}_0} \ \liminf_{k \to \infty,\, \tilde{h} \to h} \frac{\varphi(x + t_k\tilde{h}) - \varphi(x)}{t_k},$$

where $\mathcal{N}_0$ denotes the space of all positive real sequences $(t_k)_k$ converging to zero. We say that φ is directionally epidifferentiable *at x in the direction h, if $\varphi^{\downarrow}_-(x; h) = \varphi^{\downarrow}_+(x; h)$. In this case, the common value will be denoted by $\varphi^{\downarrow}(x; h)$.*

Next, we list some important properties of the directional epiderivatives:

- Clearly, the directional epiderivatives $\varphi_+^\downarrow(x;\cdot)$, $\varphi_-^\downarrow(x;\cdot)$, and $\varphi^\downarrow(x;\cdot)$ can be interpreted as epi-limits of the difference quotient functions
$$\Delta_t\,\varphi(x)(h) := \frac{\varphi(x+th)-\varphi(x)}{t} \quad \text{for} \quad t \neq 0.$$
In particular, by Lemma 2.4.3, φ is directionally epidifferentiable at x in the direction $h \in \mathbb{R}^n$ if and only if for every sequence $(t_k)_k$, $t_k \downarrow 0$, it holds
$$\begin{cases} \liminf\limits_{k\to\infty}\ \Delta_{t_k}\varphi(x)(h^k) \geq \varphi^\downarrow(x;h) & \text{for every sequence } h^k \to h, \\ \limsup\limits_{k\to\infty}\ \Delta_{t_k}\varphi(x)(h^k) \leq \varphi^\downarrow(x;h) & \text{for some sequence } h^k \to h. \end{cases}$$
As a consequence, the epiderivates $\varphi_+^\downarrow(x;\cdot)$, $\varphi_-^\downarrow(x;\cdot)$, and $\varphi^\downarrow(x;\cdot)$ are lower semicontinuous and positively homogeneous functions.

- It holds:
$$\varphi_-^\downarrow(x;h) \leq \varphi_+^\downarrow(x;h), \quad \varphi_-^\downarrow(x;h) \leq \varphi'_-(x;h), \quad \varphi_+^\downarrow(x;h) \leq \varphi'_+(x;h).$$

- If φ is Lipschitz continuous near x, then it follows $\varphi_-^\downarrow(x;h) = \varphi'_-(x;h)$ and $\varphi_+^\downarrow(x;h) = \varphi'_+(x;h)$ for all $h \in \mathbb{R}^n$.

The following lemma establishes a connection between the different tangent cones of the epigraph epi φ and the epigraphs of the epiderivatives $\varphi_+^\downarrow(x;\cdot)$ and $\varphi_-^\downarrow(x;\cdot)$.

Lemma 2.5.3 (cf. [27, Proposition 2.58]). *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper, extended real valued function and let $x \in \mathrm{dom}\ \varphi$ be given. Then, it holds*
$$\begin{aligned} T_{\mathrm{epi}\ \varphi}(x,\varphi(x)) &= \mathrm{epi}\ \varphi_-^\downarrow(x;\cdot), \\ T^i_{\mathrm{epi}\ \varphi}(x,\varphi(x)) &= \mathrm{epi}\ \varphi_+^\downarrow(x;\cdot). \end{aligned}$$

Remark 2.5.4. Suppose that $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, proper function and let $x \in \mathrm{dom}\ \varphi$ be arbitrary. Using the convexity of φ, we readily establish that the epigraph epi φ is convex. Consequently, due to (2.1.1), the contingent cone $T_{\mathrm{epi}\ \varphi}(x,\varphi(x))$ and the inner tangent cone $T^i_{\mathrm{epi}\ \varphi}(x,\varphi(x))$ coincide and are also convex sets. This implies
$$\varphi_-^\downarrow(x;h) = \varphi_+^\downarrow(x;h), \quad \forall\ h \in \mathbb{R}^n,$$
i.e., φ is directionally epidifferentiable at x. Moreover, in this case, $\varphi^\downarrow(x;\cdot)$ is a convex, lower semicontinuous, and positively homogeneous function.

Lemma 2.5.5. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex and proper function and let $x \in \mathrm{dom}\ \varphi$ be arbitrary. Then, φ is directionally differentiable at x and it holds*

$$\varphi(x+h) - \varphi(x) \geq \varphi'(x;h) \geq \varphi^\downarrow(x;h), \quad \forall\ h \in \mathbb{R}^n. \tag{2.5.1}$$

Proof. The existence of $\varphi'(x;\cdot)$ is shown in [11, Proposition 17.2]. The inequality (2.5.1) follows immediately from the convexity of φ and Remark 2.5.4. □

We continue with several calculation rules.

Lemma 2.5.6 (cf. [27, Proposition 2.136]). *Let $\varphi : \mathbb{R}^m \to (-\infty, +\infty]$ be convex, proper, and lower semicontinuous and let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable function. Suppose that Robinson's constraint qualification*

$$0 \in \operatorname{int}\{F(x) + DF(x)\mathbb{R}^n - \operatorname{dom}\,\varphi\} \tag{2.5.2}$$

is satisfied at $x \in F^{-1}(\operatorname{dom}\,\varphi)$. Then, the composite function $\varphi \circ F$ is directionally epidifferentiable at x and it holds

$$(\varphi \circ F)^{\downarrow}(x; h) = \varphi^{\downarrow}(F(x); DF(x)h), \quad \forall\ h \in \mathbb{R}^n. \tag{2.5.3}$$

Corollary 2.5.7. *Let $\varphi : \mathbb{R}^m \to (-\infty, +\infty]$ be convex, proper, and lower semicontinuous and let $f : \mathbb{R}^n \to \mathbb{R}$, $F : \mathbb{R}^n \to \mathbb{R}^m$ be two continuously differentiable functions. Suppose that Robinson's constraint qualification* (2.5.2) *is satisfied at $x \in F^{-1}(\operatorname{dom}\,\varphi)$. Then, the function $\psi := f + \varphi \circ F$ is directionally epidifferentiable at x and it holds*

$$\psi^{\downarrow}(x; h) = \nabla f(x)^\top h + \varphi^{\downarrow}(F(x); DF(x)h), \quad \forall\ h \in \mathbb{R}^n. \tag{2.5.4}$$

Proof. Let us define $\theta(y, t) := \varphi(y) + t$ and $G : \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}$, $G(y) := (F(y), f(y))$. Then, ψ can be written as the composition of the mappings θ and G and it is easy to show that the corresponding assumptions in Lemma 2.5.6 are fulfilled. Thus, ψ is directionally epidifferentiable at x and, by applying [27, Lemma 2.137] (or by a direct calculation of $\theta^{\downarrow}(y, t; \cdot)$), formula (2.5.4) can be established. □

Remark 2.5.8. Let us reconsider the situation of Corollary 2.5.7 and suppose that Robinson's constraint qualification is not necessarily satisfied at $x \in F^{-1}(\operatorname{dom}\,\varphi)$. Clearly, we then cannot expect that the composite function ψ is directionally epidifferentiable or that the chain rule (2.5.4) is valid. However, we are still able to show a somewhat weaker result. Therefore, let us define $\Upsilon(y) := \nabla f(x)^\top y + \varphi^{\downarrow}(F(x); DF(x)y)$. Then, it follows

$$\operatorname{epi}\,\psi^{\downarrow}_{-}(x; \cdot) \subseteq \operatorname{epi}\,\Upsilon.$$

In particular, we have

$$\psi^{\downarrow}_{-}(x; h) \geq \Upsilon(h) = \nabla f(x)^\top h + \varphi^{\downarrow}(F(x); DF(x)h), \quad \forall\ h \in \mathbb{R}^n.$$

Proof. Let $(h, \tau) \in \operatorname{epi}\,\psi^{\downarrow}_{-}(x; \cdot)$ be an arbitrary vector. Then, due to Lemma 2.5.3 and Definition 2.1.4, there exist sequences $(t_k)_k$, $t_k \downarrow 0$, $h^k \to h$, and $\tau_k \to \tau$ such that

$$\psi(x + t_k h^k) - \psi(x) \leq t_k \tau_k.$$

Next, a Taylor expansion of $f(x + t_k h^k)$ and $F(x + t_k h^k)$ at x yields

$$t_k \nabla f(x)^\top h + \varphi(F(x) + t_k DF(x)h^k + o(t_k)) - \varphi(F(x)) \leq t_k \tau + o(t_k).$$

Dividing both sides of the latter inequality by $t_k > 0$ and taking the limes inferior $k \to \infty$, we obtain

$$\begin{aligned}\tau &\geq \nabla f(x)^\top h + \liminf_{k\to\infty} \frac{\varphi(F(x) + t_k DF(x)h^k + o(t_k)) - \varphi(F(x))}{t_k} \\ &\geq \nabla f(x)^\top h + \liminf_{t\downarrow 0,\, \tilde{h}\to DF(x)h} \frac{\varphi(F(x) + t\tilde{h}) - \varphi(F(x))}{t} = \Upsilon(h),\end{aligned}$$

where we used Remark 2.5.4. This shows $(h, \tau) \in \operatorname{epi} \Upsilon$, which completes the proof. □

Example 2.5.9 (cf. [27, Example 2.67]). Let $S \subset \mathbb{R}^n$ be a convex, nonempty, and closed set and let us consider $\varphi \equiv \iota_S$ and $x \in S$. Then, Definition 2.1.4 and (2.1.1) imply

$$\varphi'(x; h) = \iota_{\mathcal{R}_S(x)}(h), \quad \text{and} \quad \varphi^\downarrow(x; h) = \iota_{T_S(x)}(h).$$

2.5.2. The convex subdifferential

Definition 2.5.10. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ and $x \in \operatorname{dom} \varphi$ be given. The* subdifferential *of φ is the multifunction*

$$\partial\varphi : \mathbb{R}^n \rightrightarrows \mathbb{R}^n, \quad \partial\varphi(x) := \{s \in \mathbb{R}^n : \varphi(y) - \varphi(x) \geq \langle s, y - x\rangle, \ \forall\, y \in \mathbb{R}^n\}.$$

The function φ is called subdifferentiable *at x if $\partial\varphi(x) \neq \emptyset$. The elements $s \in \partial\varphi(x)$ are called* subgradients *of φ at x.*

In the following, we list several important properties of the convex subdifferential.

Lemma 2.5.11 (cf. [27, Proposition 2.125 and 2.134]). *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex function and let $x \in \operatorname{dom} \varphi$ be arbitrary. Then, it holds:*

(i) *φ is subdifferentiable at x if and only if $\varphi^\downarrow(x; 0) > -\infty$, or, equivalently, $\varphi^\downarrow(x; 0) = 0$.*

(ii) *If φ is subdifferentiable at x, then*

$$\varphi^\downarrow(x; h) = \sup_{\lambda\in\partial\varphi(x)} \langle \lambda, h\rangle = \sigma_{\partial\varphi(x)}(h),$$

i.e., $\varphi^\downarrow(x; \cdot)$ is the support function of the subdifferential $\partial\varphi(x)$.

(iii) *Suppose that $x \in \operatorname{ri} \operatorname{dom} \varphi$, then φ is subdifferentiable at x.*

Lemma 2.5.12 (cf. [27, Proposition 2.132], [11, Proposition 16.14]). *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be convex, proper and let $x \in \operatorname{dom} \varphi$ be given. Then, it holds:*

(i) *φ is continuous at x if and only if $\partial\varphi(x)$ is nonempty and bounded.*

(ii) *If φ is continuous at x, then there exists $\varepsilon > 0$ such that $\partial\varphi(B_\varepsilon(x))$ is bounded.*

The following lemma provides an important characterization of convex, proper, lower semicontinuous, and positively homogeneous functions. A proof can be found in [11, Proposition 14.11 and Proposition 16.18].

Lemma 2.5.13. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, lower semicontinuous, and positively homogeneous function. Then, φ is the support function of the subdifferential $\partial\varphi(0)$, i.e., we have*

$$\varphi(x) = \sigma_{\partial\varphi(0)}(x), \quad \forall\ x \in \mathbb{R}^n.$$

In particular, φ is subdifferentiable at 0 and satisfies $\varphi(0) = 0$.

Let us briefly summarize and recapitulate the last results. Combing Lemma 2.5.11 (i) and (ii), we see that a convex, proper function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is subdifferentiable at some $x \in \operatorname{dom} \varphi$ if and only if its corresponding directional epiderivative $\Pi(\cdot) := \varphi^{\downarrow}(x; \cdot)$ is proper. Hence, by Remark 2.5.4, Lemma 2.5.13 is applicable to the mapping Π and, due to Lemma 2.1.8 (ii) and Lemma 2.5.11 (ii), we obtain the following essential relation

$$\partial\Pi(0) = \partial\varphi(x). \tag{2.5.5}$$

Now, let φ be also lower semicontinuous and positively homogeneous. Then, by Lemma 2.1.8 (iv) and Lemma 2.5.12 (i), φ is real valued, (i.e., $\operatorname{dom} \varphi = \mathbb{R}^n$), if and only if φ is continuous at 0. Next, we present a connection between the subdifferentiability of φ and its convex conjugate φ^*.

Lemma 2.5.14 (cf. [11, Theorem 16.23]). *Suppose that $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is convex, proper, and lower semicontinuous and let $x, s \in \mathbb{R}^n$ be arbitrary. Then, the following statements are equivalent:*

(i) $s \in \partial\varphi(x)$.

(ii) $x \in \partial\varphi^*(s)$.

(iii) $(s, -1) \in N_{\text{epi } \varphi}(x, \varphi(x))$.

(iv) $\varphi(x) + \varphi^*(s) = \langle x, s\rangle$.

Next, we present a chain rule for the subdifferential of a composition of convex functions.

Lemma 2.5.15 (cf. [11, Proposition 16.5 and Theorem 16.37]). *Let the two functions $f : \mathbb{R}^n \to (-\infty, +\infty]$ and $\varphi : \mathbb{R}^m \to (-\infty, +\infty]$ be convex, proper, and lower semicontinuous and let $A \in \mathbb{R}^{m\times n}$ be an arbitrary matrix. Furthermore, let us set $\psi(x) := f(x) + \varphi(Ax)$ and suppose that $A(\operatorname{dom} f) \cap \operatorname{dom} \varphi \neq \emptyset$. Then, it holds*

$$\partial f(x) + A^\top \partial g(Ax) \subseteq \partial\psi(x), \quad \forall\ x \in \mathbb{R}^n. \tag{2.5.6}$$

Additionally, if one of the following regularity conditions

(i) $0 \in \operatorname{int}\{A(\operatorname{dom} f) - \operatorname{dom} \varphi\}$

(ii) $A(\operatorname{ri} \operatorname{dom} f) \cap \operatorname{ri} \operatorname{dom} \varphi \neq \emptyset$

is satisfied, then it follows

$$\partial\psi(x) = \partial f(x) + A^\top \partial g(Ax), \quad \forall\ x \in \mathbb{R}^n. \tag{2.5.7}$$

We conclude this subsection with two basic and explicit examples.

Example 2.5.16. Let $S \subset \mathbb{R}^n$ be a convex set. The indicator function ι_S of S is subdifferentiable at $x \in \mathbb{R}^n$ if and only if $x \in S$. Consequently, the subdifferential of ι_S is given by

$$\partial \iota_S(x) = \begin{cases} \{s \in \mathbb{R}^n : \langle s, y - x \rangle \le 0, \forall\, y \in S\} = N_S(x) & \text{if } x \in S, \\ \emptyset & \text{otherwise.} \end{cases}$$

Example 2.5.17. Let $S \subset \mathbb{R}^n$ be a convex, nonempty, and closed set and consider the support function $\sigma_S : \mathbb{R}^n \to (-\infty, +\infty]$. Then, for all $x \in \operatorname{dom} \sigma_S$, it holds

$$\partial \sigma_S(x) = \{s \in S : \langle s, x \rangle = \sigma_S(x)\}. \tag{2.5.8}$$

Proof. Due to Example 2.2.4, the indicator function ι_S of S is a convex, proper, and lower semicontinuous function. Thus, using Theorem 2.2.3, we have $\sigma_S^*(x) = \iota_S^{**}(x) = \iota_S(x)$ for all $x \in \mathbb{R}^n$. Now, Lemma 2.5.14 implies

$$s \in \partial \sigma_S(x) \quad \Longleftrightarrow \quad \sigma_S(x) + \iota_S(s) = \langle s, x \rangle.$$

Clearly, this establishes formula (2.5.8) and we can conclude the proof. □

2.5.3. The Bouligand and the Clarke subdifferential

In this subsection, we want to derive a subdifferential calculus for vector valued functions of the type

$$F : U \to \mathbb{R}^m,$$

where $U \subset \mathbb{R}^n$ is an open, nonempty set. In this respect, let $\Omega_F \subset U$ denote the set of all points $x \in U$ at which F is Fréchet differentiable with derivative $DF(x) : \mathbb{R}^n \to \mathbb{R}^m$. Moreover, if the function F is locally Lipschitz continuous in a neighborhood $V \subset U$ of the point x, then, according to *Rademacher's Theorem*, the set $V \backslash \Omega_F$ has zero Lebesgue measure. Consequently, every locally Lipschitz continuous function $F : U \to \mathbb{R}^m$ is differentiable almost everywhere. This observation forms the basis of the definition of a generalized notion of differentiability for vector valued functions.

In the following, we will also use the term $DF(x)$ to denote the corresponding Jacobian of F at x, i.e., $J_F(x) \equiv DF(x) \in \mathbb{R}^{m \times n}$. Let us start with the definition of the Bouligand and Clarke subdifferential.

Definition 2.5.18 (Generalized derivatives). *Let $U \subset \mathbb{R}^n$ be open, nonempty, $x \in U$, and let $F : U \to \mathbb{R}^m$ be Lipschitz continuous in a neighborhood of x. The set*

$$\partial_B F(x) := \{M \in \mathbb{R}^{m \times n} : \exists\ (x^k)_k \subset \Omega_F \text{ such that } x^k \to x,\ DF(x^k) \to M\}$$

is called Bouligand subdifferential *or* B-subdifferential *of F at x. The* Clarke subdifferential *$\partial F(x)$ of F at x is defined as the convex hull of $\partial_B F(x)$, i.e., it holds*

$$\partial F(x) := \operatorname{conv}(\partial_B F(x)).$$

Moreover, the C-subdifferential *of F at x is given by $\partial_C F(x) := \partial F_1(x) \times ... \times \partial F_m(x)$.*

The next lemma presents some basic properties of the different subdifferentials. A proof can be found, e.g., in Clarke [50, Proposition 2.6.2].

Lemma 2.5.19 (cf. [238, Proposition 2.2]). *Suppose that $U \subset \mathbb{R}^n$ is open and $F : U \to \mathbb{R}^m$ is locally Lipschitz continuous near $x \in U$. Then, the following statements hold:*

(i) *The set $\partial_B F(x)$ is nonempty and compact.*

(ii) *$\partial F(x)$ and $\partial_C F(x)$ are convex, nonempty, and compact.*

(iii) *The multifunctions $\partial_B F$, ∂F, and $\partial_C F$ are locally bounded and upper semicontinuous.*

(iv) *$\partial_B F(x) \subset \partial F(x) \subset \partial_C F(x)$.*

(v) *If F is continuously differentiable near x, then it holds*

$$\partial_B F(x) = \partial F(x) = \partial_C F(x) = \{DF(x)\}.$$

For a convex functional $f : \mathbb{R}^n \to \mathbb{R}$, it can be shown that Clarke's subdifferential coincides with the convex subdifferential ∂f – up to transposition of course (see, e.g., [50, Proposition 2.2.7]). Clearly, this fact (and the discussion at the end of this subsection) indicates that the Clarke subdifferential is also connected to other differentiability concepts. For instance, in the real valued case, another directional derivative-type construction, the so-called *generalized directional derivatives*, can be used to characterize Clarke's subdifferential. We refer to [50] for more information. Moreover, in [208, Chapter 8.C and 9], these generalized directional derivatives are studied in an even more general epigraphical framework under the name *regular subderivatives*. Since these deep theoretical concepts are not specifically relevant for our later analysis, we will not go into detail here.

Let us continue with a sum and a chain rule for the Clarke subdifferential.

Lemma 2.5.20. *Let $U \subset \mathbb{R}^n$ be open, nonempty and let $F : U \to \mathbb{R}^m$ be continuously differentiable in a neighborhood of $x \in U$. In addition, suppose that $G : U \to \mathbb{R}^m$ is locally Lipschitz continuous near x and let $A \in \mathbb{R}^{m \times m}$ be an arbitrary, invertible matrix. Then, setting $\Phi \equiv F + A \cdot G$, it follows*

$$\partial\Phi(x) = DF(x) + A \cdot \partial G(x).$$

Proof. It suffices to show $\partial_B \Phi(x) = DF(x) + A \cdot \partial_B G(x)$. However, this equality follows immediately from the fact that Φ is Fréchet differentiable at some point $y \in \mathbb{R}^n$ if and only if G is Fréchet differentiable at y. □

Theorem 2.5.21. *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open, nonempty sets and suppose that $G : U \to V$ is Lipschitz continuous near $x \in U$ and $F : V \to \mathbb{R}^p$ is Lipschitz continuous in a neighborhood of $G(x) \in V$. Then, the composite function $\Phi \equiv F \circ G$ is Lipschitz continuous near x, and it holds*

$$\partial\Phi(x)h \subset \operatorname{conv}\{\partial F(G(x)) \circ \partial G(x)h\}, \quad \forall\ h \in \mathbb{R}^n. \tag{2.5.9}$$

Moreover, if G is continuously differentiable in a neighborhood of x, then formula (2.5.9) *can be further simplified:*

(i) *If F is real valued, i.e., if we have $p = 1$, then it holds*

$$\partial \Phi(x) \subset \partial F(G(x)) \circ DG(x). \tag{2.5.10}$$

(ii) *In the general case, if, in addition, the linear mapping $DG(x) : \mathbb{R}^n \to \mathbb{R}^m$ is onto, and F is directionally differentiable at every point in V, then it follows*

$$\partial \Phi(x) = \partial F(G(x)) \circ DG(x). \tag{2.5.11}$$

Proof. The first and second part of this Theorem are proven in [50, Corollary 2.6.6 and Theorem 2.3.10]. The last part can be found in [226, Lemma 2.1]. □

Remark 2.5.22. The conditions in Theorem 2.5.21 (i) and (ii) are rather restrictive and, in general, even if the inner function G is continuously differentiable, we cannot expect that an equality based representation of the Clarke subdifferential $\partial\Phi(x)$ as in (2.5.11) is available. However, if F and G satisfy the basic assumptions of Theorem 2.5.21 at some point $x \in \mathbb{R}^n$ and if G is continuously differentiable near x, then the following chain rule in terms of the C-subdifferential of F does hold:

$$\partial \Phi(x) \subset \partial_C \Phi(x) \subset \partial_C F(G(x)) \circ DG(x). \tag{2.5.12}$$

In applications, such as, e.g., *Fischer-Burmeister*-based reformulations of KKT-systems, [72, Proposition 3.1], the set $\partial_C F(G(x)) \circ DG(x)$ is often used to construct specific generalized derivatives of composite functions, since it typically has a much simpler structure than Clarke's subdifferential $\partial\Phi(x)$.

Proof. The first inclusion in (2.5.12) follows from Lemma 2.5.19 (iv). By the definition of the C-subdifferential, the condition $M \in \partial_C\Phi(x)$ means that the i-th row of M is an element of Clarke's subdifferential $\partial\Phi_i(x)$. Thus, using (2.5.10), it follows

$$M_{[i\cdot]} \in \partial\Phi_i(x) = \partial(F_i \circ G)(x) \subset \partial F_i(G(x)) \circ DG(x), \quad \forall\, i = 1, ..., m.$$

Clearly, this immediately implies (2.5.12). □

Remark 2.5.23. Another chain rule that guarantees a complete characterization of the Clarke subdifferential $\partial\Phi(x)$ and equality as in (2.5.11) is derived in [183, Proposition 7]. There, Pang et al. considered the opposite case, when the outer function F is continuously differentiable.

Finally, we present an important differentiability concept that ensures that the Clarke subdifferential reduces to a singleton and which is generally weaker than continuous differentiability. Following [50] and [208, Section 9.D], a function $F : U \to \mathbb{R}^m$ with $U \subset \mathbb{R}^n$ open and nonempty, is called *strictly differentiable* at $x \in U$ if

$$\lim_{y,z\to x,\, y\neq z} \frac{F(z) - F(y) - DF(x)(z-y)}{\|z-y\|} = 0,$$

where $DF(x)$ denotes the classical Fréchet derivative of F at x. Before we state Theorem 2.5.24, let us mention and highlight some properties of strictly differentiable functions.

- If $F : U \to \mathbb{R}^m$ is continuously differentiable in a neighborhood of $x \in U$, then F is also strictly differentiable at x.
- A function $F : U \to \mathbb{R}^m$ is strictly differentiable on an open set $V \subset U$ if and only if F is continuously differentiable on V.
- The composition of strictly differentiable functions is strictly differentiable.

Further properties and more information on strictly differentiable functions can be found in [50, Section 2.2], [208, Section 9.C–9.D], or [67, 1D.6–1D.8].

Theorem 2.5.24. *Let $U \subset \mathbb{R}^n$ be open and nonempty. A function $F : U \to \mathbb{R}^m$ is strictly differentiable at $x \in U$ if and only if F is locally Lipschitz continuous near x and Clarke's subdifferential $\partial F(x)$ reduces to a singleton.*

Proof. In the real valued case $m = 1$, this result is presented in [50, Proposition 2.2.4]. In the more general case, we have to use a connection between the Clarke subdifferential and the so-called *graphical* or *Mordukhovich coderivative* $D^*F(x) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ of F at x. Since a full definition of Mordukhovich's coderivative requires the introduction of several more specific tools and concepts, such as, e.g., the *limiting* or *basic normal cone* to the graph of F, we want to refer to [208, Section 8.G] and [158, Definition 1.32] for a detailed discussion. However, if F is locally Lipschitz continuous near x, then [208, Theorem 9.62] provides the following characterization

$$\operatorname{conv}\, D^*F(x)(y) = \operatorname{conv}\left\{M^\top y : M \in \partial_B F(x)\right\} = \left\{M^\top y : M \in \partial F(x)\right\}.$$

Thus, in this situation, Clarke's subdifferential $\partial F(x)$ is a singleton if and only if the coderivative mapping $D^*F(x)$ is single-valued. The rest of the proof now follows from [208, Exercise 9.25] or [158, Theorem 3.66]. □

2.6. Semismoothness

The concept of semimoothness was originally introduced and developed by Mifflin [148] for real valued functionals. Later, Qi [197] and Qi and Sun [199] extended this notion to general mappings between finite dimensional spaces. The importance and popularity of semismoothness can be traced back to the fact that Newton's method applied to the nonlinear and possibly nonsmooth equation

$$F(x) = 0$$

is well-defined and can be shown to converge locally at least q-superlinearily under suitable conditions, if the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is semismooth. Clearly, this generalizes the classical Newton method and enlarges the overall applicability of Newton-type methods to many different and broad classes of problems.

In the following, we state some basic definitions and give an overview of the concept of semismoothness in the context of nonsmooth equations. More details on semismoothness and the semismooth Newton method can be found in [197, 199, 198]. Extensions to the infinite dimensional setting are presented in the book [238] by Ulbrich.

In the literature, a large number of different yet equivalent definitions of semismoothness is available. Here, we will only give one of these definitions (see, e.g., [183, Theorem 5] or [238, Proposition 2.7]), that will turn out to be the most useful version for the convergence analysis of our semismooth Newton-type method later on. The original definition of Mifflin and various other, equivalent formulations, as well as corresponding proofs and discussions can be found in [199, 182, 198, 183, 238].

Definition 2.6.1 (Semismoothness). *Let $U \subset \mathbb{R}^n$ be an open and nonempty set. A function $F : U \to \mathbb{R}^m$ is said to be* semismooth *at $x \in U$, if F is Lipschitz continuous in a neighborhood of x, directionally differentiable at x, and for all $h \in \mathbb{R}^n$ it holds*

$$\sup_{M \in \partial F(x+h)} \|F(x+h) - F(x) - Mh\| = o(\|h\|) \quad \text{as} \quad h \to 0. \tag{2.6.1}$$

If F is semismooth at all $x \in U$, then F is called semismooth *(on U).*

Let us list some elementary and well-known examples of semismooth functions.

- Convex, real valued functions are semismooth, [148, Proposition 3].
- Piecewise continuously differentiable functions, such as, e.g., the ℓ_1- or ℓ_∞-norm, are semismooth. For more details we refer to [211] and [238, Section 2.5.3].
- If $F : U \to \mathbb{R}^m$, $U \subset \mathbb{R}^n$ open, is continuously differentiable in a neighborhood of some point $x \in U$, then F is semismooth at x, see [148, Proposition 4].

Next, we present the concept of α-order semismoothness, which is a natural extension of semismoothness of a function. Higher order semismoothness was introduced by Qi and Sun in [199] to achieve a better, local convergence rate of the semismooth Newton method.

Definition 2.6.2 (α-order semismoothness). *Let $U \subset \mathbb{R}^n$ be open, nonempty and let $F : U \to \mathbb{R}^m$ be given. For $0 < \alpha \leq 1$, the function F is called* α-order semismooth *at $x \in U$, if F is locally Lipschitz continuous near x, directionally differentiable at x, and for all $h \in \mathbb{R}^n$ it holds*

$$\|F(x+h) - F(x) - F'(x,h)\| = O(\|h\|^{1+\alpha}) \quad \text{as } h \to 0, \tag{2.6.2}$$

and

$$\sup_{M \in \partial F(x+h)} \|F(x+h) - F(x) - Mh\| = O(\|h\|^{1+\alpha}) \quad \text{as } h \to 0. \tag{2.6.3}$$

The function F is said to be α-order semismooth *(on U), if F is α-order semismooth at all points $x \in U$.*

Remark 2.6.3. The condition (2.6.2) is known as *α-order B-differentiability.* Again, there exist several other formulations of α-order semismoothness, see, e.g., [238, Definition 2.13 and Proposition 2.14].

Finally, let us present some calculation rules and properties of semismooth functions. We start with a useful, equivalent characterization of (α-order) semismoothness. A proof of this result can be found in [199, Corollary 2.4] and [238, Proposition 2.10 and Proposition 2.17].

Lemma 2.6.4. *Let $U \subset \mathbb{R}^n$ be open, nonempty and $0 < \alpha \leq 1$. Then, the mapping $F : U \to \mathbb{R}^m$ is (α-order) semismooth at $x \in U$ if and only if each component function $F_i : U \to \mathbb{R}$, $i = 1, ..., m$, is (α-order) semismooth at x.*

Next, we give a chain rule for semismooth functions.

Theorem 2.6.5. *Let $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ be open, nonempty sets and $0 < \alpha \leq 1$. Suppose that $G : U \to V$ is (α-order) semismooth at $x \in U$ and that $F : V \to \mathbb{R}^p$ is (α-order) semismooth at $G(x)$ with $G(U) \subset V$. Then, the composite mapping $F \circ G : U \to \mathbb{R}^p$ is (α-order) semismooth at x.*

Proof. In [148, Theorem 5], it is shown that the composition of two semismooth functions $f : \mathbb{R}^m \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{R}^m$ is again a semismooth mapping. Clearly, by using Lemma 2.6.4, this yields the more general result in Theorem 2.6.5; see also Lemma 18 in [79]. A chain rule for α-order semismooth functions is studied in [79, Theorem 19]. □

Remark 2.6.6. Additionally, let us assume that the function G is continuously differentiable in a neighborhood of x. Then, by combining Lemma 2.6.4 and Theorem 2.6.5, it can be easily shown that the composite function $F \circ G$ is also (α-order) semismooth with respect to the possibly larger set $\partial_C F(G(x))DG(x)$, i.e., it holds

$$\|(F \circ G)(x + h) - (F \circ G)(x) - Mh\| = o(\|h\|)$$

uniformly for all $M \in \partial_C F(G(x + h))DG(x + h)$, as $h \to 0$.

It is well-known that a Fréchet differentiable function is not necessarily semismooth. The following theorem shows that the combination of semismoothness and Fréchet differentiability implies some kind of higher regularity of the considered function. Theorem 2.6.7 is based on some very recent results of Movahedian [165] and will be an essential component of our second order and nonsingularity analysis in section 5.4.

Theorem 2.6.7. *Let $U \subset \mathbb{R}^n$ be open, nonempty and let $F : U \to \mathbb{R}^m$ be Fréchet differentiable and semismooth at $x \in U$. Then, F is strictly differentiable at x and $\partial F(x)$ reduces to a singleton.*

Proof. Due to Theorem 2.5.24, it suffices to show that $\partial F(x)$ reduces to a singleton. Very recently, in [165, Theorem 4.3], it has been established that Mordukhovich's coderivative and the so-called linear coderivative of F at x, see [233, 234], coincide if F is locally Lipschitz continuous, directionally differentiable, and semismooth at x. Now, applying [234, Corollary 2.10] and [233, Proposition 2.14], the Fréchet differentiability of F implies that the linear coderivative of F at x reduces to the classical Fréchet derivative $DF(x)$ and it follows $D^*F(x)(y) = \{DF(x)^\top y\}$ for all $y \in \mathbb{R}^n$. Clearly, as in the proof of Theorem 2.5.24, this yields $\partial F(x) = \{DF(x)\}$. □

3. The proximity operator

The proximity operator was originally introduced and studied by Moreau in his seminal works [161, 162, 163] and has become a popular tool in many different fields of research over the last decades. More specifically, due to its manifold applicabilities in nonsmooth optimization, the proximity operator has been intensively used to design and develop general numerical algorithms, such as fixed point and descent-based methods [54, 88, 236], proximal (quasi)-Newton approaches [14, 129], and alternating or primal-dual schemes [91, 71, 136, 42]. In the following sections, we will introduce the proximity operator, give examples and discuss several important properties. Most of the statements presented here can be found in the paper of Combettes and Wajs [54], where essential features and calculation rules for proximity operators are provided and derived. Further results can also be found in the book [11]. For more information on corresponding numerical methods and applications of the proximity operator we refer to the reviews and monographs [6, 53, 184] and the references therein.

3.1. Definitions and basic properties

Let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary symmetric, positive definite matrix and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function, then the *Moreau-Yosida regularization* or *Moreau envelope* of φ is defined as

$$\text{env}^\Lambda_\varphi : \mathbb{R}^n \to \mathbb{R}, \quad \text{env}^\Lambda_\varphi(x) := \min_{y \in \mathbb{R}^n} \ \varphi(y) + \frac{1}{2}\|x - y\|^2_\Lambda, \tag{3.1.1}$$

where $\|x\|_\Lambda := \sqrt{\langle x, x \rangle_\Lambda}$, $x \in \mathbb{R}^n$, is the norm induced by the Euclidean scalar product $\langle \cdot, \cdot \rangle_\Lambda : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $\langle x, y \rangle_\Lambda := \langle \Lambda x, y \rangle = \langle x, \Lambda y \rangle$. For every $x \in \mathbb{R}^n$ the minimum in (3.1.1) is attained at the *unique* point $\text{prox}^\Lambda_\varphi(x)$ that is characterized by the optimality condition

$$\text{prox}^\Lambda_\varphi(x) \in x - \Lambda^{-1} \cdot \partial\varphi(\text{prox}^\Lambda_\varphi(x)), \tag{3.1.2}$$

where $\partial\varphi$ denotes the convex subdifferential of φ. The function

$$\text{prox}^\Lambda_\varphi : \mathbb{R}^n \to \mathbb{R}^n, \quad \text{prox}^\Lambda_\varphi(x) := \operatorname*{arg\,min}_{y \in \mathbb{R}^n} \ \varphi(y) + \frac{1}{2}\|x - y\|^2_\Lambda$$

is called *proximity operator* of φ. Usually, the matrix Λ is chosen as a fixed one-dimensional parameter by setting $\Lambda = \frac{1}{\lambda}I$ for some $\lambda > 0$. This leads to the classical proximity operator

$$\text{prox}^{\frac{1}{\lambda}I}_\varphi(x) = \operatorname*{arg\,min}_{y \in \mathbb{R}^n} \ \varphi(y) + \frac{1}{2\lambda}\|x - y\|^2_2,$$

which is typically abbreviated by $\text{prox}_{\lambda\varphi}(x)$. In the following, we will discuss characteristic properties of the proximity operator $\text{prox}_\varphi^\Lambda$ for arbitrary parameter matrices $\Lambda \in \mathbb{S}^n_{++}$. Moreover, we will also treat the proximity operator as a function of Λ and establish more general results, which are simple extensions of the classical ones, but, to the best of our knowledge, do not seem to be completely available in the literature so far. We will use the notations

$$\text{env}_\varphi : \mathbb{R}^n \times \mathbb{S}^n \to \mathbb{R}, \quad \text{env}_\varphi(x, \Lambda) := \text{env}_\varphi^\Lambda(x)$$

and

$$\text{prox}_\varphi : \mathbb{R}^n \times \mathbb{S}^n \to \mathbb{R}^n, \quad \text{prox}_\varphi(x, \Lambda) := \text{prox}_\varphi^\Lambda(x),$$

when the Moreau envelope and the proximity operator are explicitly understood as functions of x and Λ, i.e., when the parameter matrix Λ is not fixed. Let us start with some important continuity results of proximity operators.

Lemma 3.1.1 (cf. [163], [54, Lemma 2.4]). *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function and let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary symmetric and positive definite matrix. Then, $\text{prox}_\varphi^\Lambda$ and $I - \text{prox}_\varphi^\Lambda$ are* Λ-firmly nonexpansive operators*, i.e., it holds*

$$\|T(x) - T(y)\|_\Lambda^2 \leq \langle T(x) - T(y), x - y\rangle_\Lambda, \quad \forall\ x, y \in \mathbb{R}^n,$$

for $T \equiv \text{prox}_\varphi^\Lambda$ and $T \equiv I - \text{prox}_\varphi^\Lambda$.

Proof. We follow the proof of Lemma 2.4 in [54] and set $T \equiv \text{prox}_\varphi^\Lambda$, then for $x, y \in \mathbb{R}^n$ the optimality condition (3.1.2) implies

$$\begin{cases} \varphi(\text{prox}_\varphi^\Lambda(y)) - \varphi(\text{prox}_\varphi^\Lambda(x)) \geq \langle x - \text{prox}_\varphi^\Lambda(x), \text{prox}_\varphi^\Lambda(y) - \text{prox}_\varphi^\Lambda(x)\rangle_\Lambda, \\ \varphi(\text{prox}_\varphi^\Lambda(x)) - \varphi(\text{prox}_\varphi^\Lambda(y)) \geq \langle y - \text{prox}_\varphi^\Lambda(y), \text{prox}_\varphi^\Lambda(x) - \text{prox}_\varphi^\Lambda(y)\rangle_\Lambda. \end{cases}$$

Adding those two inequalities, we obtain

$$\|\text{prox}_\varphi^\Lambda(x) - \text{prox}_\varphi^\Lambda(y)\|_\Lambda^2 + \langle y - x, \text{prox}_\varphi^\Lambda(x) - \text{prox}_\varphi^\Lambda(y)\rangle_\Lambda \leq 0.$$

Using the last result, we can easily establish the second assertion

$$\begin{aligned} \|(I - \text{prox}_\varphi^\Lambda)(x) - (I - \text{prox}_\varphi^\Lambda)(y)\|_\Lambda^2 &\leq \|x - y\|_\Lambda^2 - \langle x - y, \text{prox}_\varphi^\Lambda(x) - \text{prox}_\varphi^\Lambda(y)\rangle_\Lambda \\ &= \langle (x - \text{prox}_\varphi^\Lambda(x)) - (y - \text{prox}_\varphi^\Lambda(y)), x - y\rangle_\Lambda, \end{aligned}$$

as desired. □

Let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary and suppose that T is a Λ-firmly nonexpansive operator. Then, Lemma 3.1.1 implies

$$\|T(x) - T(y)\|_\Lambda^2 \leq \|\Lambda^{\frac{1}{2}}(T(x) - T(y))\|_2 \cdot \|\Lambda^{\frac{1}{2}}(x - y)\|_2 = \|T(x) - T(y)\|_\Lambda \cdot \|x - y\|_\Lambda$$

for all $x, y \in \mathbb{R}^n$. Thus, every Λ-firmly nonexpansive mapping is also a Λ-*nonexpansive mapping*, i.e., it is a Lipschitz continuous function with modulus 1 with respect to the norm

$\|\cdot\|_\Lambda$. Moreover, by using

$$\lambda_{\min}(\Lambda)\cdot\|z\|^2 \le \|z\|_\Lambda^2 \le \lambda_{\max}(\Lambda)\cdot\|z\|^2, \quad \forall\, z\in\mathbb{R}^n, \tag{3.1.3}$$

it immediately follows that a Λ-nonexpansive mapping is also Lipschitz continuous with respect to the Euclidean norm with modulus $\sqrt{\lambda_{\max}(\Lambda)/\lambda_{\min}(\Lambda)}$. The following lemma shows that the proximity operator preserves its local Lipschitz continuity when it is treated as a function of the parameter matrix Λ.

Lemma 3.1.2. *Let $\varphi:\ \mathbb{R}^n\to(-\infty,+\infty]$ be a convex, proper, and lower semicontinuous function and let $x\in\mathbb{R}^n$ be arbitrary but fixed. Then, for every compact subset $\mathcal{K}\subset\mathbb{S}^n_{++}$ there exists a constant $L=L(\varphi,x,\mathcal{K})$ such that*

$$\|\mathrm{prox}_\varphi^{\Lambda_1}(x)-\mathrm{prox}_\varphi^{\Lambda_2}(x)\|\le L\cdot\|\Lambda_1-\Lambda_2\|_F, \quad \forall\ \Lambda_1,\Lambda_2\in\mathcal{K}.$$

In particular, the proximity operator $\mathrm{prox}_\varphi(x,\cdot):\mathbb{S}^n\to\mathbb{R}^n$ *is Lipschitz continuous on every compact subset $\mathcal{K}\subset\mathbb{S}^n_{++}$.*

Proof. Let $\mathcal{K}\subset\mathbb{S}^n_{++}$ be a compact set. Then, due to the compactness of $\mathcal{K}$, there exist $\lambda_M\ge\lambda_m>0$ such that $\lambda_M I\succeq\Lambda\succeq\lambda_m I$ for all $\Lambda\in\mathcal{K}$. Let us define $\bar p:=\mathrm{prox}_\varphi^{\lambda_M I}(x)$ and let $\Lambda\in\mathcal{K}$ be arbitrary. Using $\mathrm{env}_\varphi^\Lambda(x)\le\mathrm{env}_\varphi^{\lambda_M I}(x)$ and applying the optimality condition (3.1.2) for $\bar p$, we obtain

$$\begin{aligned}\frac{1}{2}\|\mathrm{prox}_\varphi^\Lambda(x)-x\|^2 &\le \frac{1}{\lambda_m}(\mathrm{env}_\varphi^{\lambda_M I}(x)-\varphi(\mathrm{prox}_\varphi^\Lambda(x))\\ &\le \frac{\lambda_M}{\lambda_m}\left(\langle\bar p-x,\mathrm{prox}_\varphi^\Lambda(x)-\bar p\rangle+\frac{1}{2}\|\bar p-x\|^2\right)\\ &\le \frac{\lambda_M}{\lambda_m}\left(\|\bar p-x\|\|\mathrm{prox}_\varphi^\Lambda(x)-x\|-\frac{1}{2}\|\bar p-x\|^2\right).\end{aligned}$$

Rearranging the terms and solving the resulting inequality for $\|\mathrm{prox}_\varphi^\Lambda(x)-x\|$ yields the following bounds

$$\frac{\lambda_M}{\lambda_m}\left(1-\sqrt{1-\frac{\lambda_m}{\lambda_M}}\right)\cdot\|\bar p-x\|\le\|\mathrm{prox}_\varphi^\Lambda(x)-x\|\le\frac{\lambda_M}{\lambda_m}\left(1+\sqrt{1-\frac{\lambda_m}{\lambda_M}}\right)\cdot\|\bar p-x\|.$$

Thus, since the upper bound does only depend on λ_m, λ_M, and x, the term $\|\mathrm{prox}_\varphi^\Lambda(x)-x\|$ is bounded for all $\Lambda\in\mathcal{K}$. For convenience, let us set

$$C:=C(\varphi,x,\mathcal{K}):=\frac{\lambda_M}{\lambda_m}\left(1+\sqrt{1-\frac{\lambda_m}{\lambda_M}}\right)\cdot\|\bar p-x\|.$$

The remaining arguments of the proof follow the basic ideas and techniques presented in [236, Lemma 3]. An almost identical and a related result can also be found in [210, Lemma 3] and [129, Proposition 3.6], respectively. Now, let $\Lambda_1,\Lambda_2\in\mathcal{K}$ be arbitrary and let us define $p^i:=\mathrm{prox}_\varphi^{\Lambda_i}(x)-x$, $i=1,2$. Then, using characterization (3.1.2) for $x+p^1$ and $x+p^2$, we obtain the following inequalities:

$$\begin{cases} \varphi(x+p^2) - \varphi(x+p^1) \geq -\langle p^1, p^2 - p^1 \rangle_{\Lambda_1}, \\ \varphi(x+p^1) - \varphi(x+p^2) \geq -\langle p^2, p^1 - p^2 \rangle_{\Lambda_2}. \end{cases}$$

Combining those two inequalities, we readily get

$$\langle p^1, p^1 - p^2 \rangle_{\Lambda_1} \leq \langle p^2, p^1 - p^2 \rangle_{\Lambda_2}. \tag{3.1.4}$$

Next, adding the term $-\langle p^2, p^1 - p^2 \rangle_{\Lambda_1}$ on both sides of (3.1.4), it follows

$$\begin{aligned} \|p^1 - p^2\|^2_{\Lambda_1} &\leq \langle p^2, (\Lambda_2 - \Lambda_1)(p^1 - p^2) \rangle \\ &= \langle \Lambda_1^{-\frac{1}{2}}(\Lambda_2 - \Lambda_1) \cdot p^2, \Lambda_1^{\frac{1}{2}}(p^1 - p^2) \rangle \leq \|\Lambda_1^{-\frac{1}{2}}(\Lambda_2 - \Lambda_1) \cdot p^2\| \|p^1 - p^2\|_{\Lambda_1}, \end{aligned}$$

where we used the symmetry and invertibility of Λ_1. Consequently, due to the boundedness of p^2 and $\Lambda_1 \succeq \lambda_m I$, we can infer

$$\|\mathrm{prox}_\varphi^{\Lambda_1}(x) - \mathrm{prox}_\varphi^{\Lambda_2}(x)\| = \|p^1 - p^2\| \leq \frac{\lambda_{\max}(\Lambda_1^{-\frac{1}{2}})}{\sqrt{\lambda_m}} \|p^2\| \cdot \|\Lambda_1 - \Lambda_2\|_F \leq \frac{C}{\lambda_m} \cdot \|\Lambda_1 - \Lambda_2\|_F,$$

as desired. □

Remark 3.1.3. Let us make some comments on Lemma 3.1.2. If $(\Lambda_k)_k \in \mathbb{S}^n_{++}$ is a sequence of matrices that converges to some $\Lambda \in \mathbb{S}^n_{++}$, then Lemma 3.1.2 implies that the proximity operators $\mathrm{prox}_\varphi^{\Lambda_k}(x)$ converge to $\mathrm{prox}_\varphi^{\Lambda}(x)$ for every fixed $x \in \mathbb{R}^n$. In other words, a sequence of minimizers of the functionals $\theta_k(y) := \varphi(y) + \frac{1}{2}\|x - y\|^2_{\Lambda_k}$ converges to a minimizer of the limit functional $\theta(y) := \varphi(y) + \frac{1}{2}\|x - y\|^2_\Lambda$. This quite remarkable property can be traced back to the fact that the sequence of functionals $(\theta_k)_k$ epi- or Γ-converges to θ in our situation. Figure 3.1 illustrates this effect for two different, explicit examples.

The following corollary is an easy consequence of Lemma 3.1.2.

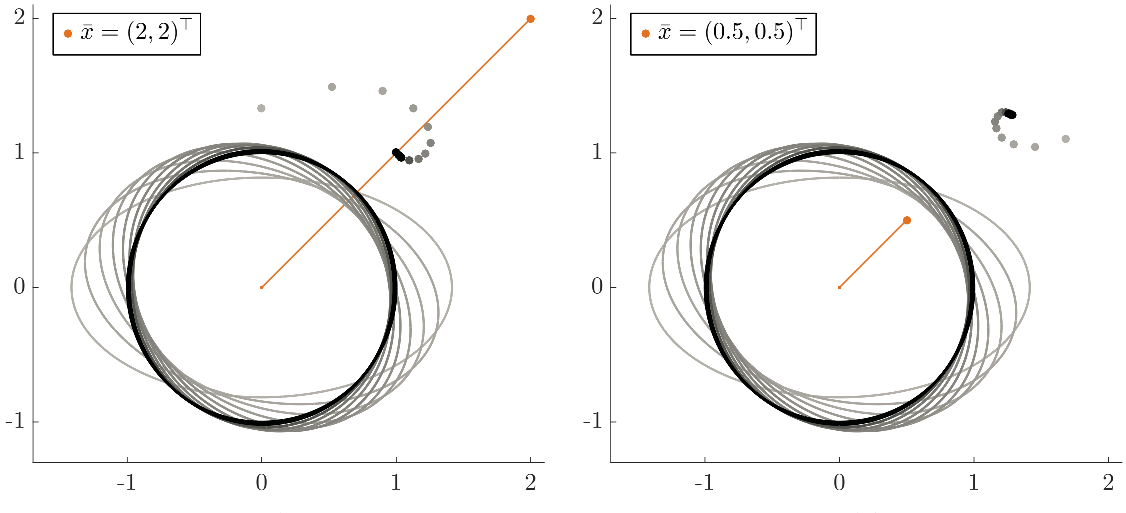
Corollary 3.1.4. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function. Then, the proximity operator $\mathrm{prox}_\varphi : \mathbb{R}^n \times \mathbb{S}^n \to \mathbb{R}^n$ is Lipschitz continuous on every compact subset $\mathcal{K} \subset \mathbb{R}^n \times \mathbb{S}^n_{++}$ and continuous on $\mathbb{R}^n \times \mathbb{S}^n_{++}$.*

Next, we discuss differentiability properties of the Moreau envelope env_φ. Let us note, that smoothness of $\mathrm{env}_\varphi^\Lambda$ was already derived by Moreau in [163, Proposition 7.d]. The proof of Lemma 3.1.5 is based on the proof in [90, Satz 6.38].

Lemma 3.1.5. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function. Then, the Moreau envelope of φ is continuously differentiable on $\mathbb{R}^n \times \mathbb{S}^n_{++}$ and its partial derivatives satisfy*

$$\nabla_x \mathrm{env}_\varphi(x, \Lambda) = \Lambda(x - \mathrm{prox}_\varphi^\Lambda(x)), \quad \nabla_\Lambda \mathrm{env}_\varphi(x, \Lambda) = \frac{1}{2}(x - \mathrm{prox}_\varphi^\Lambda(x))(x - \mathrm{prox}_\varphi^\Lambda(x))^\top$$

for all $(x, \Lambda) \in \mathbb{R}^n \times \mathbb{S}^n_{++}$. Moreover, for every arbitrary but fixed $\Lambda \in \mathbb{S}^n_{++}$, the Moreau envelope $\mathrm{env}_\varphi^\Lambda$ is convex on $\mathbb{R}^n$.



(a) (b)

Figure 3.1.: Illustration of the convergence of a sequence of proximity operators $(\mathrm{prox}_\varphi^{\Lambda_k}(\bar{x}))_k$ for different parameter matrices Λ_k and for two different choices of φ. In subfigure (a) the ℓ_1-regularization $\varphi(x) := \|x\|_1$ was used; in subfigure (b) the so-called *Burg entropy function*

$$\varphi(x) := \sum_{i=1}^{2} f(x_i), \quad f(x_i) := \begin{cases} -\ln(x_i) & \text{if } x_i > 0, \\ +\infty & \text{if } x_i \le 0, \end{cases}$$

was used (see, e.g., [54, Example 2.18]). The parameter matrices Λ_k converge to the identity matrix I. Orange point: fixed reference point. Ellipses: visualization of the different unit disks $\{x \in \mathbb{R}^2 : \|x\|_{\Lambda_k} = 1\}$. Gray points: plot of the corresponding proximity operators $\mathrm{prox}_\varphi^{\Lambda_k}(\bar{x})$.

Proof. Let us define $\theta(x, y, \Lambda) := \varphi(y) + \frac{1}{2}\|x - y\|_\Lambda^2$ and let $[h, H] \in \mathbb{R}^n \times \mathbb{S}^n$ be arbitrary. Then, for all $t > 0$ sufficiently small, the parameter matrix $\Lambda + tH$ is symmetric and positive definite. Thus, the Moreau envelope $\mathrm{env}_\varphi(x + th, \Lambda + tH)$ is well-defined and it holds

$$\begin{aligned} \mathrm{env}_\varphi(x + th, \Lambda + tH) - \mathrm{env}_\varphi(x, \Lambda) &\le \theta(x + th, \mathrm{prox}_\varphi^\Lambda(x), \Lambda + tH) - \theta(x, \mathrm{prox}_\varphi^\Lambda(x), \Lambda) \\ &= \frac{1}{2}\|(x + th) - \mathrm{prox}_\varphi^\Lambda(x)\|_{\Lambda+tH}^2 - \frac{1}{2}\|x - \mathrm{prox}_\varphi^\Lambda(x)\|_\Lambda^2 \\ &= \frac{t}{2}\|x - \mathrm{prox}_\varphi^\Lambda(x)\|_H^2 + t \cdot \langle x - \mathrm{prox}_\varphi^\Lambda(x), h\rangle_\Lambda + O(t^2). \end{aligned}$$

Please note, that the term "$\|\cdot\|_H^2$" is just used as an abbreviation for $\|y\|_H^2 = y^\top Hy$ and does not necessarily correspond to a norm since H is not assumed to be positive definite.

Now, the latter estimate shows

$$\limsup_{t\downarrow 0} \frac{\mathrm{env}_\varphi(x+th,\Lambda+tH)-\mathrm{env}_\varphi(x,\Lambda)}{t} \leq \frac{1}{2}\|x-\mathrm{prox}_\varphi^\Lambda(x)\|_H^2 + \langle \Lambda(x-\mathrm{prox}_\varphi^\Lambda(x)), h\rangle. \tag{3.1.5}$$

Similarly, we can derive a lower bound

$$\begin{aligned}\mathrm{env}_\varphi(x+th,\Lambda+tH)&-\mathrm{env}_\varphi(x,\Lambda)\\ &\geq \theta(x+th,\mathrm{prox}_\varphi^{\Lambda+tH}(x+th),\Lambda+tH)-\theta(x,\mathrm{prox}_\varphi^{\Lambda+tH}(x+th),\Lambda)\\ &= \frac{1}{2}\|(x+th)-\mathrm{prox}_\varphi^{\Lambda+tH}(x+th)\|_{\Lambda+tH}^2-\frac{1}{2}\|x-\mathrm{prox}_\varphi^{\Lambda+tH}(x+th)\|_\Lambda^2\\ &= \frac{t}{2}\|x-\mathrm{prox}_\varphi^{\Lambda+tH}(x+th)\|_H^2+t\cdot\langle x-\mathrm{prox}_\varphi^{\Lambda+tH}(x+th),h\rangle_\Lambda+O(t^2).\end{aligned}$$

This establishes

$$\liminf_{t\downarrow 0} \frac{\mathrm{env}_\varphi(x+th,\Lambda+tH)-\mathrm{env}_\varphi(x,\Lambda)}{t} \geq \frac{1}{2}\|x-\mathrm{prox}_\varphi^\Lambda(x)\|_H^2 + \langle \Lambda(x-\mathrm{prox}_\varphi^\Lambda(x)), h\rangle, \tag{3.1.6}$$

where we used the continuity of the proximity operator prox_φ. Combining (3.1.5), (3.1.6), and

$$\|x-\mathrm{prox}_\varphi^\Lambda(x)\|_H^2 = \mathrm{tr}((x-\mathrm{prox}_\varphi^\Lambda(x))(x-\mathrm{prox}_\varphi^\Lambda(x))^\top H),$$

it follows that the Moreau envelope of φ is directionally differentiable at (x,Λ) in the direction $[h,H]$ and its derivative is given by

$$\mathrm{env}_\varphi'(x,\Lambda;[h,H]) = \langle\Lambda(x-\mathrm{prox}_\varphi^\Lambda(x)),h\rangle+\frac{1}{2}\mathrm{tr}((x-\mathrm{prox}_\varphi^\Lambda(x))(x-\mathrm{prox}_\varphi^\Lambda(x))^\top H).$$

Since the direction $[h,H]\in\mathbb{R}^n\times\mathbb{S}^n$ was arbitrary and $\mathrm{env}_\varphi'(x,\Lambda;[h,H])$ is linear and continuous in $[h,H]$, the Moreau envelope env_φ is Fréchet differentiable and the gradient of env_φ satisfies

$$\nabla_x\mathrm{env}_\varphi(x,\Lambda)=\Lambda(x-\mathrm{prox}_\varphi^\Lambda(x)),\quad \nabla_\Lambda\mathrm{env}_\varphi(x,\Lambda)=\frac{1}{2}(x-\mathrm{prox}_\varphi^\Lambda(x))(x-\mathrm{prox}_\varphi^\Lambda(x))^\top.$$

Furthermore, using the continuity of the proximity operator prox_φ, we infer that env_φ is even continuously differentiable on $\mathbb{R}^n\times\mathbb{S}^n_{++}$. To prove the convexity of the Moreau envelope, we use the fact, that a continuously differentiable function is convex if and only if its gradient is a monotone mapping. Hence, let $\Lambda\in\mathbb{S}^n_{++}$ be fixed and let $x,y\in\mathbb{R}^n$ be arbitrary. Then, applying Lemma 3.1.1, it follows

$$\langle\nabla_x\mathrm{env}_\varphi(x,\Lambda)-\nabla_x\mathrm{env}_\varphi(y,\Lambda),x-y\rangle=\langle(x-\mathrm{prox}_\varphi^\Lambda(x))-(y-\mathrm{prox}_\varphi^\Lambda(y)),x-y\rangle_\Lambda\geq 0.$$

Consequently, we conclude that $\mathrm{env}_\varphi^\Lambda:\mathbb{R}^n\to\mathbb{R}$ is a convex function. □

Remark 3.1.6. We will often use the notation

$$\nabla \mathrm{env}_\varphi^\Lambda(x) := \nabla_x \mathrm{env}_\varphi(x, \Lambda)$$

to denote the partial derivative $\nabla_x \mathrm{env}_\varphi(x, \Lambda)$ or the gradient of $\mathrm{env}_\varphi^\Lambda$, when the parameter matrix Λ is fixed. Moreover, rewriting the optimality condition (3.1.2), we obtain the following, useful property of the gradient of the Moreau envelope

$$\nabla \mathrm{env}_\varphi^\Lambda(x) \in \partial\varphi(\mathrm{prox}_\varphi^\Lambda(x)). \tag{3.1.7}$$

3.2. Proximal calculus and examples

Next, we present some basic tools and rules for the computation of proximity operators. In particular, we will also discuss the *decomposition principle* of Moreau that establishes a link between the proximity operator of a function and the proximity operator of its corresponding convex conjugate.

Lemma 3.2.1. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function and let $x \in \mathbb{R}^n$, $\Lambda \in \mathbb{S}^n_{++}$ be given. Then, the following hold:*

(i) *Let us define $\psi(\cdot) := \varphi(\cdot - b)$, $b \in \mathbb{R}^n$. Then, it follows $\mathrm{prox}_\psi^\Lambda(x) = b + \mathrm{prox}_\varphi^\Lambda(x - b)$.*

(ii) *Let us define $\psi(\cdot) := \varphi(\cdot/\rho)$, $\rho \in \mathbb{R}^n \backslash \{0\}$. Then, it follows $\mathrm{prox}_\psi^\Lambda(x) = \rho\, \mathrm{prox}_\varphi^{\rho^2\Lambda}(x/\rho)$.*

Proof. The proof is exactly as in [54, Lemma 2.6]. □

The following composition formula extends Theorem 3.1 in [147] to the class of real extended valued functions and to proximity operators with matrix parameters.

Lemma 3.2.2. *Let $\varphi : \mathbb{R}^m \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function and let $x \in \mathbb{R}^n$, $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. Furthermore, let $A \in \mathbb{R}^{m \times n}$ be given and suppose that one of the following regularity conditions*

(i) $0 \in \mathrm{int}\{A\mathbb{R}^n - \mathrm{dom}\ \varphi\}$

(ii) $A\mathbb{R}^n \cap \mathrm{ri\ dom}\ \varphi \neq \emptyset$

is satisfied. Set $\psi(x) := \varphi(Ax)$, then it holds

$$\mathrm{prox}_\psi^\Lambda(x) = x - \Lambda^{-1} A^\top \Gamma v, \quad \Gamma \in \mathbb{S}^m_{++}, \tag{3.2.1}$$

if and only if v is a fixed point of the operator $(I - \mathrm{prox}_\varphi^\Gamma) \circ H_x$, where $H_x : \mathbb{R}^n \to \mathbb{R}^n$, $H_x(v) := Ax + (I - A\Lambda^{-1}A^\top\Gamma)v$.

Proof. The proof is a simple extension of the proof in [147]. For the sake of completeness, we adapt the proof given in [147] for the more general setting of Lemma 3.2.2. Due to (3.1.2) and Lemma 2.5.15, it follows

$$x - \mathrm{prox}_\psi^\Lambda(x) \in \Lambda^{-1} \cdot \partial\psi(\mathrm{prox}_\psi^\Lambda(x)) = \Lambda^{-1} \cdot A^\top \partial\varphi(A\mathrm{prox}_\psi^\Lambda(x)).$$

Hence, for any arbitrary parameter matrix $\Gamma \in \mathbb{S}^m_{++}$, there exists $v \in \Gamma^{-1} \cdot \partial\varphi(A\mathrm{prox}^\Lambda_\psi(x))$ such that

$$\mathrm{prox}^\Lambda_\psi(x) = x - \Lambda^{-1} A^\top \Gamma v$$

and we obtain

$$\begin{aligned} v \in \Gamma^{-1} \cdot \partial\varphi(A\mathrm{prox}^\Lambda_\psi(x)) \quad &\iff \quad A\mathrm{prox}^\Lambda_\psi(x) \in A\mathrm{prox}^\Lambda_\psi(x) + v - \Gamma^{-1} \cdot \partial\varphi(A\mathrm{prox}^\Lambda_\psi(x)) \\ &\iff \quad A\mathrm{prox}^\Lambda_\psi(x) = \mathrm{prox}^\Gamma_\varphi(A\mathrm{prox}^\Lambda_\psi(x) + v) \\ &\iff \quad Ax - A\Lambda^{-1}A^\top\Gamma v = \mathrm{prox}^\Gamma_\varphi(H_x(v)) \\ &\iff \quad v = (I - \mathrm{prox}^\Gamma_\varphi)(H_x(v)). \end{aligned}$$

Conversely, assume that v is a fixed point of the mapping $(I - \mathrm{prox}^\Gamma_\varphi) \circ H_x$, then the above argumentation establishes the following implication:

$$\begin{aligned} &v \in \Gamma^{-1} \cdot \partial\varphi(Ax - A\Lambda^{-1}A^\top\Gamma v) \\ &\qquad\qquad \implies \quad A^\top\Gamma v \in A^\top \partial\varphi(A(x - \Lambda^{-1}A^\top\Gamma v)) = \partial\psi(x - \Lambda^{-1}A^\top\Gamma v). \end{aligned}$$

Obviously, the last inclusion can be rearranged such that optimality condition (3.1.2) is again applicable. This finally yields $\mathrm{prox}^\Lambda_\psi(x) = x - \Lambda^{-1}A^\top\Gamma v$. □

Remark 3.2.3. If the matrices A and Λ satisfy $A\Lambda^{-1}A^\top \in \mathbb{S}^n_{++}$, then both regularity conditions in Lemma 3.2.2 are fulfilled and equation (3.2.1) can be simplified to

$$\mathrm{prox}^\Lambda_{\varphi\circ A}(x) = x - \Lambda^{-1}A^\top(A\Lambda^{-1}A^\top)^{-1}(Ax - \mathrm{prox}^{(A\Lambda^{-1}A^\top)^{-1}}_\varphi(Ax)).$$

In addition, if A is orthogonal, i.e., if it holds $A^\top A = AA^\top = I \in \mathbb{R}^{n\times n}$, then the above formula reduces to

$$\mathrm{prox}^\Lambda_{\varphi\circ A}(x) = A^\top \mathrm{prox}^{A\Lambda A^\top}_\varphi(Ax).$$

In particular, in the case $\Lambda = \lambda I$, we recover the two well-known composition formulae

$$\mathrm{prox}^{\lambda I}_{\varphi\circ A}(x) = x - A^\top(AA^\top)^{-1}(Ax - \mathrm{prox}^{\lambda(AA^\top)^{-1}}_\varphi(Ax))$$

and

$$\mathrm{prox}^{\lambda I}_{\varphi\circ A}(x) = A^\top \mathrm{prox}^{\lambda I}_\varphi(Ax),$$

respectively.

Lemma 3.2.4. *Let $(\mathcal{I}_k)_{k=1,...,N} \subset \{1, ..., n\}$ be a sequence of N distinct sets such that*

$$\bigcup_{k=1}^N \mathcal{I}_k = \{1, ..., n\} \quad \text{and} \quad n_k := |\mathcal{I}_k| \neq 0, \quad \forall\, k = 1, ..., N.$$

Furthermore, for $k = 1, ..., N$, let $(\varphi_k)_{k=1,...,N}$, $\varphi_k : \mathbb{R}^{n_k} \to (-\infty, +\infty]$, be a family of convex, proper, and lower semicontinuous functions and let $\Lambda_k \in \mathbb{S}^{n_k}_{++}$ be given. Let us define

$\varphi(x) := \sum_{k=1}^{N} \ \varphi_k(x_{\mathcal{I}_k})$ and $\Lambda \in \mathbb{R}^{n\times n}$,

$$\Lambda_{[\mathcal{I}_k\mathcal{I}_\ell]} := \begin{cases} \Lambda_k & \textit{if } k = \ell, \\ 0 & \textit{if } k \neq \ell, \end{cases} \qquad 1 \le k, \ell \le N.$$

Then, it holds

$$\text{prox}_{\varphi}^{\Lambda}(x)_{\mathcal{I}_k} = \text{prox}_{\varphi_k}^{\Lambda_k}(x_{\mathcal{I}_k}), \quad \forall\ k = 1, ..., N. \tag{3.2.2}$$

Proof. We only give a sketch of the proof and refer to [54, Lemma 2.9] for further details. Clearly, φ is a convex, proper, and lower semicontinuous function and we have $\Lambda \in \mathbb{S}^n_{++}$. Now, using

$$\partial\varphi(x_{\mathcal{I}_1}, ..., x_{\mathcal{I}_N}) = \partial\varphi_1(x_{\mathcal{I}_1}) \times ... \times \partial\varphi_N(x_{\mathcal{I}_N}),$$

(see, e.g., [11, Proposition 16.8] or [54, Lemma 2.1]), the statement (3.2.2) easily follows from a block-wise application of optimality condition (3.1.2). □

Theorem 3.2.5 (Moreau's decomposition principle). *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function and let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. Then, for all $x \in \mathbb{R}^n$, it holds*

$$x = \text{prox}_{\varphi}^{\Lambda}(x) + \Lambda^{-1} \cdot \text{prox}_{\varphi^*}^{\Lambda^{-1}}(\Lambda x), \tag{3.2.3}$$

and

$$\varphi(\text{prox}_{\varphi}^{\Lambda}(x)) + \varphi^*(\text{prox}_{\varphi^*}^{\Lambda^{-1}}(\Lambda x)) = \langle \text{prox}_{\varphi}^{\Lambda}(x), \text{prox}_{\varphi^*}^{\Lambda^{-1}}(\Lambda x)\rangle, \tag{3.2.4}$$

where φ^* denotes the convex conjugate of φ.

Proof. A proof of (3.2.3) can be found in [14, Lemma 5−B.1]. To prove the second assertion, we proceed as in [54, Lemma 2.10]. By reformulating the equation (3.2.3) and using (3.1.7), we obtain

$$\text{prox}_{\varphi^*}^{\Lambda^{-1}}(\Lambda x) = \Lambda(x - \text{prox}_{\varphi}^{\Lambda}(x)) = \nabla \text{env}_{\varphi}^{\Lambda}(x) \in \partial\varphi(\text{prox}_{\varphi}^{\Lambda}(x)).$$

Due to Lemma 2.5.14, this is equivalent to (3.2.4). □

The following result was established by Yu in [261].

Theorem 3.2.6. *Let $\varphi_1, \varphi_2 : \mathbb{R}^n \to (-\infty, +\infty]$ be two convex, proper and lower semicontinuous functions and let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. If φ_1 and φ_2 satisfy*

$$\partial\varphi_2(x) \subseteq \partial\varphi_2(\text{prox}_{\varphi_1}^{\Lambda}(x)), \quad \forall\ x \in \mathbb{R}^n, \tag{3.2.5}$$

then the following composition rule holds for all $x \in \mathbb{R}^n$:

$$\text{prox}_{\varphi_1+\varphi_2}^{\Lambda}(x) = (\text{prox}_{\varphi_1}^{\Lambda} \circ \text{prox}_{\varphi_2}^{\Lambda})(x). \tag{3.2.6}$$

Proof. The proof is just a minor, mainly "notational" extension of the proof of Theorem 1 in [261]. Therefore, we will omit the proof. □

Examples

In this subsection we want to derive some explicit formulae for proximity operators by using the calculation rules we have just presented. We mainly concentrate on examples that will be relevant in the subsequent sections or in our numerical considerations at the end of this thesis. More computational results and a broader overview of explicitly known proximity operators can be found in [54, 6, 53].

Example 3.2.7 (Classical projection). Let $K \subset \mathbb{R}^n$ be a convex, nonempty, and closed set and let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. Then, it holds

$$\operatorname{prox}^{\Lambda}_{\iota_K}(x) = \operatorname*{arg\,min}_{y \in \mathbb{R}^n} \ \iota_K(y) + \frac{1}{2}\|x-y\|^2_\Lambda = \operatorname*{arg\,min}_{y \in K} \ \frac{1}{2}\|x-y\|^2_\Lambda =: \mathcal{P}^{\Lambda}_K(x).$$

In particular, if $\Lambda = \lambda I$ for some $\lambda > 0$, then the proximity operator of the indicator function ι_K coincides with the classical, orthogonal projection onto the set K.

Example 3.2.8 (Norms and homogeneous functions). Let $|||\cdot||| : \mathbb{R}^n \to \mathbb{R}$ be a norm on $\mathbb{R}^n$ and let $\mu > 0$, $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. Then, by applying Theorem 3.2.5 and Example 2.2.5, we obtain

$$\text{(3.2.7)} \qquad \operatorname{prox}^{\Lambda}_{\mu|||\cdot|||}(x) = x - \mu\Lambda^{-1} \cdot \operatorname{prox}^{\mu\Lambda^{-1}}_{|||\cdot|||^*}(\Lambda x/\mu) = x - \Lambda^{-1} \cdot \mathcal{P}^{\Lambda^{-1}}_{B_{|||\cdot|||_\circ}(0,\mu)}(\Lambda x).$$

In the special case $\Lambda = \frac{1}{\lambda}I$, $\lambda > 0$, formula (3.2.7) can be used to derive and calculate a large number of important proximity operators. For instance, if we consider the ℓ_1- or ℓ_2-norm, then (3.2.7) reduces to the two well-known *shrinkage operators*

$$\operatorname{prox}_{\mu\lambda\|\cdot\|_1}(x) = \operatorname{prox}^{\frac{1}{\lambda}I}_{\mu\|\cdot\|_1}(x) = x - \mathcal{P}_{[-\mu\lambda,\mu\lambda]}(x) = \operatorname{sign}(x) \odot \max\{|x| - \mu\lambda, 0\},$$

$$\operatorname{prox}_{\mu\lambda\|\cdot\|_2}(x) = \operatorname{prox}^{\frac{1}{\lambda}I}_{\mu\|\cdot\|_2}(x) = x - \mathcal{P}_{B_{\|\cdot\|_2}(0,\mu\lambda)}(x) = \frac{x}{\|x\|_2} \cdot \max\{\|x\|_2 - \mu\lambda, 0\},$$

where the application of the sign function $\operatorname{sign}(\cdot)$ and the absolute value $|\cdot|$ is understood component-wise. Now, suppose that $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, proper, lower semicontinuous, and positively homogeneous function. Using Lemma 2.5.13, Theorem 2.2.3 and 3.2.5, we obtain the following extension of formula (3.2.7)

$$\text{(3.2.8)} \qquad \operatorname{prox}^{\Lambda}_{\mu\varphi}(x) = x - \mu\Lambda^{-1} \cdot \operatorname{prox}^{\mu\Lambda^{-1}}_{\sigma^*_{\partial\varphi(0)}}(\Lambda x/\mu) = x - \Lambda^{-1} \cdot \mathcal{P}^{\Lambda^{-1}}_{\mu\partial\varphi(0)}(\Lambda x).$$

Example 3.2.9 (Constraints and ℓ_1-norm). Let $a, b \in [-\infty, +\infty]^n$ be such that $a_i \leq b_i$ for all $i = 1, ..., n$. Additionally, let us exclude the degenerate situations $a_i = b_i = -\infty$ or $a_i = b_i = +\infty$, i.e., let us suppose that the sets $[a_i, b_i] \cap \mathbb{R}$ are nonempty for all i. Now, let $\lambda \in \mathbb{R}^n_{++}$ and $\mu > 0$ be arbitrary and let us set $\Lambda = \operatorname{diag}(1 \oslash \lambda) \in \mathbb{S}^n_{++}$. The aim of this example is to compute the proximity operator of the composite function

$$\varphi(x) := \mu\|x\|_1 + \iota_{[a,b]}(x).$$

Clearly, we are in the setting of Lemma 3.2.4, i.e., the proximity operator $\operatorname{prox}_{\varphi}^{\Lambda}(x)$ can be computed component-wise. Moreover, due to

$$\partial|x_i| = \begin{cases} \{+1\} & \text{if } x_i > 0, \\ [-1,+1] & \text{if } x_i = 0, \\ \{-1\} & \text{if } x_i < 0, \end{cases} \quad \operatorname{prox}_{\iota_{[a_i,b_i]}}^{1/\lambda_i}(x_i) = \mathcal{P}_{[a_i,b_i]}(x_i) = \begin{cases} b_i & \text{if } x_i > b_i, \\ x_i & \text{if } x_i \in [a_i,b_i] \cap \mathbb{R}, \\ a_i & \text{if } x_i < a_i, \end{cases}$$

we immediately see that Theorem 3.2.6 is applicable when $0 \in [a_i, b_i]$. Since the absolute value reduces to a differentiable function when $a_i > 0$ or $b_i < 0$, the proximity operator $\operatorname{prox}_{\mu|\cdot|+\iota_{[a_i,b_i]}}^{1/\lambda_i}(x_i)$ can be computed directly in these cases; it holds:

$$\operatorname{prox}_{\mu|\cdot|+\iota_{[a_i,b_i]}}^{1/\lambda_i}(x_i) = \begin{cases} \mathcal{P}_{[a_i,b_i]}(x_i - \mu\lambda_i) & \text{if } a_i > 0, \\ \mathcal{P}_{[a_i,b_i]}(\operatorname{prox}_{\mu|\cdot|}^{1/\lambda_i}(x_i)) & \text{if } 0 \in [a_i,b_i], \\ \mathcal{P}_{[a_i,b_i]}(x_i + \mu\lambda_i) & \text{if } b_i < 0. \end{cases}$$

In summary, after some more (easy) manipulations, we obtain the composition formula

$$\operatorname{prox}_{\varphi}^{\Lambda}(x) = \mathcal{P}_{[a,b]}(\operatorname{prox}_{\mu\|\cdot\|_1}^{\Lambda}(x)) = \mathcal{P}_{[a,b]}(x - \mathcal{P}_{[-\mu\lambda,\mu\lambda]}(x)).$$

The following example concludes this subsection and is a bit more sophisticated.

Example 3.2.10 (Epigraphical projection). We want to analyze the proximity operator that is characterized by the optimization problem

$$\min_{(y,\gamma)\in\mathbb{R}^n\times\mathbb{R}} \frac{\alpha}{2}\|x-y\|_2^2 + \frac{\beta}{2}(t-\gamma)^2 \quad \text{s.t.} \quad \|Ay-b\|_2 \le \gamma, \tag{3.2.9}$$

where $\alpha, \beta > 0$, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$ are given. Clearly, if $\alpha = \beta$, then the optimal solution of the latter minimization problem is given by the projection onto the set $K := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \|Ax-b\|_2 \le t\}$. Projections of this type usually occur as subproblems or subroutines in so-called *basis pursuit* problems which have the following form:

$$\min_{x\in\mathbb{R}^n} \Omega(x) \quad \text{s.t.} \quad \|Ax-b\|_2 \le \sigma. \tag{3.2.10}$$

Here, $\Omega : \mathbb{R}^n \to \mathbb{R}$ is a general sparsity-inducing penalization and $\sigma > 0$ estimates the level of noise in the (possibly noisy) measurements b. Later, in our numerical comparison in chapter 7, we will use the computational results of this example to derive an *Alternating Direction Method of Multipliers (ADMM)* for an optimization problem of the form (3.2.10) with a group-sparse penalty function. Now, defining

$$\Lambda := \begin{pmatrix} \alpha I_n & 0 \\ 0 & \beta \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}, \quad B = \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{(m+1)\times(n+1)},$$

and

$$\varphi(y,\gamma) := \iota_{\text{epi } \|\cdot\|_2}(y - b, \gamma),$$

we see that the optimal solution of (3.2.9) is given by the proximity operator $\mathrm{prox}^{\Lambda}_{\varphi\circ B}(x,t)$. Moreover, if the matrix A satisfies $AA^\top = I$, then, using Lemma 3.2.1 (i) and Remark 3.2.3, we can derive a closed form formula for this proximity operator. In particular, it holds

$$\mathrm{prox}^{\Lambda}_{\varphi\circ B}(x,t) = \begin{pmatrix} x \\ t \end{pmatrix} - \begin{pmatrix} A^\top & 0 \\ 0 & 1 \end{pmatrix} \left(\begin{pmatrix} Ax - b \\ t \end{pmatrix} - \mathcal{P}^{\Lambda}_{\mathrm{epi}\ \|\cdot\|_2}(Ax-b,t) \right)$$

and

$$\mathcal{P}^{\Lambda}_{\mathrm{epi}\ \|\cdot\|_2}(y,\gamma) = \begin{cases} (y,\gamma) & \text{if } \|y\|_2 \le \gamma, \\ \frac{\alpha\|y\|_2+\beta\gamma}{2\|y\|_2}(y,\|y\|_2) & \text{if } -\frac{\alpha}{\beta}\|y\|_2 \le \gamma \le \|y\|_2, \\ (0,0) & \text{if } \gamma \le -\frac{\alpha}{\beta}\|y\|_2. \end{cases}$$

A detailed verification of the epigraph formula can be found, e.g., in [87, Proposition 3.3].

3.3. Semismoothness and second order properties

As motivated in our introduction, proximity operators play an essential role in deriving first order necessary conditions for optimization problems of the form

$$\min_{x\in\mathbb{R}^n} \; f(x) + \varphi(x), \tag{3.3.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is smooth and $\varphi : \mathbb{R}^n \to (-\infty,+\infty]$ is a convex, proper, and lower semicontinuous function. In particular, as we will see in the next chapter, the proximal framework can be utilized to reformulate these optimality conditions as nonsmooth, proximal-based equations that include the proximity operator of the mapping φ. Consequently, this subsection is dedicated to analyze the semismoothness and the second order properties of the proximity operator $\mathrm{prox}^{\Lambda}_{\varphi}$ and to investigate whether the semismooth Newton method can be applied to solve the problem (3.3.1).

Unfortunately, in general, the proximity operator cannot be expected to be a semismooth function. More specifically, Kruskal [127], and Shapiro [214], constructed a three- and two-dimensional example of a convex, closed set K such that the corresponding projection operator $\mathcal{P}_K(x)$ is not directionally differentiable at some point $x \notin K$.

Although the analysis of the differentiability properties of the metric projection $\mathcal{P}_K$ is a very classical field of research, see, e.g., [266, 80, 213, 212], general results that establish and guarantee the semismoothness of the projection $\mathcal{P}_K$ are rather limited and often necessitate a special and specific structure of the convex, closed set $K \subset \mathbb{R}^n$. For instance, in [75], Facchinei and Pang studied the metric projection onto sets of the form

$$K := \{x \in \mathbb{R}^n : g(x) \le 0\}, \tag{3.3.2}$$

where each component function $g_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, ..., m$, is assumed to be convex and twice continuously differentiable. Under the *Sequentially Bounded Constraint Qualification* (SBCQ) and the *Constant Rank Constraint Qualification* (CRCQ), Facchinei and Pang established directional differentiability and piecewise smoothness of the projection operator

$\mathcal{P}_K$, respectively. Moreover, in a similar fashion and applying the CRCQ, Sun and Han [228] and Mifflin et al. [149] have shown that the proximity operator of a piecewise C^2-function and of the maximum of a finite collection of convex C^2-functions is again a piecewise smooth and hence, a semismooth function. Extensions using weaker constraint qualifications were considered, e.g., in [146]. For a more detailed discussion of the metric projection, we refer to [75, Chapter 4] and the references therein.

In general, if the nonsmooth function φ does not possess a certain piecewise structure, much fewer results are available. However, Meng, Sun, and Zhao [145] showed that, under a mild regularity condition, semismoothness of the proximity operator $\mathrm{prox}_\varphi^{\lambda I}$ can be traced back to semismoothness of the metric projection onto the epigraph of φ. In [146], this result was further refined and it was shown that piecewise smoothness of the epigraphical projection $\mathcal{P}_{\mathrm{epi}\,\varphi}$ implies piecewise smoothness of the corresponding proximity operator. (See also the recent work [48] for another connection between the proximity operator $\mathrm{prox}_\varphi^{\lambda I}$ and the epigraphical projection $\mathcal{P}_{\mathrm{epi}\,\varphi}$). Thus, the analysis of the semismoothness of the proximity operator can be completely shifted to a respective investigation of the projection operator $\mathcal{P}_{\mathrm{epi}\,\varphi}$. Unfortunately, in many situations, the epigraph epi φ will not be representable as a set of the form (3.3.2) with a smooth function g and the known results for metric projections are not applicable. Nonetheless, these different results and observations initiated a “renewed” discussion and an intensive study of differentiability properties of certain epigraphical projections. In particular, in the area of low-rank matrix optimization and matrix cone programming, new and profound results were established by Ding et al. [63, 64]; see also [118, 44, 119] for recent applications.

In the following, we will briefly introduce the classes of so-called *semialgebraic* and *tame* functions. In their seminal work [21], Bolte, Daniilidis and Lewis, showed that semialgebraic and tame functions are semismooth. Moreover, these two classes of functions provide an extensive calculus and, remarkably, it also follows that the proximity operator of a semialgebraic function is an α-order semismooth function for some $\alpha > 0$. The material in the following subsection is primarily based on [114, 21] and on the observations in [63].

Semialgebraic and tame functions

We will now sketch the main definitions and theorems for tame functions and present some basic calculation rules for semialgebraic functions. More details can be found in [114, 21]. Furthermore, for a more algebraic and geometric interpretation of semialgebraic functions and tameness, we refer to the work of van den Dries and Coste [244, 57, 58].

Definition 3.3.1 (o-minimal structure, cf. [57, 21]). *An* o-minimal structure *on* $(\mathbb{R}, +, \cdot)$ *is a sequence* $\mathcal{O} = (\mathcal{O}_n)_n$ *of collections* $\mathcal{O}_n$*,* $n \in \mathbb{N}$*, of* definable subsets *of* $\mathbb{R}^n$ *satisfying the following axioms*

(i) *For every* $n \in \mathbb{N}$*, the collection* $\mathcal{O}_n$ *is closed under Boolean operations (finite intersections, unions, and complements).*

(ii) *If* $A \in \mathcal{O}_n$ *and* $B \in \mathcal{O}_m$*, then* $A \times B$ *belongs to* $\mathcal{O}_{n+m}$*.*

(iii) *If* $\pi : \mathbb{R}^{n+1} \to \mathbb{R}$*,* $\pi(x_1, ..., x_n, x_{n+1}) := (x_1, ..., x_n)$*, is the canonical projection onto* $\mathbb{R}^n$*, then for any set* $A \in \mathcal{O}_{n+1}$ *it holds* $\pi(A) \in \mathcal{O}_n$*.*

(iv) *$\mathcal{O}_n$ contains the family of algebraic subsets of $\mathbb{R}^n$. In particular, every set of the form $\{x \in \mathbb{R}^n : p(x) = 0\}$, where $p : \mathbb{R}^n \to \mathbb{R}$ is a polynomial function, belongs to $\mathcal{O}_n$.*

(v) *The elements of $\mathcal{O}_1$ are exactly the finite unions of open intervals and points.*

A mapping $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is said to be definable *in $\mathcal{O}$ if its graph is definable in $\mathcal{O}$ as a subset of $\mathbb{R}^n \times \mathbb{R}^m$.*

Definition 3.3.2 (Tame functions, cf. [21, Definition 2]). *A set $A \subset \mathbb{R}^n$ is called* tame *if for every $r > 0$, there exists an o-minimal structure $\mathcal{O}$ over $(\mathbb{R}, +, \cdot)$, such that the intersection $A \cap [-r, r]^n$ is definable in this structure. A mapping $F : U \subset \mathbb{R}^n \to \mathbb{R}^m$ is called* tame *if its graph* gra F *is tame as a subset of $\mathbb{R}^n \times \mathbb{R}^m$.*

Let us mention that a tame function is not necessarily definable in an o-minimal structure $\mathcal{O}$. For instance, the sine function, $\sin : \mathbb{R} \to \mathbb{R}$, is tame but not definable in every o-minimal structure. This can be easily seen by noticing that the set $\pi(\text{gra} \sin \cap \mathbb{R} \times \{0\})$ violates part (v) of Definition 3.3.1.

The class $\mathcal{SA}$ of so-called *semialgebraic* objects is of special interest since it forms the smallest o-minimal structure on $(\mathbb{R}, +, \cdot)$, see [57, Exercise 1.7]. Here, a set $A \subset \mathbb{R}^n$ is said to be *semialgebraic* if it can be written as a finite intersection and union of polynomial sets, i.e., if

$$A = \bigcup_{j=1}^{p} \bigcap_{i=1}^{q} \{x \in \mathbb{R}^n : p_{ij}(x) = 0,\ q_{ij}(x) < 0\}, \quad p, q \in \mathbb{N},$$

where $p_{ij}, q_{ij} : \mathbb{R}^n \to \mathbb{R}$ are polynomial functions on $\mathbb{R}^n$. A mapping is called *semialgebraic* if its graph is semialgebraic. Let us note that for semialgebraic objects, the projection axiom (iii) in Definition 3.3.1 is a consequence of the *Tarski-Seidenberg principle* [244, 58]. Before we discuss the connection between semialgebraic functions, semismoothness, and the proximity operator, we want to state some basic properties of semialgebraic sets and functions. Clearly, by Definition 3.3.1, the finite union, intersection, and difference of semialgebraic sets is also a semialgebraic set. Besides, it holds:

- The closure, the interior, and the boundary of a semialgebraic set are semialgebraic.
- Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a semialgebraic mapping and let $B \subset \mathbb{R}^m$ be a semialgebraic set. Then, the set $F^{-1}(B)$ is semialgebraic.
- The sum and composition of semialgebraic functions is again a semialgebraic function.
- Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be semialgebraic, then the function $\varphi : \mathbb{R}^n \to \mathbb{R}$, $\varphi(x) := \inf_{y \in \mathbb{R}^m} f(x, y)$ is also semialgebraic.

The proof of these properties heavily relies on the fact that the class $\mathcal{SA}$ is an o-minimal structure. In particular, item (iii) in Definition 3.3.1 will constantly be used to derive and formulate rather elegant proofs. To facilitate the understanding of the concept and of the overall mechanism of semialgebraic sets and mappings, we want to briefly verify the latter three statements.

Proof. First of all, by induction, it easily follows that if $A \subset \mathbb{R}^{n+m}$ is a semialgebraic set and if

$$\pi : \mathbb{R}^{n+m} \to \mathbb{R}^n, \quad \pi(z_1, ..., z_n, z_{n+1}, ..., z_{n+m}) := (z_1, ..., z_n),$$

is the projection onto the first n components of $z \in \mathbb{R}^{n+m}$, then $\pi(A)$ is also a semialgebraic set. (Actually, this is one of several, basic formulations of the Tarski-Seidenberg Theorem). Now, let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a semialgebraic function and let $B \in \mathbb{R}^m$ be a semialgebraic set. Then, the set $F^{-1}(B)$ can be rewritten as follows

$$F^{-1}(B) = \pi(\text{gra } F \cap \mathbb{R}^n \times B), \quad \pi(z_1, ..., z_n, z_{n+1}, ..., z_{n+m}) := (z_1, ..., z_n).$$

Thus, the preimage $F^{-1}(B)$ is semialgebraic. Next, let us suppose that $G : \mathbb{R}^n \to \mathbb{R}^m$ is another semialgebraic mapping. Then, it holds

$$\text{gra } F + G = \pi(\mathcal{B}_1 \cap \mathcal{B}_2 \cap \mathcal{B}_3), \quad \pi(z_1, ..., z_{n+m}, z_{n+m+1}, ..., z_{n+3m}) := (z_1, ..., z_{n+m}),$$

where the sets $\mathcal{B}_i$, $i = 1, 2, 3$, are given by $\mathcal{B}_1 := \{(x, y, u, v) \in \mathbb{R}^{n+3m} : y = u + v\}$, $\mathcal{B}_2 := \{(x, y, u, v) \in \mathbb{R}^{n+3m} : (x, u) \in \text{gra } F\}$, and $\mathcal{B}_3 := \{(x, y, u, v) \in \mathbb{R}^{n+3m} : (x, v) \in \text{gra } G\}$. Since the sets $\mathcal{B}_2$, $\mathcal{B}_3$ are semialgebraic and $\mathcal{B}_1$ is an algebraic set, it follows that $F + G$ is a semialgebraic mapping. The argumentation for the composition of two functions is similar. Specifically, if $H : \mathbb{R}^p \to \mathbb{R}^n$ is a semialgebraic function, then the graph of $F \circ H$ can be written as

$$\text{gra } F \circ H = \pi(\{(x, y, z) \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^n : (x, z) \in \text{gra } H,\ (z, y) \in \text{gra } F\})$$

with $\pi(z_1, ..., z_{p+m}, z_{p+m+1}, ..., z_{p+m+n}) := (z_1, ..., z_{p+m})$. Hence, the composition $F \circ H$ is again a semialgebraic function. Finally, let us prove the last statement and let us consider the graph of the marginal function φ:

$$\begin{aligned} \text{gra } \varphi &= \{(x, \tau) \in \mathbb{R}^n \times \mathbb{R} : \tau = \varphi(x) = \inf_y\ f(x, y)\} \\ &= \{(x, \tau) \in \mathbb{R}^n \times \mathbb{R} : \forall\ \varepsilon > 0,\ \exists\ y \in \mathbb{R}^m, \text{ such that } f(x, y) \le \tau + \varepsilon\}. \end{aligned}$$

Now, let us define $b : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$, $b(x, \tau, \varepsilon, y) := f(x, y) - \tau - \varepsilon$, and

$$\mathcal{B} := \{(x, \tau, \varepsilon, y) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}^m : b(x, \tau, \varepsilon, y) \le 0\}.$$

Then apparently, the set $\mathcal{B}$ can be written as

$$\mathcal{B} = \pi_3(\text{gra } b \cap \mathbb{R}^{m+n+2} \times \mathbb{R}_-),$$

where $\pi_3(z_1, ..., z_{m+n+2}, z_{m+n+3}) = (z_1, ..., z_{m+n+2})$ is the canonical projection onto the first $n + m + 2$ components of $z \in \mathbb{R}^{m+n+3}$. Since f is assumed to be a semialgebraic function, it immediately follows that the graph of b and the set $\mathcal{B}$ are semialgebraic sets. Moreover, by setting

$$\pi_1(z_1, ..., z_{n+1}, z_{n+2}) = (z_1, ..., z_{n+1}), \quad \pi_2(z_1, ..., z_{n+2}, z_{n+3}, ..., z_{m+n+2}) = (z_1, ..., z_{n+2}),$$

the graph gra φ can be represented as follows

$$\text{gra } \varphi = \mathbb{R}^n \times \mathbb{R} \setminus \pi_1 \left(\{(x,\tau,\varepsilon) \in \mathbb{R}^n \times \mathbb{R} \times \mathbb{R} : \varepsilon > 0\} \setminus \pi_2(\mathcal{B}) \right).$$

Let us note that this construction basically expresses the quantifiers "$\forall$" and "$\exists$" as set operations. Since the involved sets are all semialgebraic, this clearly proves our claim. □

The next result establishes semismoothness of semialgebraic and tame functions and was first presented by Bolte, Daniilidis, and Lewis in [21, Theorem 1 and Remark 4]; see also [114] for an alternative proof.

Theorem 3.3.3. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous mapping. Then, the following statements hold:*

(i) *If the function F is tame, then F is semismooth.*

(ii) *If F is semialgebraic, then F is α-order semismooth for some $\alpha > 0$.*

As in the work of Ding [63], we can now derive semismoothness of the proximity operator $\text{prox}_\varphi^\Lambda$ of a semialgebraic function.

Corollary 3.3.4. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, lower semicontinuous, and semialgebraic function and let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary parameter matrix. Then, the proximity operator $\text{prox}_\varphi^\Lambda : \mathbb{R}^n \to \mathbb{R}^n$ is a semialgebraic mapping and α-order semismooth for some $\alpha > 0$.*

Proof. Due to Theorem 3.3.3, it suffices to show that the proximity operator is a semialgebraic function. Since the Moreau envelope $\text{env}_\varphi^\Lambda$ is the marginal function of the semialgebraic mapping

$$\theta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad \theta(x,y) := \varphi(y) + \frac{1}{2}\|x-y\|_\Lambda^2,$$

it immediately follows that $\text{env}_\varphi^\Lambda$ is semialgebraic. Furthermore, the graph of the proximity operator $\text{prox}_\varphi^\Lambda$ can be represented as follows

$$\text{gra prox}_\varphi^\Lambda = \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^n : \theta(x,y) = \text{env}_\varphi^\Lambda(x)\}.$$

Since the functions θ and $\text{env}_\varphi^\Lambda$ are semialgebraic, this clearly implies that the proximity operator $\text{prox}_\varphi^\Lambda$ is a semialgebraic mapping. □

Using the calculus of semialgebraic mappings and the *Courant-Fischer max-min principle*, it is possible to show that the absolute value of a real number and the k-th eigenvalue or singular value of a matrix are semialgebraic functions. This implies that the ℓ_p-norms and the Schatten-p norms are semialgebraic for all $p \in [1,\infty) \cap \mathbb{Q}$ and $p = \infty$. Specifically, the ℓ_1-norm, the TV-regularization and the nuclear norm are examples of semialgebraic functions. We refer to Karow [122, Section 3.1] for detailed proofs.

Thus, in summary, we have seen that the class of semialgebraic functions is rather broad and enjoys a rich calculus. Moreover, nearly every application and example considered in this thesis can be treated within the framework of semialgebraic mappings. Finally, let us also note that the abstract and general results of this subsection can be further extended to the larger o-minimal structure of the so-called *globally subanalytic* functions, see [20] and the references therein.

Second order properties of the proximity operator

In the following paragraph, motivated by the results in [110, 145], we want to discuss and derive certain second order properties of the proximity operator which will be obligatory for the second order analysis of optimization problems later on.

As usual, let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. In Lemma 3.1.1 and the subsequent discussion, it was shown that the proximity operator $\text{prox}^\Lambda_\varphi$ is a Lipschitz continuous function. Thus, due to the Theorem of Rademacher, the function $\text{prox}^\Lambda_\varphi$ is Fréchet differentiable almost everywhere. Let $\Omega^\Lambda_\varphi \subset \mathbb{R}^n$ denote the set of all points at which the proximity operator $\text{prox}^\Lambda_\varphi$ is Fréchet differentiable. Then, the following statements are true:

- The function $\text{env}^\Lambda_\varphi$ is twice Fréchet differentiable on Ω^Λ_φ.
- For all $x \in \Omega^\Lambda_\varphi$ the matrix $\Lambda D\text{prox}^\Lambda_\varphi(x)$ is symmetric and positive semidefinite.
- For all $x \in \Omega^\Lambda_\varphi$ the matrix $\Lambda(I - D\text{prox}^\Lambda_\varphi(x))$ is symmetric and positive semidefinite.

Let us briefly verify the latter properties. The first result follows immediately from Lemma 3.1.5. By setting

$$T(x) := \frac{1}{2}\|x\|^2_\Lambda - \text{env}^\Lambda_\varphi(x),$$

the symmetry of the matrix $\Lambda D\text{prox}^\Lambda_\varphi(x)$ follows from the identity $\nabla^2 T(x) = \Lambda D\text{prox}^\Lambda_\varphi(x)$, for $x \in \Omega^\Lambda_\varphi$, and the well-known fact that a twice Fréchet differentiable function possesses a symmetric Hessian; see, e.g., [62, Theorem 8.12.2 and 8.12.3]. Now, let $x \in \Omega^\Lambda_\varphi$, $h \in \mathbb{R}^n$, be arbitrary and let $t > 0$ be sufficiently small. Then, due to Lemma 3.1.1, we have

$$\begin{aligned} 0 \le \|\text{prox}^\Lambda_\varphi(x + th) - \text{prox}^\Lambda_\varphi(x)\|^2_\Lambda &\le \langle \text{prox}^\Lambda_\varphi(x + th) - \text{prox}^\Lambda_\varphi(x), \Lambda th\rangle \\ &= \langle D\text{prox}^\Lambda_\varphi(x) \cdot th, \Lambda th\rangle + o(t^2), \end{aligned}$$

where we used the Fréchet differentiability of $\text{prox}^\Lambda_\varphi$ on Ω^Λ_φ. Dividing both sides of the latter inequality by t^2 and taking the limit $t \downarrow 0$, we establish

$$\langle h, \Lambda D\text{prox}^\Lambda_\varphi(x)h\rangle \ge 0, \quad \forall\ h \in \mathbb{R}^n.$$

To prove the third assertion, we first use

$$\nabla^2 \text{env}^\Lambda_\varphi(x) = \Lambda(I - D\text{prox}^\Lambda_\varphi(x)), \quad \forall\ x \in \Omega^\Lambda_\varphi.$$

Thus, as in the second part, the matrix $\Lambda(I - D\text{prox}^\Lambda_\varphi(x))$ has to be symmetric. Moreover, the positive semidefiniteness of $\Lambda(I - D\text{prox}^\Lambda_\varphi(x))$ is a direct consequence of the convexity of the Moreau envelope $\text{env}^\Lambda_\varphi$.

Let us recall that the Bouligand subdifferential of the proximity operator $\operatorname{prox}_{\varphi}^{\Lambda}$ at $x \in \mathbb{R}^n$ is defined as follows

$$\partial_B \operatorname{prox}_{\varphi}^{\Lambda}(x) = \{M \in \mathbb{R}^{n\times n} : \exists\ (x^k)_k \subset \Omega_{\varphi}^{\Lambda} \text{ such that } x^k \to x,\, D\operatorname{prox}_{\varphi}^{\Lambda}(x^k) \to M\}.$$

Now, a simple continuity argument and Lemma 2.5.20 show that the last two properties do also hold for every generalized derivative $M \in \partial_B \operatorname{prox}_{\varphi}^{\Lambda}(x)$ of the Bouligand subdifferential of $\operatorname{prox}_{\varphi}^{\Lambda}$. In the following Lemma, we summarize our observations and present an analogue and final result for the Clarke subdifferential of the proximity operator $\operatorname{prox}_{\varphi}^{\Lambda}$. Let us mention that Meng, Sun, and Zhao [145] have already established a similar result for metric projections onto convex, nonempty, and closed sets.

Lemma 3.3.5. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function and let $\Lambda \in \mathbb{S}^n_{++}$ and $x \in \mathbb{R}^n$ be arbitrary. Then, for every $V \in \partial \operatorname{prox}_{\varphi}^{\Lambda}(x) \subset \mathbb{R}^{n\times n}$, the following statements are true:*

(i) *The matrices ΛV and $\Lambda(I - V)$ are symmetric and positive semidefinite.*

(ii) *It holds $\langle Vh, \Lambda(I-V)h\rangle \geq 0$ for all $h \in \mathbb{R}^n$.*

Proof. The first part is an immediate consequence of $\partial \operatorname{prox}_{\varphi}^{\Lambda}(x) = \operatorname{conv} \partial_B \operatorname{prox}_{\varphi}^{\Lambda}(x)$ and Lemma 2.5.20. The proof of the second part is identical to the proof of [145, Proposition 1] and will be omitted here. □

Let us note, that the symmetry of the matrix ΛV, $V \in \partial \operatorname{prox}_{\varphi}^{\Lambda}(x)$, can be used to calculate the transposed of V. In particular, it holds

$$V^\top = V^\top \Lambda \Lambda^{-1} = (\Lambda V)^\top \Lambda^{-1} = \Lambda V \Lambda^{-1}, \quad \text{and} \quad V = \Lambda^{-1} V^\top \Lambda. \tag{3.3.3}$$

We conclude this section with a structural property of the directional derivative of the proximity operator. This result is a straightforward extension of [132, Corollary 2.6].

Lemma 3.3.6. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function and let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. Suppose that $\operatorname{prox}_{\varphi}^{\Lambda}$ is directionally differentiable at some point $x \in \mathbb{R}^n$, then it holds*

$$(\operatorname{prox}_{\varphi}^{\Lambda})'(x;h) \in N_{\partial\varphi(\operatorname{prox}_{\varphi}^{\Lambda}(x))}(\nabla \operatorname{env}_{\varphi}^{\Lambda}(x)), \quad \forall\ h \in \mathbb{R}^n.$$

Proof. For the sake of completeness, let us recapitulate the proof presented in [132]. Let $h \in \mathbb{R}^n$ be arbitrary and $t > 0$. Due to the monotonicity of the convex subdifferential $\partial\varphi$ and $\nabla \operatorname{env}_{\varphi}^{\Lambda}(x+th) \in \partial\varphi(\operatorname{prox}_{\varphi}^{\Lambda}(x+th))$, it holds

$$\langle \nabla \operatorname{env}_{\varphi}^{\Lambda}(x+th) - \lambda, \operatorname{prox}_{\varphi}^{\Lambda}(x+th) - \operatorname{prox}_{\varphi}^{\Lambda}(x)\rangle \geq 0, \quad \forall\ \lambda \in \partial\varphi(\operatorname{prox}_{\varphi}^{\Lambda}(x)).$$

Using the continuity of $\nabla \operatorname{env}_{\varphi}^{\Lambda}$ and the directional differentiability of $\operatorname{prox}_{\varphi}^{\Lambda}$, it follows

$$\begin{aligned}&\langle \nabla \operatorname{env}_{\varphi}^{\Lambda}(x) - \lambda, (\operatorname{prox}_{\varphi}^{\Lambda})'(x;h)\rangle \\ &\qquad = \lim_{t\downarrow 0} \left\langle \nabla \operatorname{env}_{\varphi}^{\Lambda}(x+th) - \lambda, \frac{\operatorname{prox}_{\varphi}^{\Lambda}(x+th) - \operatorname{prox}_{\varphi}^{\Lambda}(x)}{t} \right\rangle \geq 0\end{aligned}$$

for all $\lambda \in \partial\varphi(\mathrm{prox}_\varphi^\Lambda(x))$ and hence, $(\mathrm{prox}_\varphi^\Lambda)'(x; h) \in N_{\partial\varphi(\mathrm{prox}_\varphi^\Lambda(x))}(\nabla \mathrm{env}_\varphi^\Lambda(x))$. □

4. A globalized semismooth Newton method for nonsmooth optimization problems

In this chapter, we propose and investigate a semismooth Newton method for general nonsmooth optimization problems of the form

$$(\mathcal{P}) \qquad \min_{x \in \mathbb{R}^n} \; f(x) + \varphi(x) =: \psi(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, possibly nonconvex and $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous mapping.

The proposed algorithm generalizes the semismooth Newton method for ℓ_1-regularized optimization problems that was presented and analyzed by Milzarek and Ulbrich in [157]. In particular, by exploiting the properties of the proximity operator, we will modify and extend the algorithmic framework and the convergence theory given in [157] to the more general class of nonsmooth minimization problems $(\mathcal{P})$.

As in [157], we will combine the efficiency of filter globalization techniques with the fast local convergence properties of the semismooth Newton method [197, 199] to construct an overall globally and locally fast converging algorithm. Our approach is primarily based on the idea to obtain trial steps from semismooth Newton steps for a nonsmooth reformulation

$$(\mathcal{E}) \qquad F^\Lambda(x) = x - \mathrm{prox}_\varphi^\Lambda(x - \Lambda^{-1} \nabla f(x)) = 0, \quad \Lambda \in \mathbb{S}^n_{++}$$

of the first order optimality conditions of $(\mathcal{P})$. The acceptance of these steps is controlled by a multidimensional filter globalization technique. If the semismooth Newton step is not accepted, then a suitably chosen descent step is performed. The main requirement is that these alternative steps ensure global convergence in the case where only finitely many semismooth Newton steps are taken. Here, we choose a proximal gradient method with an Armijo-type line search, which was first introduced by Fukushima and Mine [88], for this purpose. The nonsmooth function $F^\Lambda : \mathbb{R}^n \to \mathbb{R}^n$ arising in $(\mathcal{E})$ will be derived in section 4.1.

We use a globalization technique that is based on a multidimensional filter framework. Originally, the filter concept was developed by Fletcher and Leyffer [83] in order to globalize SQP methods for nonlinear programming problems without using penalty functions. The original version of the filter method works with a two dimensional filter, where each entry consists of the objective function value and a measure for the constraint violation at a given point. The filter globalization concept has rapidly established itself as one of the most important and efficient globalization techniques in nonlinear programming. For further details we refer to [84, 82, 240]. Gould, Leyffer, and Toint modified this concept in [95] and proposed a multidimensional filter to globalize (Gauss-)Newton-based methods for nonlinear equations and least squares problems. In [96] Gould, Sainvitu, and Toint adapted this approach to an unconstrained minimization problem by applying the method to the gradient of the objective function. Our method can be viewed as an extension of this idea to the general setting of the nonsmooth optimization problem ($\mathcal{P}$).

Under assumptions comparable to those of other state-of-the-art methods, we prove for our algorithm that every accumulation point of the generated sequence is a stationary point. Furthermore, under suitable second order conditions, transition to q-superlinear local convergence is shown. In contrast to many other analyses, we consider not only the case of convex f, but also address the general situation of a nonconvex function f. Moreover, in the subsequent chapter, we also provide a profound and detailed discussion of abstract and different second order-type conditions for problem ($\mathcal{P}$). In particular, for a certain class of nonsmooth functions φ, we will show that the semismoothness of the proximity operator $\text{prox}_{\varphi}^{\Lambda}$, a (no gap) second order sufficient condition and the strict complementarity condition guarantees fast local convergence of the semismooth Newton method.

Let us note that the following sections are essentially based on the work [157] and that several parts have already appeared in similar form in [157] for the ℓ_1-regularized setting.

This chapter is organized as follows. In section 4.1 we specify different optimality conditions for the nonsmooth minimization problem ($\mathcal{P}$) and derive the nonsmooth equation ($\mathcal{E}$). In the sections 4.2.1–4.2.3 we state the assumptions under which we prove convergence and discuss some preliminaries concerning the properties of the proximal gradient method as well as the theoretic introduction and examination of the multidimensional filter method. We then continue with the presentation of the main approach. In section 4.3 we prove our results on global and local convergence of the algorithm. Finally, the mentioned second order framework for the nonsmooth problem ($\mathcal{P}$) can be found in chapter 5.

4.1. First order optimality conditions

We now derive first order optimality conditions for the nonsmooth optimization problem ($\mathcal{P}$). Therefore, suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous mapping and let us assume that $\bar{x} \in \text{dom}\ \varphi$ is a local solution of problem ($\mathcal{P}$). Then, for all $d \in \mathbb{R}^n$ and all $t > 0$ sufficiently small it holds

$$\psi(\bar{x} + td) - \psi(\bar{x}) \geq 0.$$

This shows that the directional epiderivative $\psi_{-}^{\downarrow}(\bar{x}; d)$ must be nonnegative for all $d \in \mathbb{R}^n$. Furthermore, by using the convexity of φ, it follows

$$\begin{aligned}
\psi_{-}^{\downarrow}(\bar{x}; d) &\leq \liminf_{t \downarrow 0} \frac{f(\bar{x} + td) - f(\bar{x}) + \varphi((1-t)\bar{x} + t(\bar{x} + d)) - \varphi(\bar{x})}{t} \\
&\leq \varphi(\bar{x} + d) - \varphi(\bar{x}) + \liminf_{t \downarrow 0} \frac{f(\bar{x} + td) - f(\bar{x})}{t} \\
&= \varphi(\bar{x} + d) - \varphi(\bar{x}) + \nabla f(\bar{x})^{\top} d,
\end{aligned}$$

which directly implies $-\nabla f(\bar{x}) \in \partial\varphi(\bar{x})$. Next, let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary symmetric and positive definite matrix. Then, the latter condition is clearly equivalent to

$$\bar{x} \in \bar{x} - \Lambda^{-1}\nabla f(\bar{x}) - \Lambda^{-1} \cdot \partial\varphi(\bar{x})$$

and by invoking equation (3.1.2), this just means

$$\bar{x} = \mathrm{prox}^{\Lambda}_{\varphi}(\bar{x} - \Lambda^{-1}\nabla f(\bar{x})). \tag{4.1.1}$$

Finally, let us assume that $\bar{x}$ satisfies the fixed point-type equation (4.1.1) and let us set $\bar{z} := \bar{x} - \Lambda^{-1}\nabla f(\bar{x})$. Then, rearranging the terms in (4.1.1) and in the definition of $\bar{z}$, we obtain

$$\bar{x} = \mathrm{prox}^{\Lambda}_{\varphi}(\bar{z}) \quad \text{and} \quad \nabla f(\mathrm{prox}^{\Lambda}_{\varphi}(\bar{z})) + \Lambda(\bar{z} - \mathrm{prox}^{\Lambda}_{\varphi}(\bar{z})) = 0.$$

Of course, in the general, nonconvex case, we cannot expect that these different conditions are sufficient to guarantee optimality. However, they can be used to characterize stationarity of a feasible point.

Definition 4.1.1 (Stationarity). *A feasible point $\bar{x} \in \mathrm{dom}\ \varphi$ is called* stationary point *of the problem $(\mathcal{P})$ if it holds*

$$\psi^{\downarrow}_{-}(\bar{x}; d) \geq 0, \quad \forall\ d \in \mathbb{R}^n. \tag{4.1.2}$$

In the following, we collect our different reformulations of the stationarity condition (4.1.2) and summarize our previous discussion.

Lemma 4.1.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. Furthermore, assume that $\bar{x} \in \mathrm{dom}\ \varphi$ is a stationary point of problem $(\mathcal{P})$. Then, the following conditions are mutually equivalent:*

(i) *For all $d \in \mathbb{R}^n$ it holds $\psi^{\downarrow}_{-}(\bar{x}; d) \geq 0$.*

(ii) *It holds $0 \in \nabla f(\bar{x}) + \partial\varphi(\bar{x})$.*

(iii) *For any $\Lambda \in \mathbb{S}^n_{++}$, the following fixed point-type equation is satisfied*

$$F^{\Lambda}(\bar{x}) := \bar{x} - \mathrm{prox}^{\Lambda}_{\varphi}(\bar{x} - \Lambda^{-1}\nabla f(\bar{x})) = 0.$$

(iv) *Let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. Then, the vector $\bar{z} = \bar{x} - \Lambda^{-1}\nabla f(\bar{x})$ is a zero of the so-called* normal map *$F^{\Lambda}_{\mathrm{nor}} : \mathbb{R}^n \to \mathbb{R}^n$,*

$$F^{\Lambda}_{\mathrm{nor}}(\bar{z}) := \nabla f(\mathrm{prox}^{\Lambda}_{\varphi}(\bar{z})) + \Lambda(\bar{z} - \mathrm{prox}^{\Lambda}_{\varphi}(\bar{z})) = 0.$$

Proof. So far, we have already shown (i) $\Rightarrow$ (ii) $\Leftrightarrow$ (iii) $\Rightarrow$ (v). To complete the proof, let us suppose that $\bar{z}$ is a zero of the Normal map $F^{\Lambda}_{\mathrm{nor}}$. By setting $\bar{x} := \mathrm{prox}^{\Lambda}_{\varphi}(\bar{z})$, we readily obtain $\bar{z} = \bar{x} - \Lambda^{-1}\nabla f(\bar{x})$ and $F^{\Lambda}(\bar{x}) = \bar{x} - \mathrm{prox}^{\Lambda}_{\varphi}(\bar{z}) = 0$. Since the conditions (ii) and (iii) are already known to be equivalent, we also see that this result does not depend on the specific choice of Λ. Now, let us turn to the direction "(ii) $\Rightarrow$ (i)". The inclusion $-\nabla f(\bar{x}) \in \partial\varphi(\bar{x})$ implies that φ is subdifferentiable at $\bar{x}$. In this case, $\Pi(\cdot) := \varphi^{\downarrow}(\bar{x}; \cdot)$ is a convex, proper, lower semicontinuous, and positively homogeneous mapping and by (2.5.5), we obtain $\partial\Pi(0) = \partial\varphi(x)$. Thus, applying Remark 2.5.8, we can infer

$$0 \leq \nabla f(\bar{x})^\top (d - 0) + \Pi(d) - \Pi(0) = \nabla f(\bar{x})^\top d + \varphi^{\downarrow}(\bar{x}; d) \leq \psi^{\downarrow}_{-}(\bar{x}; d), \quad \forall\, d \in \mathbb{R}^n.$$

This concludes the proof. □

As already mentioned, our strategy and overall goal is to develop a globally and locally fast converging, semismooth Newton method to solve the nonsmooth equation

$$F^{\Lambda}(x) = x - \operatorname{prox}^{\Lambda}_{\varphi}(x - \Lambda^{-1}\nabla f(x)) = 0. \tag{4.1.3}$$

Clearly, Lemma 4.1.2 justifies this approach since each solution of (4.1.3) corresponds to a stationary point of our initial problem ($\mathcal{P}$).

In the remainder of this section, we will discuss several useful properties of the nonsmooth function F^{Λ}. The next statement shows that $\|F^{\Lambda}(x)\|$ does not grow too much when the parameter matrix Λ changes. This result was first established by Tseng and Yun in [236].

Lemma 4.1.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. Moreover, let $\Lambda_1, \Lambda_2 \in \mathbb{S}^n_{++}$ be two arbitrary symmetric, positive definite matrices. Then, for all $x \in \mathbb{R}^n$ and for $W := \Lambda_2^{-\frac{1}{2}}\Lambda_1\Lambda_2^{-\frac{1}{2}}$, it follows*

$$\|F^{\Lambda_1}(x)\| \leq \frac{1 + \lambda_{\max}(W) + \sqrt{1 - 2\lambda_{\min}(W) + \lambda_{\max}(W)^2}}{2} \frac{\lambda_{\max}(\Lambda_2)}{\lambda_{\min}(\Lambda_1)} \|F^{\Lambda_2}(x)\|.$$

Proof. We refer to [236, Lemma 3] for a detailed proof. Let us briefly remark that in [236, Lemma 3] the additional restriction "$x \in \operatorname{dom}\varphi \equiv \operatorname{dom} P$" is made. This assumption is not really necessary since, in the proof the "crucial", possibly extended valued function "$\varphi \equiv P$" is only evaluated at appropriate proximity operators that are always contained in $\operatorname{dom}\varphi \equiv \operatorname{dom} P$. □

Remark 4.1.4. Let $\Lambda \in \mathbb{S}^n_{++}$ be given and let $(\Lambda_k)_k \subset \mathbb{S}^n_{++}$ be a family of symmetric, positive definite matrices. Suppose that there exist constants $\lambda_M \geq \lambda_m > 0$ such that

$$\lambda_M I \succeq \Lambda_k \succeq \lambda_m I, \quad \forall\, k \in \mathbb{N}.$$

Then, it easily follows

$$\frac{\lambda_{\max}(\Lambda)}{\lambda_m} I \succeq \Lambda_k^{-\frac{1}{2}}\Lambda\Lambda_k^{-\frac{1}{2}} \succeq \frac{\lambda_{\min}(\Lambda)}{\lambda_M} I \quad \text{and} \quad \frac{\lambda_M}{\lambda_{\min}(\Lambda)} I \succeq \Lambda^{-\frac{1}{2}}\Lambda_k\Lambda^{-\frac{1}{2}} \succeq \frac{\lambda_m}{\lambda_{\max}(\Lambda)} I,$$

for all $k \in \mathbb{N}$, and, due to Lemma 4.1.3, we obtain the following bounds

$$\underline{\lambda} \cdot \|F^{\Lambda}(x)\| \leq \|F^{\Lambda_k}(x)\| \leq \overline{\lambda} \cdot \|F^{\Lambda}(x)\| \tag{4.1.4}$$

for all $k \in \mathbb{N}$, $x \in \mathbb{R}^n$ and some constants $\underline{\lambda}$, $\overline{\lambda} > 0$, which do not depend on k or Λ_k. Thus, if the parameter matrices Λ_k remain in a bounded set, the latter inequalities imply:

$$F^\Lambda(x^k) \to 0 \qquad \iff \qquad F^{\Lambda_k}(x^k) \to 0, \quad k \to \infty.$$

As a consequence, the parameter matrix Λ is allowed to change in each iteration. Hence, adaptive update schemes such as the well-known *Barzilai-Borwein step size rule*, [10], or other techniques can be applied.

Lemma 4.1.5. *Suppose that f and φ satisfy the assumptions in Lemma 4.1.3 and let $\Lambda_1, \Lambda_2 \in \mathbb{S}^n_{++}$ be arbitrary. Then, for all $x \in \mathbb{R}^n$, it holds*

$$\|F^{\Lambda_1}(x) - F^{\Lambda_2}(x)\| \leq \frac{1}{\lambda_{\min}(\Lambda_1)} \|(\Lambda_2 - \Lambda_1)F^{\Lambda_2}(x)\|. \tag{4.1.5}$$

Proof. The proof uses the same techniques and ideas as the proof of [236, Lemma 3]; see also [210, Lemma 3] and [129, Proposition 3.6] for related results. Using $x - F^{\Lambda_i}(x) = \mathrm{prox}_\varphi^{\Lambda_i}(x - \Lambda_i^{-1}\nabla f(x)) \in \mathrm{dom}\ \varphi$, for $i = 1, 2$, and the characterization of the proximity operator (3.1.2), we obtain

$$\begin{cases} \varphi(x - F^{\Lambda_2}(x)) - \varphi(x - F^{\Lambda_1}(x)) \geq \langle \Lambda_1 F^{\Lambda_1}(x) - \nabla f(x), F^{\Lambda_1}(x) - F^{\Lambda_2}(x)\rangle, \\ \varphi(x - F^{\Lambda_1}(x)) - \varphi(x - F^{\Lambda_2}(x)) \geq \langle \Lambda_2 F^{\Lambda_2}(x) - \nabla f(x), F^{\Lambda_2}(x) - F^{\Lambda_1}(x)\rangle. \end{cases}$$

Adding those two inequalities yields

$$\langle \Lambda_1 F^{\Lambda_1}(x) - \Lambda_2 F^{\Lambda_2}(x), F^{\Lambda_1}(x) - F^{\Lambda_2}(x)\rangle \leq 0$$

and

$$\begin{aligned} \|F^{\Lambda_1}(x) - F^{\Lambda_2}(x)\|^2_{\Lambda_1} &\leq \langle (\Lambda_2 - \Lambda_1)F^{\Lambda_2}(x), F^{\Lambda_1}(x) - F^{\Lambda_2}(x)\rangle \\ &\leq \|(\Lambda_2 - \Lambda_1)F^{\Lambda_2}(x)\| \|F^{\Lambda_1}(x) - F^{\Lambda_2}(x)\|. \end{aligned}$$

Finally, by applying (3.1.3), we establish inequality (4.1.5). □

The next lemma provides a connection between the nonsmooth function $F^\Lambda(x)$ and the normal map $F^\Lambda_{\mathrm{nor}}(x)$ and generalizes a result of Facchinei and Pang for variational inequalities; see [75, Proposition 1.5.14].

Lemma 4.1.6. *Let f and φ satisfy the assumptions in Lemma 4.1.3 and let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary symmetric, positive definite matrix. Moreover, let $x \in \mathrm{dom}\ \varphi$ be given and suppose that φ is subdifferentiable at x. Then, it holds*

$$\|F^\Lambda(x)\|_\Lambda \leq \mathrm{dist}(-\nabla f(x), \partial\varphi(x))_{\Lambda^{-1}} = \inf_z \{\|F^\Lambda_{\mathrm{nor}}(z)\|_{\Lambda^{-1}} : x = \mathrm{prox}^\Lambda_\varphi(z)\}.$$

Proof. Let $x - y \in \mathrm{dom}\ \varphi$ be arbitrary. As in the proof of Lemma 4.1.5 we obtain

$$\begin{aligned} \varphi(x-y) - \varphi(x - F^\Lambda(x)) &\geq \langle \Lambda F^\Lambda(x) - \nabla f(x), F^\Lambda(x) - y\rangle \\ &= \|F^\Lambda(x)\|^2_\Lambda - \langle F^\Lambda(x) - \Lambda^{-1}\nabla f(x), y\rangle_\Lambda - \langle \nabla f(x), F^\Lambda(x)\rangle. \end{aligned}$$

Now, setting $y = 0$ it follows

$$\begin{aligned}\|F^\Lambda(x)\|_\Lambda^2 &\leq \langle \nabla f(x), F^\Lambda(x)\rangle + \varphi(x) - \varphi(\mathrm{prox}_\varphi^\Lambda(x - \Lambda^{-1}\nabla f(x))) \\ &\leq \langle \nabla f(x) + v, F^\Lambda(x)\rangle \leq \|\nabla f(x) + v\|_{\Lambda^{-1}}\|F^\Lambda(x)\|_\Lambda\end{aligned}$$

for all $v \in \partial\varphi(x)$. Thus, by taking the infimum over all such $v \in \partial\varphi(x)$ and by using

$$x = \mathrm{prox}_\varphi^\Lambda(z) \quad\Longleftrightarrow\quad x \in z - \Lambda^{-1}\partial\varphi(x) \quad\Longleftrightarrow\quad \Lambda(z - x) \in \partial\varphi(x),$$

we establish the following estimate

$$\begin{aligned}\|F^\Lambda(x)\|_\Lambda &\leq \inf_{v \in \partial\varphi(x)} \|\nabla f(x) + v\|_{\Lambda^{-1}} \\ &= \mathrm{dist}(-\nabla f(x), \partial\varphi(x))_{\Lambda^{-1}} = \inf_z \{\|F_{\mathrm{nor}}^\Lambda(z)\|_{\Lambda^{-1}} : x = \mathrm{prox}_\varphi^\Lambda(z)\},\end{aligned}$$

as desired. □

4.2. Algorithmic framework

In this section, we present the different algorithmic components of our globalized semismooth Newton method in detail.

In particular, in subsection 4.2.2, we propose and investigate a proximal gradient method that will be used as an underlying base algorithm and that was first analyzed by Fukushima, Mine [88] and Tseng, Yun [236]. As in [236], we incorporate an Armijo-type linesearch to guarantee descent and global convergence of the iterates generated by the proximal gradient method. Afterwards, in section 4.2.3, we introduce a multidimensional filter framework that controls the acceptance of the Newton iterates and suitably connects the proximal gradient method and the semismooth Newton approach. In contrast to most other convergence analyses for convex composite algorithms, we will consider both the general nonconvex case and the convex case. Taking account of the possible effects of the nonconvexity, we augment our basic combination of semismooth Newton and proximal gradient steps by adding certain growth conditions. In subsection 4.2.4, we present these growth conditions and the full algorithm in detail. Moreover, we also give examples that illustrate the application of the semismooth Newton method in practise.

Besides, we also propose a specialized method for convex problems that has an easier structure and converges under similar assumptions. More specifically, if f is convex and if the nonsmooth function φ is positively homogeneous and real valued, then an abstract convergence result can be established without specifying any algorithmic details. This extends a similar result of Milzarek and Ulbrich [157] for convex ℓ_1-regularized problems.

We start with a brief discussion of our basic assumptions.

4.2.1. Assumptions

The following conditions summarize our assumptions for proving that every accumulation point of the proposed algorithm is a stationary point.

Assumption 4.2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be given and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function. Furthermore, let us assume that there exists an open, convex set* $\operatorname{dom} \varphi \subseteq \Omega \subseteq \mathbb{R}^n$ *such that:*

(A.1) *The function f is continuously differentiable on Ω.*

(A.2) *The gradient mapping $\nabla f : \Omega \to \mathbb{R}^n$ is Lipschitz continuous on* $\operatorname{dom} \varphi$ *with modulus $L_f > 0$.*

(A.3) *The mapping f is twice continuously differentiable on Ω.*

We will also utilize the following condition. Let $(\Lambda_k)_k \subset \mathbb{S}^n_{++}$ be a family of symmetric, positive definite parameter matrices, then we assume:

(B) *There exist $0 < \lambda_m \leq \lambda_M$ such that $\lambda_M I \succeq \Lambda_k \succeq \lambda_m I$ for all $k \in \mathbb{N}$.*

Next, we consider a specialized version of Assumption 4.2.1:

Assumption 4.2.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be given and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. Furthermore, let us assume that there exists an open, convex set* $\operatorname{dom} \varphi \subseteq \Omega \subseteq \mathbb{R}^n$ *such that:*

(C.1) *The function $f : \Omega \to \mathbb{R}$ is convex.*

(C.2) *The objective function ψ is coercive on Ω.*

(C.3) *The function $\varphi : \mathbb{R}^n \to \mathbb{R}$ is real valued and positively homogeneous.*

In the following, we want show that for a general class of algorithms comprising those investigated in this thesis, the assumptions (A.1), (B), and (C.1)–(C.3) imply boundedness of an arbitrary sequence $(x^k)_k$ of iterates satisfying $F^{\Lambda_k}(x^k) \to 0$, as $k \to \infty$. We start with a discussion of the compactness of the level sets of ψ.

Lemma 4.2.3. *Let $\Omega \subset \mathbb{R}^n$ and $f : \Omega \to \mathbb{R}$ satisfy condition* (C.1)*, then the following statements are equivalent:*

(i) *For every $\alpha \in \mathbb{R}$ the level set* $\operatorname{lev}_\alpha \psi$ *is compact.*

(ii) *There exists $\bar{\alpha} \in \mathbb{R}$ such that the level set* $\operatorname{lev}_{\bar{\alpha}} \psi$ *is nonempty and compact.*

(iii) *For some $\bar{\alpha} \in \mathbb{R}$, the level set* $\operatorname{lev}_{\bar{\alpha}} \psi$ *is nonempty and for every $\bar{x} \in \operatorname{lev}_{\bar{\alpha}} \psi$ there exists $\vartheta, R > 0$ such that*

$$\psi(x) \geq \psi(\bar{x}) + \vartheta \|x - \bar{x}\|, \quad \forall\, x \in \Omega \setminus B_R(0).$$

(iv) *The function ψ is coercive on Ω, i.e.,* $\lim_{x \in \Omega,\, \|x\| \to \infty} \psi(x) = +\infty$.

Proof. At first, let us suppose, that the level sets $\mathrm{lev}_\alpha\ \psi$, $\alpha \in \mathbb{R}$, are all compact. Since the function φ is proper, there exists $\xi \in \mathrm{dom}\ \varphi$ and, consequently, the level set $\mathrm{lev}_{\psi(\xi)}\ \psi$ is nonempty and compact. Now, let us assume that $\mathrm{lev}_{\bar\alpha}\ \psi$ is nonempty and compact for a certain $\bar\alpha \in \mathbb{R}$. Due to the compactness of $\mathrm{lev}_{\bar\alpha}\ \psi$ and the Theorem of Weierstrass, we can expand the level set $\mathrm{lev}_{\bar\alpha}\ \psi$ by a small strip without leaving Ω, i.e., there exists $\delta > 0$ such that

$$\mathcal{L}_\delta := \{x + d\ :\ x \in \mathrm{lev}_{\bar\alpha}\ \psi,\ d \in \mathbb{R}^n,\ \|d\| < \delta\}$$

is still contained in Ω (but not necessarily in $\mathrm{dom}\ \varphi$). Let $\Gamma := \partial\mathcal{L}_\delta$ denote the boundary of $\mathcal{L}_\delta$ and let us fix an arbitrary point $\bar x \in \mathrm{lev}_{\bar\alpha}\ \psi$. Using the compactness of Γ and Weierstrass's Theorem, we can find $\varepsilon > 0$ such that

$$\psi(x) \geq \inf_{y\in\Gamma}\ \psi(y) > \psi(\bar x) + \varepsilon, \quad \forall\ x \in \Gamma.$$

(In the case $\Gamma \cap \mathrm{dom}\ \varphi = \emptyset$, this inequality is obviously fulfilled). Further, there exists $r > 0$ with $\max_{x\in\Gamma} \|x - \bar x\| = r$. Next, let $t \in (0, 1)$ and $z \in \Gamma$ be arbitrary and let us consider the point $x = \bar x + \frac{1}{t}(z - \bar x)$. Then, the convexity of ψ implies

$$\psi(z) = \psi(tx + (1-t)\bar x) \leq t\psi(x) + (1-t)\psi(\bar x).$$

By combining the latter properties with the estimate $\frac{1}{t} = \frac{\|x-\bar x\|}{\|z-\bar x\|} \geq \frac{1}{r}\|x - \bar x\|$, we can establish the following growth rate:

$$\text{(4.2.1)} \qquad \psi(x) > \frac{1}{t}(\psi(\bar x) + \varepsilon) + \left(1 - \frac{1}{t}\right)\psi(\bar x) \geq \psi(\bar x) + \frac{\varepsilon}{r}\|x - \bar x\|, \quad \forall\ x \in \Omega \setminus \overline{\mathcal{L}_\delta}.$$

Clearly, this shows that condition (iii) is satisfied. Moreover, since (iii) immediately implies condition (iv), we only need to verify the direction "(iv) ⇒ (i)" to finish the proof of this Lemma. However, since the equivalence of condition (i) and (iv) is already well-known for a more general setting, see, e.g., [11, Proposition 11.11], we are done here. □

Now, we want to prove that every sequence $(x^k)_k \subset \Omega$ that satisfies $F^{\Lambda_k}(x^k) \to 0$, as $k \to \infty$, has to be automatically bounded. We will see that the real valuedness and positive homogeneity of φ plays a central role in the proof of this claim.

Lemma 4.2.4. *Let $\Omega \subset \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$, and the mapping $\varphi : \mathbb{R}^n \to \mathbb{R}$ satisfy the conditions* (A.1)*,* (C.1)*–*(C.3)*, and let the sequences $(x^k)_k \subset \Omega$, $(\Lambda_k)_k \subset \mathbb{S}^n_{++}$ be arbitrary. Moreover, suppose that condition* (B) *holds and that the sequence $(F^{\Lambda_k}(x^k))_k$ converges to* 0 *as $k \to \infty$. Then, $(x^k)_k$ remains in a bounded set $\Omega_0 \subset \mathbb{R}^n$.*

Proof. Due to Lemma 4.2.3, there exist a point $\bar x \in \mathrm{dom}\ \varphi = \mathbb{R}^n$ and a suitable constant $\vartheta > 0$ such that

$$\text{(4.2.2)} \qquad f(x) - f(\bar x) \geq \varphi(\bar x) - \varphi(x) + \vartheta\|\bar x - x\|$$

for all $x \in \mathbb{R}^n$ sufficiently large. To prove that $(x^k)_k$ is bounded, assume in contrary that there exists a subsequence $(x^k)_{K_1}$ of $(x^k)_k$ with $\|x^k\| \to \infty$ as $K_1 \ni k \to \infty$. Then, we have $(F^{\Lambda_k}(x^k))_{K_1} \to 0$ and, by applying the calculation rule (3.2.8), assumption (B), and the compactness of $\partial\varphi(0)$ (see Lemma 2.5.12 (i) and the discussion after Lemma 2.5.13), there exists another subsequence $(x^k)_{K_2}$ of $(x^k)_{K_1}$ such that

$$\text{(4.2.3)}\qquad \nabla f(x^k) = \Lambda_k F^{\Lambda_k}(x^k) - \mathcal{P}^{\Lambda_k^{-1}}_{\partial\varphi(0)}(\Lambda_k x^k - \nabla f(x^k)) \to g^* \quad \text{and} \quad g^* \in \partial\varphi(0),$$

as $K_2 \ni k \to \infty$. Now, let $k \in K_2$ be sufficiently large and let us define

$$\begin{aligned} p^k &:= x^k - F^{\Lambda_k}(x^k) = \operatorname{prox}^{\Lambda_k}_{\varphi}(x^k - \Lambda_k^{-1}\nabla f(x^k)), \\ q^k &:= \Lambda_k F^{\Lambda_k}(x^k) - \nabla f(x^k) = \mathcal{P}^{\Lambda_k^{-1}}_{\partial\varphi(0)}(\Lambda_k x^k - \nabla f(x^k)). \end{aligned}$$

Then, we obtain the following estimate

$$\begin{aligned} &\vartheta + \frac{\varphi(\bar{x}) - \varphi(F^{\Lambda_k}(x^k))}{\|\bar{x} - x^k\|} \\ &\qquad \le \vartheta + \frac{\varphi(x^k - F^{\Lambda_k}(x^k)) + \varphi(\bar{x}) - \varphi(x^k)}{\|\bar{x} - x^k\|} \\ &\qquad \le \frac{\varphi(p^k) + f(x^k) - f(\bar{x})}{\|\bar{x} - x^k\|} \le \frac{\varphi(p^k) + \langle \nabla f(x^k), x^k - \bar{x}\rangle}{\|\bar{x} - x^k\|} \\ &\qquad = \frac{\varphi(p^k) - \langle q^k, x^k - F^{\Lambda_k}(x^k)\rangle - \langle q^k, F^{\Lambda_k}(x^k) - \bar{x}\rangle + \langle \Lambda_k F^{\Lambda_k}(x^k), x^k - \bar{x}\rangle}{\|\bar{x} - x^k\|} \\ &\qquad \le \frac{\varphi(p^k) - (\varphi(p^k) + \varphi^*(q^k))}{\|\bar{x} - x^k\|} + \|q^k\| \cdot \frac{\|F^{\Lambda_k}(x^k) - \bar{x}\|}{\|\bar{x} - x^k\|} + \lambda_M \cdot \|F^{\Lambda_k}(x^k)\|, \end{aligned}$$

where we used the subadditivity of φ, the convexity of f, Moreau's decomposition principle (3.2.4) and inequality (4.2.2). Next, let us further define

$$q_{\max} := \max_{q \in \partial\varphi(0)} \|q\|.$$

Due to $\varphi^*(q^k) = \sigma^*_{\partial\varphi(0)}(q^k) = \iota_{\partial\varphi(0)}(q^k) = 0$ and by taking the limit $K_2 \ni k \to \infty$, we establish the contradiction

$$\begin{aligned} \vartheta = \lim_{K_2 \ni k \to \infty} &\left\{ \vartheta + \frac{\varphi(\bar{x}) - \varphi(F^{\Lambda_k}(x^k))}{\|\bar{x} - x^k\|} \right\} \\ &\le \lim_{K_2 \ni k \to \infty} \left\{ \lambda_M \cdot \|F^{\Lambda_k}(x^k)\| + q_{\max} \frac{\|F^{\Lambda_k}(x^k) - \bar{x}\|}{\|\bar{x} - x^k\|} \right\} = 0. \end{aligned}$$

Hence, the sequence $(x^k)_k$ is bounded, as desired. □

Remark 4.2.5. If the set Ω is bounded, i.e., if the effective domain dom φ is bounded, then we can use a much easier and more direct argumentation. In particular, due to the simple fact

$$p^k = \operatorname{prox}^{\Lambda_k}_{\varphi}(x^k - \Lambda_k^{-1}\nabla f(x^k)) \in \operatorname{dom}\varphi, \quad \forall\, k \in \mathbb{N},$$

we immediately see that the sequence $(x^k)_k$ has to be bounded in this situation. Moreover, this argument does also clearly not depend on the convexity of f.

Lemma 4.2.6. *Let $\Omega \subset \mathbb{R}^n$, $f : \Omega \to \mathbb{R}$, and the function $\varphi : \mathbb{R}^n \to \mathbb{R}$ satisfy the conditions* (A.1), (C.1)–(C.3), *and consider the sequences $(x^k)_k \subset \Omega$, $(\Lambda_k)_k \in \mathbb{S}^n_{++}$. Suppose that assumption* (B) *holds and let the sets $\mathcal{K}_P \cup \mathcal{K}_N$ be a disjoint partitioning of $\mathbb{N}$ such that $\psi(x^k) \leq \psi(x^{k-1})$ for all $k \in \mathcal{K}_P$ and either $\mathcal{K}_N$ is finite (or empty) or $F^{\Lambda_k}(x^k) \to 0$ for $\mathcal{K}_N \ni k \to \infty$. Then, the sequence of iterates $(x^k)_k$ remains in a compact set $\Omega_0 \subset \mathbb{R}^n$.*

Proof. Due to Lemma 4.2.3 all level sets of the objective function ψ are compact. Hence, it suffices to show that all iterates are contained in an appropriate level set $\mathrm{lev}_\alpha\, \psi$. If the set $\mathcal{K}_N$ is finite, then, due to $\psi(x^k) \leq \psi(x^{k-1})$ for all $k \in \mathcal{K}_P$, we obtain $x^k \in \mathrm{lev}_\alpha\, \psi$ for all $k \geq 0$, where $\alpha := \max\{\psi(x^r)\,;\, r \in \{0\} \cup \mathcal{K}_N\}$. Next, Lemma 4.2.4 shows that $(x^k)_{\mathcal{K}_N}$ is bounded, if $\mathcal{K}_N$ contains infinitely many elements. Further, there holds

$$\|\nabla f(x^k)\| = \|\Lambda_k F^{\Lambda_k}(x^k) + \mathcal{P}^{\Lambda_k^{-1}}_{\partial\varphi(0)}(\Lambda_k x^k - \nabla f(x^k))\| \leq \lambda_M \|F^{\Lambda_k}(x^k)\| + q_{\max},$$

where $q_{\max} := \max_{q \in \partial\varphi(0)} \|q\|$ and we used the compactness of $\partial\varphi(0)$. By applying the subdifferential inequality for the convex functions f and φ and Example 2.5.17, we obtain

$$\begin{aligned}\psi(x^k) - \psi(y) &\leq \langle \nabla f(x^k), x^k - y\rangle + \sigma_{\partial\varphi(0)}(x^k) - \varphi(y)\\ &\leq \|\nabla f(x^k)\|\|x^k - y\| + q_{\max}\|x^k - y\| \leq (\lambda_M \|F^{\Lambda_k}(x^k)\| + 2q_{\max})\|x^k - y\|\end{aligned}$$

for some fixed $y \in \mathrm{dom}\ \varphi$. Since the sequence $(x^k)_{\mathcal{K}_N}$ is bounded, we see that

$$\alpha := \max_{k \in \{0\} \cup \mathcal{K}_N} \psi(x^k) < \infty.$$

From $\psi(x^k) \leq \psi(x^{k-1})$ for all $k \in \mathcal{K}_P$ we thus conclude that $x^k \in \mathrm{lev}_\alpha\, \psi$ for all $k \geq 0$. □

4.2.2. A proximal gradient method with an Armijo-type linesearch technique

We now consider a globalized proximal gradient method that we use for the step generation whenever the semismooth Newton step is not accepted. It is advantageous to analyze this method separately before proceeding the development and investigation of the final overall algorithm.

Let x^k denote the current iterate and let $d^k := -F^{\Lambda_k}(x^k)$ be a direction that is generated by the fixed point-type equation $(\mathcal{E})$. Then, the globalized proximal point method calculates $x^{k+1} = x^k + \sigma_k d^k$, where the step size σ_k is controlled by a quasi-Armijo rule. The details are formulated in Algorithm 1. We use the following notation:

$$u(x^k) := x^k - \Lambda_k^{-1}\nabla f(x^k), \quad \Delta^k := -(\nabla f(x^k))^\top F^{\Lambda_k}(x^k) + \varphi(\mathrm{prox}_\varphi^{\Lambda_k}(u(x^k))) - \varphi(x^k).$$

In the following, we will show that Algorithm 1 is a globally convergent descent method. Further properties of Algorithm 1 will be discussed later together with the convergence analysis of our main approach.

Let us mention that over the last years the proximal gradient method has established itself as one of the most common and basic first order methods for convex composite problems.

Algorithm 1: Proximal gradient method with quasi-Armijo rule

```
S0 Initialization: Choose $x^0 \in \operatorname{dom}\varphi$, $\Lambda_0 \in \mathbb{S}^n_{++}$, and $\beta, \gamma \in (0,1)$. Set iteration $k := 0$.
   while $F^{\Lambda_k}(x^k) \neq 0$ do
S1     Compute a new direction $d^k = -F^{\Lambda_k}(x^k) = \operatorname{prox}_\varphi^{\Lambda_k}(u(x^k)) - x^k$.
S2     Choose a maximal quasi-Armijo step size $\sigma_k \in \{1, \beta, \beta^2, ....\}$ with
           $\psi(x^k + \sigma_k d^k) \leq \psi(x^k) + \sigma_k \gamma \Delta^k$.
S3     Set $x^{k+1} = x^k + \sigma_k d^k$ and choose $\Lambda_{k+1} \in \mathbb{S}^n_{++}$.
       $k \leftarrow k + 1$.
```

Moreover, besides Algorithm 1, many variants and alternative approaches have been proposed and developed. For instance, the so-called *BB-methods*, which combine nonmonotone linesearch techniques and the Barzilai-Borwein spectral approach [267, 10], represent a popular and frequently used class of extended proximal gradient methods. Here, characteristically, the parameter matrix Λ_k is chosen via $\Lambda_k = (\lambda^k_{BB1})^{-1} I$ or $\Lambda_k = (\lambda^k_{BB2})^{-1} I$, where

$$\lambda^k_{BB1} := \frac{(\bar{x}^{k-1})^\top \bar{x}^{k-1}}{(\bar{x}^{k-1})^\top \bar{g}^{k-1}}, \qquad \lambda^k_{BB2} := \frac{(\bar{x}^{k-1})^\top \bar{g}^{k-1}}{(\bar{g}^{k-1})^\top \bar{g}^{k-1}},$$

and $\bar{x}^{k-1} := x^k - x^{k-1}$, $\bar{g}^{k-1} := \nabla f(x^k) - \nabla f(x^{k-1})$. These nonmonotone variants of Algorithm 1 have been successfully implemented in many different algorithms, as, e.g., SpaRSA [256], FPC-AS, [254], TVAL3 [136], or curvilinear search methods [252, 251]. In this work, we will focus on the very basic version of the proximal gradient method that guarantees descent in the objective function at each iteration to facilitate the convergence analysis of the augmented semismooth Newton method.

Let us start with the following descent property.

Lemma 4.2.7 (Descent directions). *Let assumption* (A.1) *hold and let the sequences* $(x^k)_k$ *and* $(d^k)_k$ *be generated by Algorithm 1. Then for all* $k \geq 0$ *it holds*

$$\Delta^k \leq -\|d^k\|^2_{\Lambda_k} = -\|F^{\Lambda_k}(x^k)\|^2_{\Lambda_k}.$$

Proof. At first, let us note that the update formula in step **S3** of Algorithm 1 and $\sigma_k \in [0,1]$ imply $x^k \in \operatorname{dom}\varphi$ for all $k \in \mathbb{N}$. Thus, due to assumption (A.1), the terms $u(x^k)$ and $\nabla f(x^k)$ are well-defined for all k. Next, using the characterization (3.1.2) of the proximity operator $\operatorname{prox}_\varphi^{\Lambda_k}(u(x^k))$ and Lemma 3.1.5, we obtain

$$\begin{aligned}\Delta^k &\leq -\langle \nabla f(x^k), F^{\Lambda_k}(x^k)\rangle + \langle \nabla \operatorname{env}_\varphi^{\Lambda_k}(u(x^k)), \operatorname{prox}_\varphi^{\Lambda_k}(u(x^k)) - x^k\rangle \\ &= -\langle \nabla f(x^k), F^{\Lambda_k}(x^k)\rangle - \langle \Lambda_k F^{\Lambda_k}(x^k) - \nabla f(x^k), F^{\Lambda_k}(x^k)\rangle = -\|F^{\Lambda_k}(x^k)\|^2_{\Lambda_k},\end{aligned}$$

as desired. □

Lemma 4.2.8. *Suppose that the assumptions* (A.1)–(A.2) *and condition* (B) *hold and let the sequences* $(x^k)_k$ *and* $(d^k)_k$ *be generated by Algorithm 1. Then, there exists a constant* $\zeta > 0$ *for all* $k \in \mathbb{N}$ *with*

$$\psi(x^k + \sigma d^k) \le \psi(x^k) + \sigma\gamma\Delta^k \quad \textit{for all } \sigma \in [0, \zeta]. \tag{4.2.4}$$

Proof. Apparently $\sigma = 0$ fulfills (4.2.4). So, let us consider $\sigma \in (0, 1]$ sufficiently small, then we obtain for arbitrary but fixed $k \in \mathbb{N}$

$$\begin{aligned}\frac{\psi(x^k + \sigma d^k) - \psi(x^k)}{\sigma} - \gamma\Delta^k &\le \frac{f(x^k + \sigma d^k) - f(x^k)}{\sigma} - \nabla f(x^k)^\top d^k + (1-\gamma)\Delta^k \\ &\le \int_0^1 (\nabla f(x^k + \sigma t d^k) - \nabla f(x^k))^\top d^k \,\mathrm{d}t - \lambda_m(1-\gamma)\|d^k\|^2 \\ &\le \left(\frac{L_f}{2}\sigma - \lambda_m(1-\gamma)\right)\|d^k\|^2,\end{aligned}$$

where we used the convexity of φ, Lemma 4.2.7 and assumption (B). At this point, let us emphasize that the convexity of the set dom φ and $\sigma, t \in [0, 1]$ imply

$$x^k + \sigma t d^k = (1 - t\sigma)x^k + t\sigma \cdot \mathrm{prox}_\varphi^{\Lambda_k}(u(x^k)) \in \mathrm{dom}\ \varphi \subset \Omega, \quad \forall\ k \in \mathbb{N}.$$

Hence, the expression $\nabla f(x^k + \sigma t d^k)$ is well-defined for all k. Moreover, the quasi-Armijo condition (4.2.4) is satisfied whenever

$$\sigma \le \zeta := \min\left\{\frac{2\lambda_m(1-\gamma)}{L_f}, 1\right\}.$$

□

Remark 4.2.9. Lemma 4.2.8 shows that every step size sequence $(\sigma_k)_k$ generated by Algorithm 1, is uniformly bounded from below (whenever our assumptions hold). To be more precise, we have

$$\sigma_k \ge \beta\zeta > 0, \quad \forall\ k \ge 0. \tag{4.2.5}$$

The following convergence result was first established by Tseng and Yun [236]. In contrast to many other convergence analyses, Lipschitz continuity of the gradient ∇f or boundedness of the iterates is not required.

Theorem 4.2.10 (Global convergence). *Suppose that the assumptions* (A.1) *and* (B) *are satisfied and let the sequences* $(x^k)_k$ *and* $(\Lambda_k)_k$ *be generated by Algorithm 1. Then, every accumulation point* x^* *of* $(x^k)_k$ *satisfies* $F^\Lambda(x^*) = 0$, $\Lambda \in \mathbb{S}^n_{++}$, *and is thus a stationary point of the nonsmooth problem* $(\mathcal{P})$.

Proof. This theorem is an application of a more general result of Tseng and Yun; see [236, Theorem 1b]. For the sake of completeness, we also provide a simplified proof in the Appendix A.1. □

4.2.3. A multidimensional filter framework

We adopt the multidimensional filter globalization concept of [95, 96] and tailor it to a semismooth Newton method for solving the equation

$$F^{\Lambda}(x) = 0, \quad \Lambda \in \mathbb{S}^n_{++}. \tag{4.2.6}$$

We will apply the filter to accept or reject semismooth Newton steps. The *filter value* corresponding to a point $x \in \mathbb{R}^n$ is given by $\theta(x)$, where the *filter function* $\theta : \mathbb{R}^n \to \mathbb{R}^p_+$ is continuous and satisfies

$$c_\theta \|F^{\Lambda}(x)\|_\infty \leq \|\theta(x)\|_\infty \leq C_\theta \|F^{\Lambda}(x)\|_\infty \tag{4.2.7}$$

with constants $0 < c_\theta < C_\theta$. This ensures that $(\theta(x^k))_k$ is bounded if and only if the sequence $(F^{\Lambda}(x^k))_k$ is bounded and that the filter function θ and the function F^{Λ} have the same set of zeros. Hence, $\bar{x}$ is a stationary point of the minimization problem $(\mathcal{P})$ if and only if $\theta(\bar{x}) = 0$. Next, we give a typical example for a filter function θ.

Example 4.2.11. A standard approach for choosing θ (with many possible variants) is to decompose $\{1, \ldots, n\}$ into p possibly overlapping nonempty sets $\mathcal{I}_j$ with $\bigcup_{j=1}^{p} \mathcal{I}_j = \{1, 2, ..., n\}$. The function θ is then defined as

$$\theta_j(x) := \frac{1}{\sqrt{|\mathcal{I}_j|}} \|F^{\Lambda}_{\mathcal{I}_j}(x)\| = \left(\frac{1}{|\mathcal{I}_j|} \sum\nolimits_{i \in \mathcal{I}_j} F^{\Lambda}_i(x)^2 \right)^{1/2}, \quad \forall\, j \in \{1, ..., p\}. \tag{4.2.8}$$

This choice satisfies condition (4.2.7) with $c_\theta := 1/\sqrt{\max_j |\mathcal{I}_j|}$ and $C_\theta := 1$. The selection of p and of the set $\mathcal{I}_j$ can be based on the characteristics of the problem. For instance, if we choose $p = n$ and $\mathcal{I}_j = \{j\}$, then we obtain

$$\theta(x) = (|F^{\Lambda}_1(x)|, |F^{\Lambda}_2(x)|, ..., |F^{\Lambda}_n(x)|)^\top.$$

Remark 4.2.12. Since we are also interested in situations where the parameter matrix Λ depends on the current iteration k, it is natural to ask whether the filter concept does also work for sequences of the form $(F^{\Lambda_k}(x^k))_k$. Clearly, if the parameter matrices $(\Lambda_k)_k$ satisfy assumption (B), then Lemma 4.1.3 is applicable and the boundedness condition (4.2.7) will also hold for $F^{\Lambda_k}(x^k)$,

$$c_\theta \|F^{\Lambda_k}(x^k)\|_\infty \leq \|\theta(x^k)\|_\infty \leq C_\theta \|F^{\Lambda_k}(x^k)\|_\infty,$$

but of course with other constants $0 < c_\theta \leq C_\theta$. Furthermore, in Example 4.2.11, the matrix Λ can also be replaced by an arbitrary and changing parameter matrix $\Lambda_k \in \mathbb{S}^n_{++}$, $k \in \mathbb{N}$. Thus, the filter function need not be necessarily restricted to fixed parameter matrices. However, to emphasize the dependency of the filter function and the filter values on Λ, we will also work with the following notational variant

$$\theta : \mathbb{R}^n \times \mathbb{S}^n_{++} \to \mathbb{R}^p_+, \quad (x, \Lambda) \mapsto \theta(x, \Lambda).$$

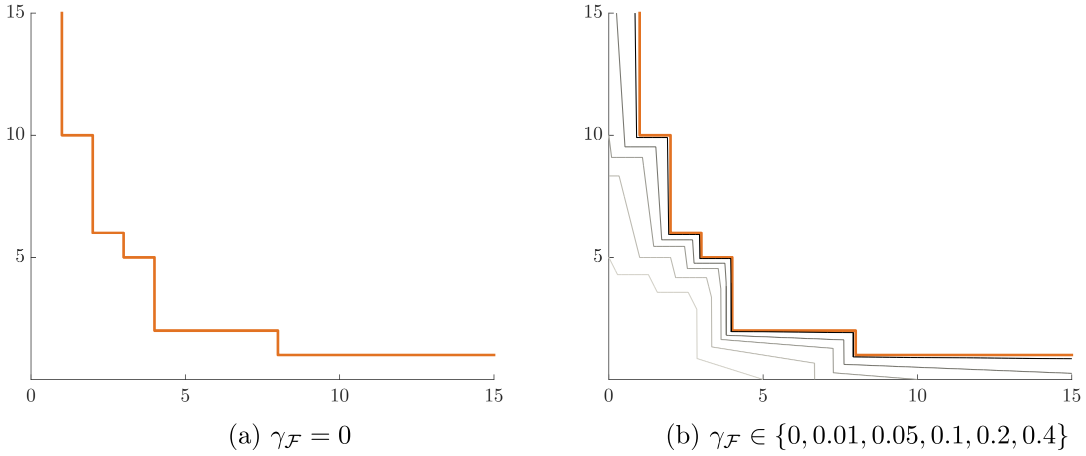


Figure 4.1.: Example of a two-dimensional filter and of the filter acceptance criterion. In subfigure (a), the acceptance test (4.2.11) is illustrated for $\gamma_{\mathcal{F}} = 0$. In particular, each point that lies above the orange line is dominated by a filter entry and is not acceptable to the filter. In subfigure (b), the same situation is shown for different values of $\gamma_{\mathcal{F}}$.

In particular, this extension of the filter function will turn out to be useful when comparing different filter values $\theta(x^k) \equiv \theta(x^k, \Lambda_k)$ and $\theta(x^{k+1}) \equiv \theta(x^{k+1}, \Lambda_{k+1})$. Since we will always work with bounded parameter matrices, our following discussion focuses on the basic definition of the filter function that does not explicitly include the parameter matrix Λ. Moreover, if the filter concept is used within an algorithmic framework, we will adhere to the convention $\theta(x^k) \equiv \theta(x^k, \Lambda_k)$.

Now, assume that a filter function $\theta : \mathbb{R}^n \to \mathbb{R}^p_+$ has been chosen (e.g., according to (4.2.8) for some $p \leq n$). At iteration k, the *filter* $\mathcal{F}_k \subset \mathbb{R}^p_+$ is a finite collection of *filter entries* $q \in \mathbb{R}^p_+$, where usually (and in our context always) each $q \in \mathcal{F}_k$ corresponds to a point $x \in \mathbb{R}^n$, via $q = \theta(x)$ and the points x are selected iterates x^ℓ, $\ell < k$, of the method to be globalized. In our case, these points are a subset of the iterates generated by semismooth Newton steps for (4.2.6). Similar to [95], we define an acceptance criterion for a point x.

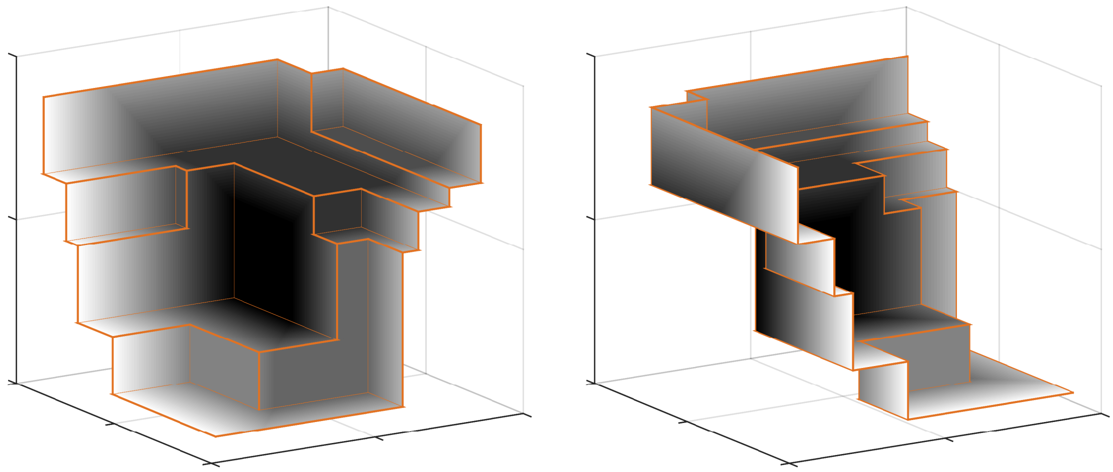
Definition 4.2.13 (Filter acceptance criterion). *A point $x \in \mathbb{R}^n$ is said to be* acceptable to the filter $\mathcal{F}_k \subset \mathbb{R}^p_+ \setminus \{0\}$ *if*

$$\max_{1 \leq j \leq p} \big(q_j - \theta_j(x)\big) \geq \gamma_{\mathcal{F}}\, \delta(q, \theta(x)) \tag{4.2.9}$$

holds for all $q \in \mathcal{F}_k$. Here, $\gamma_{\mathcal{F}} \in (0, 1)$ is fixed and $\delta : \mathbb{R}^p_+ \times \mathbb{R}^p_+ \to \mathbb{R}_+$ is continuous and satisfies for all $q \in \mathbb{R}^p_+$

$$\delta(q, q) = 0 \quad \Longrightarrow \quad q = 0. \tag{4.2.10}$$

If the new iterate $x^{k+1} \in \mathbb{R}^n$ is acceptable to the current filter $\mathcal{F}_k$, we can, if we wish, update the filter by adding $\theta(x^{k+1})$ to the filter: $\mathcal{F}_{k+1} := \mathcal{F}_k \cup \{\theta(x^{k+1})\}$. If the filter is



(a) (b)

Figure 4.2.: Example of a three-dimensional filter. The subfigures visualize the acceptance criterion (4.2.11) for $\gamma_{\mathcal{F}} = 0$ from two different perspectives. The coloring changes with the norm of the respective points.

not updated, then we set $\mathcal{F}_{k+1} := \mathcal{F}_k$. After each update the filter can be scanned for redundant entries that no longer have influence on the acceptance rule and consequently can be removed. More details can be found in [95]. Returning to the acceptance rule, there are many suitable choices for the function δ. Here, we will work with $\delta(q, \theta(x)) := \|\theta(x)\|_\infty$. Then, the corresponding acceptance test

$$\max_{1\le j\le p} \big(q_j - \theta_j(x)\big) \ge \gamma_{\mathcal{F}} \|\theta(x)\|_\infty, \quad \forall\, q \in \mathcal{F}_k \tag{4.2.11}$$

ensures the uniform boundedness of the filter entries. The filter concept yields convergence in the following sense.

Lemma 4.2.14. *Let $\theta : \mathbb{R}^n \to \mathbb{R}^p_+$ be a filter function and let $\delta : \mathbb{R}^p_+ \times \mathbb{R}^p_+ \to \mathbb{R}_+$ satisfy condition (4.2.10). Furthermore, let $(x^k)_K$ be an infinite subsequence of iterates such that $(\theta(x^k))_{k\in K}$ is bounded, x^k is acceptable to $\mathcal{F}_{k-1}$ for all $k \in K \setminus \{0\}$ and the filter is updated, i.e., $\mathcal{F}_k = \mathcal{F}_{k-1} \cup \{\theta(x^k)\}$, for all $k \in K$. Then it holds*

$$\lim_{K\ni k\to\infty} \theta(x^k) = 0.$$

Proof. Since the sequence $(\theta(x^k))_{k\in K}$ is bounded, there exists a subsequence $(\theta(x^{k_\ell}))_\ell$, $k_\ell \in K$, that converges to an accumulation point $\theta^* \in \mathbb{R}^p_+$. Since $\theta(x^{k_\ell})$ is acceptable to $\mathcal{F}_{k_\ell-1} \supseteq \mathcal{F}_{k_{\ell-1}} \supseteq \{\theta(x^{k_{\ell-1}})\}$, there holds

$$\max_{1\le j\le p} \big(\theta_j(x^{k_{\ell-1}}) - \theta_j(x^{k_\ell})\big) \ge \gamma_{\mathcal{F}}\, \delta(\theta(x^{k_{\ell-1}}), \theta(x^{k_\ell})).$$

Taking the limit $\ell \to \infty$, we obtain

$$0 = \max_{1 \le j \le p} \, (\theta_j^* - \theta_j^*) \ge \gamma_{\mathcal{F}} \, \delta(\theta^*, \theta^*),$$

where we used the continuity of δ. Applying the second part of condition (4.2.10), the last equation implies $\theta^* = 0$. □

Remark 4.2.15. If we choose $\delta(q, \theta(x)) := \|\theta(x)\|_\infty$ as in (4.2.11), Lemma 4.2.14 holds without explicitly assuming the boundedness of the filter entries. In fact, since x^k is acceptable to $\mathcal{F}_{k-1}$ we then have, for all $q \in \mathcal{F}_{k-1}$, that

$$\gamma_{\mathcal{F}} \|\theta(x^k)\|_\infty \le \max_{1 \le j \le p} \big(q_j - \theta_j(x)\big) \le \|q\|_\infty.$$

Lemma 4.2.14 can be regarded as the essence of the multidimensional filter framework [95, 96]. The general idea to apply this concept is as follows: If a globally convergent base algorithm is given (in our case the proximal gradient method with quasi-Armijo step size rule) and an additional method for computing steps (in our case the semismooth Newton) shall be incorporated, then we can use the filter to control acceptance of the latter steps while resorting to steps of the base algorithm, otherwise. Then, any subsequence of points generated by filter steps tends to stationarity. If only finitely many filter steps are taken, then global convergence follows from the properties of the base algorithm. This implies that there exists a subsequence approaching stationarity. To prove that every accumulation point is stationary, the tricky part is the situation where infinitely many filter steps take place but only finitely many iterates resulting from filter steps are contained in the convergent subsequence. In this case it is required to show that the intermediate filter steps do not affect the convergence of the base algorithm.

We will formulate an algorithm of the described type in the next section. In the following sections we then will prove that all limit points are stationary along the lines just described.

4.2.4. The full algorithm

We now derive a semismooth Newton method for the minimization problem $(\mathcal{P})$ for both convex and nonconvex f. The method uses the following nonsmooth equation form of the optimality conditions:

$$F^\Lambda(x) = x - \operatorname{prox}_\varphi^\Lambda(x - \Lambda^{-1}\nabla f(x)) = 0, \quad \Lambda \in \mathbb{S}^n_{++}. \tag{4.2.12}$$

This results in the nonsmooth Newton system

$$M(x^k)s^k = -F^{\Lambda_k}(x^k), \quad \Lambda_k \in \mathbb{S}^n_{++}, \tag{4.2.13}$$

where $M(x^k)$ denotes a generalized derivative of F^{Λ_k} in x^k and the parameter matrices Λ_k may be chosen differently or adaptively in each iteration.

In chapter 3, we have seen that the class of functions φ that guarantees semismoothness of the proximity operator $\operatorname{prox}_\varphi^{\Lambda_k}$ and thus of F^{Λ_k} is quite large and contains many well-known and important examples, such as the ℓ_1-, ℓ_2- or the nuclear norm. However, further structural information about the proximity operator $\mathrm{prox}_{\varphi}^{\Lambda_k}$ are needed to explicitly construct and choose an appropriate set of generalized derivatives $M(x^k)$. In the following, we want to assume that a suitable realization of $M(x^k)$ is always available.

Before we state the full algorithm, let us consider several concrete examples to illustrate the semismooth Newton step (4.2.15) and the construction of $M(x^k)$.

Example 4.2.16 (ℓ_1-optimization). In the following, we consider an ℓ_1-regularized optimization problem of the form

$$\min_{x \in \mathbb{R}^n} \; f(x) + \mu \|x\|_1,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function and $\mu > 0$ is a regularization parameter. Let $\lambda \in \mathbb{R}^n_{++}$ be arbitrary and let us set $\Lambda := \mathrm{diag}(1 \oslash \lambda)$. Then, due to the separability of the ℓ_1-norm and by using Example 3.2.8 and equation (3.2.7), we obtain

$$F^\Lambda(x) = \Lambda^{-1} \nabla f(x) + \mathcal{P}_{[-\mu\lambda, \mu\lambda]^n}(x - \Lambda^{-1}\nabla f(x)) = 0. \tag{4.2.14}$$

This results in the nonsmooth Newton system

$$M(x^k) s^k = -F^\Lambda(x^k), \tag{4.2.15}$$

where $M(x^k)$ denotes a generalized derivative of F^Λ in x^k. Since the function F^Λ is piecewise continuously differentiable, it is also semismooth at all $x \in \mathbb{R}^n$, see, e.g., [211]. Furthermore, setting $u(x) := x - \Lambda^{-1}\nabla f(x)$ and applying Remark 2.5.22, it holds

$$\begin{aligned} \partial F^\Lambda(x) &= \Lambda^{-1}\nabla^2 f(x) + \partial(\mathcal{P}_{[-\mu\lambda,\mu\lambda]^n} \circ u)(x) \\ &\subset \Lambda^{-1}\nabla^2 f(x) + \partial_C \mathcal{P}_{[-\mu\lambda,\mu\lambda]^n}(u(x)) \circ (I - \Lambda^{-1}\nabla^2 f(x)). \end{aligned}$$

In particular, choosing

$$M(x) := (I - D(x))\Lambda^{-1}\nabla^2 f(x) + D(x), \tag{4.2.16}$$

where the diagonal matrix $D(x)$ is defined component-wise

$$D(x)_{[ii]} \begin{cases} = 0 & \text{if } |u_i(x)| > \mu\lambda_i, \\ \in \{0,1\} & \text{if } |u_i(x)| = \mu\lambda_i, \\ = 1 & \text{if } |u_i(x)| < \mu\lambda_i, \end{cases} \quad \forall\; i = 1, ..., n,$$

we have that $M(x)$ is the Jacobian of one of the smooth active pieces that define F^Λ at x. Denoting by $\partial_{PW} F^\Lambda(x)$ the collection of all these $M(x)$, there holds $\partial_B F^\Lambda(x) \subset \partial_{PW} F^\Lambda(x)$ and F^Λ is semismooth w.r.t. $\partial_{PW} F^\Lambda$. More specifically, there holds

$$\|F^\Lambda(x+s) - F^\Lambda(x) - M(x+s)\| = o(\|s\|)$$

uniformly for all $M(x+s) \in \partial_{PW} F^\Lambda(x+s)$ as $s \to 0$; see [124] and Remark 2.6.6. In our numerical comparison, we will work with the unique choice for $M(x)$ that results when we select $D(x)_{[ii]} = 1$ in the case $|u_i(x)| = \mu\lambda_i$.

Example 4.2.17 (Constrained ℓ_1-optimization). Next, we discuss the following ℓ_1-type optimization problem,

$$\min_x \ f(x) + \mu\|x\|_1 + \iota_{[a,b]}(x).$$

This problem is a simple extension of the previous example with additional box constraints. Again, $f : \mathbb{R}^n \to \mathbb{R}$ is supposed to be twice continuously differentiable and we have $a, b \in [-\infty, +\infty]^n$, $\mu > 0$. Moreover, we set $\Lambda := \operatorname{diag}(1 \oslash \lambda)$ for $\lambda \in \mathbb{R}^n_{++}$. Then, by using Example 3.2.9, the proximity operator $\operatorname{prox}^\Lambda_\varphi(x)$, $\varphi(x) := \mu\|x\|_1 + \iota_{[a,b]}(x)$, is given by

$$\operatorname{prox}^\Lambda_\varphi(x) = \mathcal{P}_{[a,b]}(x - \mathcal{P}_{[-\mu\lambda,\mu\lambda]^n}(x)).$$

Thus, setting $u(x) := x - \Lambda^{-1}\nabla f(x)$, we obtain

$$F^\Lambda(x) = x - \mathcal{P}_{[a,b]}(u(x) - \mathcal{P}_{[-\mu\lambda,\mu\lambda]^n}(u(x))).$$

Since this function is again piecewise continuously differentiable, it is semismooth on $\mathbb{R}^n$ and we can reuse the basic construction of Example 4.2.16. In particular, since the projections in $F^\Lambda(x)$ are applied component-wise, we can utilize [50, Theorem 2.3.9] and Theorem 2.5.21 (i) to construct a generalized derivative of F^Λ at x. It holds

$$\begin{aligned}\partial(\mathcal{P}_{[a_i,b_i]} \circ w_i)(x) &\subset \operatorname{conv}(\partial\mathcal{P}_{[a_i,b_i]}(w_i(x)) \circ \partial w_i(x)) \\ &\subset \operatorname{conv}(\partial\mathcal{P}_{[a_i,b_i]}(w_i(x)) \cdot (1 - \partial\mathcal{P}_{[-\mu\lambda_i,\mu\lambda_i]}(u_i(x))) \cdot \nabla u_i(x)^\top) = \mathcal{D}_i(x)\nabla u_i(x)^\top,\end{aligned}$$

where $w(x) := u(x) - \mathcal{P}_{[-\mu\lambda,\mu\lambda]^n}(u(x))$ and the set $\mathcal{D}_i(x) \subset \mathbb{R}$ is defined via

$$\delta \in \mathcal{D}_i(x) \quad :\Longleftrightarrow \quad \delta \begin{cases} = 0 & \text{if } w_i(x) \notin [a_i, b_i] \ \vee \ |u_i(x)| < \mu\lambda_i, \\ = 1 & \text{if } w_i(x) \in (a_i, b_i) \ \wedge \ |u_i(x)| > \mu\lambda_i, \\ \in [0,1] & \text{otherwise.} \end{cases}$$

Clearly, due to $\partial F^\Lambda(x) \subset \partial_C F^\Lambda(x) \subset I - \mathcal{D}_1(x)\nabla u_1(x)^\top \times ... \times \mathcal{D}_n(x)\nabla u_n(x)^\top$, this leads to the following possible choice of generalized derivatives

$$M(x) = I - D(x)(I - \Lambda^{-1}\nabla^2 f(x)), \quad D(x) = \operatorname{diag}(\delta_1, ..., \delta_n)$$

with $\delta_i \in \mathcal{D}_i(x)$ for all $i = 1, ..., n$ and the corresponding semismooth Newton system has the form

$$M(x^k)s^k = -F^\Lambda(x^k).$$

Furthermore, similar to Remark 2.6.6 or [79, Theorem 19], it can be shown that the function F^Λ is also semismooth w.r.t. the possibly larger set $I - \mathcal{D}_1(x)\nabla u_1(x)^\top \times ... \times \mathcal{D}_n(x)\nabla u_n(x)^\top$.

Example 4.2.18 (Group sparse optimization). Here, we consider optimization problems with a group sparse penalty term

$$\min_{x\in\mathbb{R}^n} \ f(x) + \sum_{i=1}^s \omega_i\|x_{g_i}\|_2, \quad \omega_i > 0, \quad i = 1, ..., s,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is again a twice continuously differentiable function and the index sets g_i, $i = 1, ..., s$ form a disjoint partitioning of the set $\{1, ..., n\}$. Moreover, let $\lambda \in \mathbb{R}^s_+$ be an arbitrary vector and let us define the parameter matrix $\Lambda \in \mathbb{S}^n_{++}$,

$$\Lambda_{[g_i g_j]} = \begin{cases} \frac{1}{\lambda_i} I & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \quad \forall\, 1 \leq i, j \leq s. \tag{4.2.17}$$

Then, due to Lemma 3.2.4, the proximity operator associated with $\varphi(x) := \sum_{i=1}^s \omega_i \|x_{g_i}\|_2$ and Λ can be computed group-wise via

$$\text{prox}_\varphi^\Lambda(x)_{g_i} = \text{prox}_{\omega_i \|\cdot\|_2}^{\lambda_i^{-1} I}(x_{g_i}), \quad \forall\, i = 1, ..., s.$$

Consequently, by using Example 3.2.8 and equation (3.2.7), we obtain

$$F^\Lambda(x)_{g_i} = \lambda_i \nabla f(x)_{g_i} + \mathcal{P}_{B_{\|\cdot\|_2}(0, \omega_i \lambda_i)}(x_{g_i} - \lambda_i \nabla f(x)_{g_i}), \quad \forall\, i = 1, ..., s,$$

and the semismooth Newton equation is given by

$$M(x^k) s^k = -F^\Lambda(x^k).$$

As usual, the matrix $M(x^k)$ denotes a suitable generalized derivative of F^Λ at x^k. Furthermore, since the mapping φ is the sum of semialgebraic functions, Corollary 3.3.4 implies that F^Λ is semismooth on $\mathbb{R}^n$. Now, let us set $u \equiv u(x) := x - \Lambda^{-1} \nabla f(x)$. In our numerical comparison in chapter 7, we will work with the following generalized derivatives

$$M(x) := (I - D(x)) \Lambda^{-1} \nabla^2 f(x) + D(x),$$

where the matrix $D(x)$ is defined block-wise via

$$D(x)_{[g_i g_j]} = 0, \quad D(x)_{[g_i g_i]} \begin{cases} = I & \text{if } \|u_{g_i}\|_2 < \omega_i \lambda_i, \\ \in \left\{ I - t u_{g_i} u_{g_i}^\top : t \in \left[0, \frac{1}{(\omega_i \lambda_i)^2}\right] \right\} & \text{if } \|u_{g_i}\|_2 = \omega_i \lambda_i, \\ = \frac{\omega_i \lambda_i}{\|u_{g_i}\|_2} I - \frac{\omega_i \lambda_i}{\|u_{g_i}\|_2^3} u_{g_i} u_{g_i}^\top & \text{if } \|u_{g_i}\|_2 > \omega_i \lambda_i, \end{cases}$$

for all $1 \leq i, j \leq s$ and $i \neq j$. In particular, it immediately follows

$$D(x)_{[g_i g_i]} \in \partial_C \mathcal{P}_{B_{\|\cdot\|_2}(0, \omega_i \lambda_i)}(u_{g_i}), \quad \forall\, i = 1, ..., s.$$

Hence, as in Example 4.2.16 and by Remark 2.5.22 and 2.6.6, this specific choice yields a suitable set of generalized derivatives and does again not affect the semismoothness of F^Λ.

We now continue with the description of our algorithmic approach.

For a given semismooth Newton step s^k the decision on accepting $x^k + s^k$ as new iterate is based on the filter framework presented in the last section. Consequently, we accept the trial point $x^k + s^k$ whenever it is acceptable for the current filter $\mathcal{F}_k$, i.e., whenever the filter value $\theta(x^k + s^k) \equiv \theta(x^k + s^k, \Lambda_{k+1})$ satisfies the acceptance test (4.2.11).

Algorithm 2: Semismooth Newton Method with Multi-Dimensional Filter Globalization

```
S0 Initialization: Choose an initial point $x^0 \in \mathrm{dom}\ \varphi$, $\Lambda_0 \in \mathbb{S}^n_{++}$, $\tau > 0$, $\beta, \gamma \in (0,1)$
   (quasi-Armijo parameters), $\gamma_{\mathcal{F}} \in (0,1), \mathcal{F}_{-1} = \emptyset$ (filter parameters) and $\alpha_i > 0$ for
   $i \in \{1,2,3\}$, $\eta \in (0,1)$. Set $k := 0$, $\psi_0 := \infty$, $\rho_0 := \infty$. Set iteration $k := 0$.
   while $F^{\Lambda_k}(x^k) \neq 0$ do
S1     If $k = 0$ or $x^k$ was obtained in step S3, add $\theta(x^k)$ to the filter: $\mathcal{F}_k = \mathcal{F}_{k-1} \cup \{\theta(x^k)\}$.
       Otherwise, set $\mathcal{F}_k = \mathcal{F}_{k-1}$. Choose $\Lambda_{k+1} \in \mathbb{S}^n_{++}$.
S2     Compute the semismooth Newton step $s^k$ via $M(x^k)s^k = -F^{\Lambda_k}(x^k)$. If this is not
       possible go to step S4.
S3     Set $x^{k+1} = x^k + s^k$ and check if $x^{k+1}$ lies in dom $\varphi$ and is acceptable for the filter $\mathcal{F}_k$:
           $\max_{1 \leq j \leq p} \left(q_j - \theta_j(x^{k+1})\right) \geq \gamma_{\mathcal{F}} \max_{1 \leq j \leq p} \theta_j(x^{k+1}), \quad \forall\, q \in \mathcal{F}_k.$
       If $x^{k+1}$ is acceptable for $\mathcal{F}_k$ and either $f$ is convex or (4.2.18) holds or (4.2.19) holds,
       set $\psi_{k+1} = \psi_k$, $\rho_{k+1} = \min\{\rho_k, \|F^{\Lambda_{k+1}}(x^{k+1})\|\}$ and skip step S4 and S5.
S4     Compute the direction $d^k = -F^{\Lambda_k}(x^k)$ and choose a maximal quasi-Armijo step
       $\sigma_k \in \{1, \beta, \beta^2, \beta^3, ...\} \subset (0,1]$ satisfying
           $\psi(x^k + \sigma_k d^k) \leq \psi(x^k) + \sigma_k \gamma \Delta^k.$
S5     Set $x^{k+1} = x^k + \sigma_k d^k$, $\psi_{k+1} = \psi(x^{k+1})$, and $\rho_{k+1} = \rho_k$.
       $k \leftarrow k + 1$.
```

In the convex case, if the trial point $x^k + s^k$ satisfies all conditions and is contained in dom φ, we accept the Newton step, set $x^{k+1} = x^k + s^k$, update the filter $\mathcal{F}_{k+1} = \mathcal{F}_k \cup \{\theta(x^k)\}$ and start the next iteration. Otherwise we reject the Newton step and perform a step of the globalized proximal gradient method $x^{k+1} = x^k + \sigma_k d^k$. In the nonconvex case, we require additional conditions for accepting a Newton step. Details are given below. The resulting method is summarized in Algorithm 2.

The algorithm contains two conditions (4.2.18) and (4.2.19) in step **S4**, and requires that one of the two has to hold if f is not convex. We now introduce these condition. Before doing so, however, we stress that these conditions are *only required* in the nonconvex case and only if we want to prove that *every accumulation* point of $(x^k)_k$ is stationary. If we are satisfied with the existence of at least one stationary accumulation point of $(x^k)_k$, then the conditions (4.2.18) or (4.2.19) are not required. We now state these two limited growth conditions for $\|F^{\Lambda_{k+1}}(x^k + s^k)\|$ and for $\psi(x^k + s^k)$:

$$\|F^{\Lambda_{k+1}}(x^{k+1})\| \leq \eta \rho_k \text{ and } \psi(x^{k+1}) \leq \psi(x^k) + \alpha_1 \sqrt{\|F^{\Lambda_k}(x^k)\| \|F^{\Lambda_{k+1}}(x^{k+1})\|}, \tag{4.2.18}$$

$$\psi(x^{k+1}) \leq \psi_k + \alpha_2 \|F^{\Lambda_{k+1}}(x^{k+1})\|^{2+\alpha_3}. \tag{4.2.19}$$

Let us further note that the latter growth conditions are chosen to control the ψ-descent of the Newton steps $x^k + s^k$ in a way such that global convergence of Algorithm 2 can be established without any additional restrictions. In the convex setting, if the conditions (4.2.18) and (4.2.19) are not used in Algorithm 2, we have to assume boundedness of the sequence of iterates $(x^k)_k$ and existence of an optimal solution of $(\mathcal{P})$ to cope with this missing controllability. For our subsequent analysis, we introduce the sets

$$\begin{aligned} \mathcal{K}_P &:= \{k : x^k \text{ was generated by the proximal gradient method}\}, \\ \mathcal{K}_N &:= \{k : x^k \text{ was generated by the Newton method}\}. \end{aligned}$$

Consequently, this means $k+1 \in \mathcal{K}_P$ if and only if $x^{k+1} = x^k + \sigma_k d^k$ was obtained in step **S5** and $k+1 \in \mathcal{K}_N$ if and only if $x^{k+1} = x^k + s^k$ was obtained in step **S3** before going to step **S5** .

It holds

$$\rho_k = \min_{i \in \mathcal{K}_N \cap \{1,\ldots,k\}} \|F^{\Lambda_i}(x^i)\|$$

and $\psi_k = \psi(x^{\ell_P(k)})$, where $\ell_P(k) := \max(\mathcal{K}_P \cap \{1, ..., k\})$ is the index of the last proximal gradient iteration. The trial point $x^k + s^k$ is accepted as new iterate if it is acceptable for the current filter $\mathcal{F}_k$ and additionally satisfies one of the above conditions (4.2.18) or (4.2.19).

Remark 4.2.19. The condition $x^{k+1} \in \operatorname{dom} \varphi$ in step **S3** of Algorithm 2 is needed to ensure well-definedness of the underlying proximal gradient method and of the growth conditions. In particular, it may happen that the Newton method generates an infeasible iterate x^{k+1} with $\psi(x^{k+1}) = \infty$ that is acceptable to the filter. Clearly, without the additional constraint $x^{k+1} \in \operatorname{dom} \varphi$, the quasi-Armijo condition in step **S4** then cannot be satisfied and consequently, our algorithmic mechanism cannot guarantee global convergence in this case. Let us note that if the domain dom φ is closed, then feasibility of the Newton iterates can be achieved via the following projection operation

$$x^{k+1} = \mathcal{P}_{\operatorname{dom} \varphi}(x^k + s^k).$$

Since the metric projection $\mathcal{P}_{\operatorname{dom} \varphi}$ is a nonexpansive operator, this will not affect the local convergence properties of the semismooth Newton method, see, e.g., [237]. Of course, this infeasibility problem arises from our specific globalization strategy. In chapter 6, we introduce a *D-gap function* for *generalized variational inequalities* that can be used as a real valued, smooth merit function for the optimality system (4.2.12) and for problem $(\mathcal{P})$. Recently, in [187], Patrinos et al. proposed another merit function-based globalization for the semismooth Newton method when f is strongly convex. These two different merit function approaches circumvent the unfavourable feasibility issues and can be used to design alternative base algorithms to substitute the proximal gradient method.

Remark 4.2.20. Clearly, the new parameter matrix Λ_{k+1} can also be calculated after step **S3** (or **S5**) of Algorithm 2. In this case, the acceptance test in step **S3** is performed w.r.t. to the old parameter matrix Λ_k, i.e., the filter values $\theta(x^{k+1}) \equiv \theta(x^{k+1}, \Lambda_k)$ then depend on Λ_k and the growth conditions have to be adjusted accordingly. While this allows a computation of the next parameter matrix Λ_{k+1} based on the new iterate x^{k+1}, this also requires the computation of $F^{\Lambda_k}(x^{k+1})$ and $F^{\Lambda_{k+1}}(x^{k+1})$. If the parameter matrices $(\Lambda_k)_k$ stay in a compact set $\mathcal{K} \subset \mathbb{S}^n_{++}$, i.e., if assumption (B) is satisfied, this adjustment will not affect our convergence theory.

4.3. Convergence analysis

4.3.1. Global convergence

This section focuses on the analysis of the convergence behaviour of our main approach.

Lemma 4.3.1. *Let the assumptions* (A.1) *and* (B) *hold and let* $(x^k)_k$ *and* $(\Lambda_k)_k$ *be generated by Algorithm 2. Consider a subsequence* $(x^k)_K$ *that converges to* x^* *and contains infinitely many iterates resulting from semismooth Newton steps, i.e.,* $|K \cap \mathcal{K}_N| = \infty$*. Then* x^* *is a stationary point.*

Proof. By assumption, there exists an infinite set $K \subset \mathcal{K}_N$ such that $(x^k)_K \to x^*$. Due to the structure of Algorithm 2 and by Remark 4.2.15, all assumptions of Lemma 4.2.14 are satisfied. Hence,

$$\lim_{K \ni k \to \infty} \theta(x^k) = \lim_{K \ni k \to \infty} \theta(x^k, \Lambda_k) = 0.$$

Thus, by (4.2.7) or Remark 4.2.12, we conclude $(F^{\Lambda_k}(x^k))_K \to 0$ and Remark 4.1.4 and the continuity of the proximity operator yield $F^{\Lambda}(x^*) = 0$, for some arbitrary $\Lambda \in \mathbb{S}^n_{++}$. □

Theorem 4.3.2. *Let the assumptions* (A.1)–(A.2) *and condition* (B) *be satisfied and let the sequences* $(x^k)_k$*,* $(\Lambda_k)_k$ *be generated by Algorithm 2. Furthermore, suppose that the sequence of iterates* $(x^k)_k$ *stays in a bounded set* $\Omega_0 \subseteq \Omega$*. Then, every accumulation point of* $(x^k)_k$ *is a stationary point of problem* $(\mathcal{P})$*.*

Proof. Let $x^* \in \mathbb{R}^n$ be any accumulation point of the sequence (x^k) and let $(x^k)_K$ be a corresponding subsequence that converges to x^*.

The strategy of the proof is based on a rigorous discussion of the occurrence and number of Newton and proximal gradient iterates in (x^k) and $(x^k)_K$. We start with two simpler cases that can be readily established by using Theorem 4.2.10 and Lemma 4.3.1. The remaining part of the proof is concerned with the case that $(x^k)_K$ contains only finitely many Newton iterations while infinitely many Newton iterates were generated. Here, we want to show that the possibly negative effect of the intermediate Newton steps is controlled by the filter conditions in a way that convergence of the subsequence $(x^k)_K$ can always be guaranteed. Consequently, we analyze several subcases that correspond to different filter (acceptance) constellations of the intermediate steps.

The proof of Theorem 4.3.2 is similar to the proof of [157, Theorem 4.2], where convergence of Algorithm 2 is shown for the specific choice $\varphi(x) \equiv \mu\|x\|_1$. In contrast to [157], the boundedness of the sequence of iterates $(x^k)_k$ is only needed in one specific sub-case in order to guarantee and control boundedness of the terms $\|F^{\Lambda_k}(x^k)\|$ for $k \in \mathcal{K}_P$. In Remark 4.3.11, we present two alternative strategies that allow to circumvent this additional boundedness assumption.

Case 1: $|K \cap \mathcal{K}_N| = \infty$. Then the claim follows directly from Lemma 4.3.1.

Case 2: $|K \cap \mathcal{K}_N| < \infty$, $|\mathcal{K}_N| < \infty$. In this case we just compute a finite number of Newton iterations, i.e., there exists $k_0 \in \mathbb{N}$ such that $k \in \mathcal{K}_P$ holds for all $k \geq k_0$. Hence, we can apply the convergence result of Theorem 4.2.10 for the quasi-Armijo proximal gradient method to complete the proof in this case.

The discussion of the remaining case heavily depends on the acceptance criteria in step **S3** of the algorithm and consequently we have to distinguish several more subcases.

Case 3: $|K \cap \mathcal{K}_N| < \infty$, $|\mathcal{K}_N| = \infty$. Since we perform infinitely many Newton steps, the sequence $(F^{\Lambda_k}(x^k))_{\mathcal{K}_N}$ converges to zero by Lemma 4.3.1. But since only finitely many such x^k are contained in the sequence $(x^k)_k$, the challenge is to show that the convergence of the proximal gradient method is not disrupted by the intermediate Newton steps, which might not always result in ψ-descent.

At first, we introduce several useful constants and derive preparatory estimates that will be needed in our subsequent investigation. Assumption (A.2) yields the Lipschitz constant L_f for ∇f on Ω. Thus, for all $\lambda_M I \succeq \Lambda \succeq \lambda_m I$, we obtain the Lipschitz constant $1 + L_f \lambda_m^{-1}$ for $u(x) = x - \Lambda^{-1}\nabla f(x)$ on Ω. Since the proximity operator $\operatorname{prox}_\varphi^\Lambda$ is Λ-nonexpansive, an easy computation yields that $F^\Lambda(x) = x - \operatorname{prox}_\varphi^\Lambda(u(x))$ is Lipschitz continuous on Ω with modulus

$$C_0 := 1 + (\lambda_m^{-1}\lambda_M)^{\frac{1}{2}} + L_f(\lambda_m^{-3}\lambda_M)^{\frac{1}{2}}.$$

Hence, due to Assumption (B) and Remark 4.1.4, there exists a constant $\overline{\lambda} = \overline{\lambda}(\lambda_m, \lambda_M)$ such that

$$\text{(4.3.1)}\quad \|F^{\Lambda_{k+1}}(x^{k+1})\| \leq \|F^{\Lambda_{k+1}}(x^k)\| + C_0 \cdot \sigma_k \|d^k\| \leq (\overline{\lambda} + C_0)\|F^{\Lambda_k}(x^k)\| =: C_1 \|F^{\Lambda_k}(x^k)\|$$

for all $k+1 \in \mathcal{K}_P$. Moreover, for all $k+1 \in \mathcal{K}_N$, using that x^{k+1} is acceptable to $\mathcal{F}_k$, we obtain from Remark 4.2.15:

$$\text{(4.3.2)}\quad \|F^{\Lambda_{k+1}}(x^{k+1})\| \leq \sqrt{n}\|F^{\Lambda_{k+1}}(x^{k+1})\|_\infty \leq \frac{\sqrt{n}}{c_\theta}\|\theta(x^{k+1})\|_\infty \leq \frac{\sqrt{n}}{\gamma_{\mathcal{F}} c_\theta}\|q\|_\infty, \quad \forall\, q \in \mathcal{F}_k.$$

Let $k_0 := \min \mathcal{K}_N$. Then, there holds $\theta(x^{k_0}) \in \mathcal{F}_k$ for all $k > k_0$, and thus

$$\text{(4.3.3)}\qquad \|F^{\Lambda_k}(x^k)\| \leq \frac{\sqrt{n}}{\gamma_{\mathcal{F}} c_\theta}\|\theta(x^{k_0})\|_\infty =: C_2, \quad \forall\, k \in \mathcal{K}_N,\ k > k_0.$$

Additionally, using the boundedness of $(x^k)_k$, Remark 4.1.4, and the Lipschitz continuity of the residual function F^Λ, there also exists $C_3 > 0$ such that

$$\text{(4.3.4)}\qquad \|F^{\Lambda_k}(x^k)\| \leq C_3, \quad \forall\, k \in \mathcal{K}_P, \quad k > k_0.$$

Further, let $(i_j)_{j\geq 0}$ enumerate all elements of the set $\{k \in \mathcal{K}_P : k > k_0\}$ in increasing order. Then the set $J := \{j : i_j \in K\}$ contains infinitely many indices. Defining

$$\Sigma(r) := \sum_{j=0}^{r-1}(\psi(x^{i_j}) - \psi(x^{i_{j+1}}))$$

we will use the telescope sum

$$\psi(x^{i_0}) - \psi(x^*) \geq \psi(x^{i_0}) - \liminf_{J \ni r \to \infty} \psi(x^{i_r}) = \liminf_{J \ni r \to \infty} \Sigma(r).$$

Our approach consists in deriving a lower bound for the right hand side that would exceed the left hand side as $J \ni r \to \infty$ unless $(F^{\Lambda_k}(x^k))_{K \cap \mathcal{K}_P} \to 0$.

Therefore, we will discuss the difference $\psi(x^{i_j}) - \psi(x^{i_{j+1}})$ of two consecutive proximal gradient iterates (with possibly other iterates in between). We define the index subsets

$$\mathcal{K}_N^a := \{k \in \mathcal{K}_N : x^k \text{ satisfies (4.2.18)}\}, \quad \mathcal{K}_N^b := \{k \in \mathcal{K}_N : x^k \text{ satisfies (4.2.19)}\}.$$

and the function

$$n_a : \mathbb{N} \to \mathbb{N}, \quad n_a(k) := |\mathcal{K}_N^a \cap \{k_0 + 1, ..., k\}|.$$

Note that k_0 is the index of the very first iterate x^{k_0} obtained by a Newton step. We then have

$$\|F^{\Lambda_k}(x^k)\| \leq \eta^{n_a(k)} \rho_{k_0}, \quad \forall\ k \in \mathcal{K}_N^a. \tag{4.3.5}$$

We now consider $j \geq 0$ and derive a lower bound for $\psi(x^{i_j}) - \psi(x^{i_{j+1}})$. To this end, we distinguish several cases. For abbreviation, let $\ell := i_j$, $k := i_{j+1} - 1$.

Sub-case 1: $k = \ell$ (i.e., $k, k+1 \in \mathcal{K}_P$):
Applying Lemma 4.2.7, Lemma 4.2.8, and inequality (4.3.1), we obtain

$$\psi(x^\ell) - \psi(x^{k+1}) \geq -\sigma_k \gamma \Delta_k \geq \sigma_k \gamma \lambda_m \|d^k\|^2 \geq \zeta \gamma \lambda_m \left\|F^{\Lambda_k}(x^k)\right\|^2 \geq \frac{\zeta \gamma \lambda_m}{C_1^2} \left\|F^{\Lambda_{k+1}}(x^{k+1})\right\|^2.$$

Denoting by J_1 the set of all $j \geq 0$ for which this case occurs (recall $\ell = i_j$), we see that $\Sigma_1(r) := \sum_{j \in J_1,\ j<r} (\psi(x^{i_j}) - \psi(x^{i_{j+1}}))$ is bounded below as $r \to \infty$. Furthermore, $\liminf_{J \ni r \to \infty} \Sigma_1(r) < \infty$ requires either $|J_1| < \infty$ or

$$\liminf_{J_1 \ni j \to \infty} F^{\Lambda_{i_{j+1}}}(x^{i_{j+1}}) = \lim_{J_1 \ni j \to \infty} F^{\Lambda_{i_{j+1}}}(x^{i_{j+1}}) = 0.$$

Sub-case 2: $\ell < k \in \mathcal{K}_N^b$:
Using the same arguments as in sub-case 1, (4.2.19), and $\psi_{k-1} = \psi(x^\ell)$, we obtain:

$$\begin{aligned} \psi(x^\ell) - \psi(x^{k+1}) &= \psi(x^\ell) - \psi(x^k) + \psi(x^k) - \psi(x^{k+1}) \\ &\geq \psi(x^\ell) - \psi_{k-1} - \alpha_2 \left\|F^{\Lambda_k}(x^k)\right\|^{(2+\alpha_3)} + \zeta \gamma \lambda_m \|F^{\Lambda_k}(x^k)\|^2 \\ &\geq \frac{1}{C_1^2} \left(\zeta \gamma \lambda_m - \alpha_2 \left\|F^{\Lambda_k}(x^k)\right\|^{\alpha_3} \right) \left\|F^{\Lambda_{k+1}}(x^{k+1})\right\|^2. \end{aligned}$$

Since $(F^{\Lambda_k}(x^k))_{\mathcal{K}_N} \to 0$, we see that there exists $k_1 \geq k_0$ such that $\alpha_2 \|F^{\Lambda_k}(x^k)\|^{\alpha_3} < \zeta \gamma \lambda_m / 2$ for all $k \in \mathcal{K}_N$ with $k \geq k_1$. Hence, if J_2 denotes the set of all $j \geq 0$ for which this sub-case 2 occurs, we see that $\Sigma_2(r) := \sum_{j \in J_2, j<r} (\psi(x^{i_j}) - \psi(x^{i_{j+1}}))$ is bounded below for $r \to \infty$ and that $\liminf_{J \ni r \to \infty} \Sigma_2(r) < \infty$ requires either $|J_2| < \infty$ or

$$\liminf_{J_2 \ni j \to \infty} F^{\Lambda_{i_{j+1}}}(x^{i_{j+1}}) = \lim_{J_2 \ni j \to \infty} F^{\Lambda_{i_{j+1}}}(x^{i_{j+1}}) = 0.$$

Sub-case 3: $\ell + 1, \dots, r_b - 1 \in \mathcal{K}_N$, $r_b = r_b(j) \in \mathcal{K}_N^b$, $r_b + 1, \dots, k \in \mathcal{K}_N^a$: Let $r_a = r_a(j)$ be defined as $r_a := \max(\{k_0\} \cup (\mathcal{K}_N^a \cap \{k_0 + 1, \dots, r_b - 1\}))$. We obtain

$$\begin{aligned}
\psi(x^\ell) - \psi(x^{k+1}) &= \psi(x^\ell) - \psi(x^{r_b}) + \sum_{i=r_b}^{k-1} (\psi(x^i) - \psi(x^{i+1})) + \psi(x^k) - \psi(x^{k+1}) \\
&\geq \psi(x^\ell) - \psi_{r_b - 1} - \alpha_2 \|F^{\Lambda_{r_b}}(x^{r_b})\|^{2+\alpha_3} \\
&\quad - \alpha_1 \sum_{i=r_b}^{k-1} \|F^{\Lambda_i}(x^i)\|^{\frac{1}{2}} \|F^{\Lambda_{i+1}}(x^{i+1})\|^{\frac{1}{2}} + \zeta\gamma\lambda_m \|F^{\Lambda_k}(x^k)\|^2 \\
&\geq -\alpha_2 \|F^{\Lambda_{r_b}}(x^{r_b})\|^{2+\alpha_3} - \alpha_1 \sqrt{\rho_{k_0} C_2} \sum_{i=r_b}^{k-1} \eta^{\frac{n_a(i+1)}{2}} + \frac{\zeta\gamma\lambda_m}{C_1^2} \|F^{\Lambda_{k+1}}(x^{k+1})\|^2,
\end{aligned}$$

where we used estimates as in the first case, inequalities (4.3.2), (4.3.3), (4.3.5) and the growth conditions (4.2.18), (4.2.19). Since the iterate x^{r_b} is acceptable for the filter, we can use inequality (4.3.2) with $q = \theta(x^{r_a})$. This yields

$$\|F^{\Lambda_{r_b}}(x^{r_b})\| \leq \frac{\sqrt{n}}{\gamma_{\mathcal{F}} c_\theta} \|\theta(x^{r_a})\|_\infty \leq \frac{\sqrt{n} C_\theta}{\gamma_{\mathcal{F}} c_\theta} \|F^{\Lambda_{r_a}}(x^{r_a})\| \leq \frac{\sqrt{n} C_\theta}{\gamma_{\mathcal{F}} c_\theta} \eta^{n_a(r_a)} \rho_{k_0},$$

and thus

$$\begin{aligned}
\psi(x^\ell) - \psi(x^{k+1}) \geq{}& -\alpha_2 \left(\frac{\sqrt{n} C_\theta}{\gamma_{\mathcal{F}} c_\theta} \eta^{n_a(r_a)} \rho_{k_0} \right)^{2+\alpha_3} - \alpha_1 \sqrt{\rho_{k_0} C_2} \sum_{i=r_b}^{k-1} \eta^{\frac{n_a(i+1)}{2}} \\
&+ \frac{\zeta\gamma\lambda_m}{C_1^2} \|F^{\Lambda_{k+1}}(x^{k+1})\|^2.
\end{aligned}$$

Let J_3 denote all $j \geq 0$ that fall into this sub-case 3. Then, we have $n_a(r_a(j)) \neq n_a(r_a(j'))$ for all $j, j' \in J_3$, $j \neq j'$. Hence

$$\sum_{j \in J_3} \eta^{n_a(r_a(j))(2+\alpha_3)} \leq \sum_{k=0}^{\infty} \eta^{k(2+\alpha_3)} = \frac{1}{1 - \eta^{2+\alpha_3}}.$$

Furthermore, $n_a(r_b(j) + 1) < n_a(r_b(j) + 2) < \dots < n_a(k) = n_a(i_{j+1} - 1)$, and thus,

$$\sum_{j \in J_3} \sum_{i=r_b(j)}^{i_{j+1}-2} \eta^{\frac{n_a(i+1)}{2}} \leq \sum_{k=0}^{\infty} \eta^{\frac{k}{2}} = \frac{1}{1 - \sqrt{\eta}}.$$

We thus see that $\Sigma_3(r) := \sum_{j \in J_3, j < r} (\psi(x^{i_j}) - \psi(x^{i_{j+1}}))$ is bounded below as $r \to \infty$ and that $\liminf_{J \ni r \to \infty} \Sigma_3(r) < \infty$ requires either $|J_3| < \infty$ or

$$\liminf_{J_3 \ni j \to \infty} F^{\Lambda_{i_{j+1}}}(x^{i_{j+1}}) = \lim_{J_3 \ni j \to \infty} F^{\Lambda_{i_{j+1}}}(x^{i_{j+1}}) = 0.$$

Sub-case 4: $\ell+1,\dots,k \in \mathcal{K}_N^a$.
This is the same situation as in sub-case 3, except that there does not exist an iterate $i \in \{\ell+1,\dots,k\}$ with $i \in \mathcal{K}_N^b$. Similarly as before, but easier, we obtain

$$\begin{aligned}\psi(x^\ell)-\psi(x^{k+1}) &= \sum_{i=\ell}^{k-1}(\psi(x^i)-\psi(x^{i+1}))+\psi(x^k)-\psi(x^{k+1})\\ &\geq -\alpha_1\sum_{i=\ell}^{k-1}\|F^{\Lambda_i}(x^i)\|^{\frac{1}{2}}\|F^{\Lambda_{i+1}}(x^{i+1})\|^{\frac{1}{2}}+\zeta\gamma\lambda_m\|F^{\Lambda_k}(x^k)\|^2\\ &\geq -\alpha_1\sqrt{\rho_{k_0}\max\{C_2,C_3\}}\sum_{i=\ell}^{k-1}\eta^{\frac{n_a(i+1)}{2}}+\frac{\zeta\gamma\lambda_m}{C_1^2}\|F^{\Lambda_{k+1}}(x^{k+1})\|^2,\end{aligned}$$

where we additionally used the estimate (4.3.4) and $\ell = i_j \in \mathcal{K}_P$. Let J_4 denote all $j \geq 0$ with $i_j \geq k_0$ that fall into this sub-case 4. Then, we have $n_a(i_j+1) < n_a(i_j+2) < \dots < n_a(i_{j+1}-1)$, and thus,

$$\sum_{j\in J_4}\sum_{i=i_j}^{i_{j+1}-2}\eta^{\frac{n_a(i+1)}{2}} \leq \sum_{k=0}^{\infty}\eta^{\frac{k}{2}} = \frac{1}{1-\sqrt{\eta}}.$$

This shows that $\Sigma_4(r) := \sum_{j\in J_4, j<r}(\psi(x^{i_j})-\psi(x^{i_{j+1}}))$ is bounded below as $r\to\infty$ and that $\liminf_{J\ni r\to\infty}\Sigma_4(r)<\infty$ requires either $|J_4|<\infty$ or

$$\liminf_{J_4\ni j\to\infty} F^{\Lambda_{i_{j+1}}}(x^{i_{j+1}}) = \lim_{J_4\ni j\to\infty} F^{\Lambda_{i_{j+1}}}(x^{i_{j+1}}) = 0.$$

Taking all cases together, it follows from

$$\psi(x^{i_0})-\psi(x^*) \geq \liminf_{J\ni r\to\infty}\sum_{c=1}^{4}\Sigma_c(r) \geq \sum_{c=1}^{4}\liminf_{J\ni r\to\infty}\Sigma_c(r)$$

that $(F^{\Lambda_{i_{j+1}}}(x^{i_{j+1}}))_{j\geq 0}\to 0$, since otherwise the limit on the right hand side would be $+\infty$ (note that all $\Sigma_c(r)$ were shown to be bounded below as $r\to\infty$). Using Remark 4.1.4 and since K contains infinitely many indices i_j, we conclude $F^\Lambda(x^*)=0$ for some arbitrary $\Lambda\in\mathbb{S}^n_{++}$. □

For convex problems, Algorithm 2 can be shown to converge globally without the growth conditions (4.2.18) and (4.2.19) if the iterates stay in a compact set. In particular, if the objective function ψ is coercive and φ is real valued and positively homogeneous, this boundedness condition is guaranteed by Lemma 4.2.6. Since every norm is a positively homogeneous function, the following theorem generalizes the convergence result in [157] for ℓ_1-regularized problems (i.e., in [157] the authors implicitly relied on the fact that the ℓ_1-norm is a positively homogeneous function).

Theorem 4.3.3. *Let the assumptions* (A.1), (B), *and* (C.1) *be satisfied and let the sequences* $(x^k)_k$ *and* $(\Lambda_k)_k$ *be generated by Algorithm 2. Furthermore, suppose that problem* ($\mathcal{P}$) *possesses an optimal solution* $\bar{x}\in\operatorname{dom}\varphi$*, the growth conditions* (4.2.18) *and* (4.2.19) *are not* *used in Algorithm 2 and that the sequence $(x^k)_k$ stays in a compact set $\Omega_0 \subset \Omega$. Then, every accumulation point of $(x^k)_k$ is a stationary point and thus, a globally optimal solution of the problem $(\mathcal{P})$.*

Proof. Let $x^* \in \mathbb{R}^n$ be an arbitrary accumulation point of the sequence $(x^k)_k$ and let $(x^k)_K$ be a subsequence converging to x^*. Clearly, as in Theorem 4.3.2, the two simple cases $|K \cap \mathcal{K}_N| = \infty$ and $|K \cap \mathcal{K}_N| < \infty$, $|\mathcal{K}_N| < \infty$ are already covered by Lemma 4.3.1 and Theorem 4.2.10. Thus, let us discuss the remaining, more difficult case.

So, let us assume $|K \cap \mathcal{K}_N| < \infty$, $|\mathcal{K}_N| = \infty$. Since we perform infinitely many Newton steps, the subsequence $(F^{\Lambda_k}(x^k))_{\mathcal{K}_N}$ converges to zero by Lemma 4.3.1. By assumption, the sequence $(x^k)_k$ remains in a compact set Ω_0 and there exists an optimal solution $\bar{x} \in \operatorname{dom} \varphi$ of problem $(\mathcal{P})$. We now want to show $(\psi(x^k))_{\mathcal{K}_N} \to \psi(\bar{x})$. Assume that $(\psi(x^k))_{\mathcal{K}_N}$ does not converge to $\psi(\bar{x})$. Since $\bar{x}$ is a global minimum of the objective function ψ, there then exist $\varepsilon > 0$ and a subsequence $L \subset \mathcal{K}_N$ with

$$\psi(x^\ell) \geq \psi(\bar{x}) + \varepsilon, \quad \forall\, \ell \in L.$$

From the bounded sequence $(x^\ell)_{\ell \in L}$ we can choose a further subsequence $\tilde{L} \subset L$ satisfying

$$(x^\ell)_{\ell \in \tilde{L}} \to \tilde{x} \quad \text{and} \quad \psi(\tilde{x}) = \liminf_{\tilde{L} \ni \ell \to \infty} \psi(x^\ell) \geq \psi(\bar{x}) + \varepsilon.$$

Assumption (B) and the continuity of the proximity operator yields $F^\Lambda(\tilde{x}) = 0$ for some $\Lambda \in \mathbb{S}^n_{++}$, and thus $\tilde{x}$ is a global solution, which results in the contradiction $\psi(\tilde{x}) = \psi(\bar{x})$. Hence, we have proved $(\psi(x^k))_{\mathcal{K}_N} \to \psi(\bar{x})$.

Next, using the feasibility of the iterates x^k and the descent property $\psi(x^k) < \psi(x^{k-1})$ for all $k \in \mathcal{K}_P$, (see Lemma 4.2.7), we obtain $\psi(x^*) = \psi(\bar{x})$ and thus the limit point x^* is an optimal solution, hence also a stationary point. □

Remark 4.3.4. As we have already mentioned and as we have shown in Lemma 4.2.6, the assumptions (C.2)–(C.3) are sufficient to ensure boundedness of the iterates x^k, $k \in \mathbb{N}$. Moreover, in this case, the coercivity of the objective function ψ also guarantees existence of an optimal solution $\bar{x} \in \operatorname{dom} \varphi$ of problem $(\mathcal{P})$. Accordingly, the boundedness of the effective domain $\operatorname{dom} \varphi$ yields the same implications. Another, alternative condition, which guarantees the boundedness of the sequence of iterates $(x^k)_k$, is the coercivity of the nonsmooth function $F^\Lambda : \mathbb{R}^n \to \mathbb{R}^n$ for some $\Lambda \in \mathbb{S}^n_{++}$. In fact, this condition implies that the Newton iterates stay in a compact set and by using the descent property of the proximal gradient steps, boundedness of the whole sequence $(x^k)_k$ can be established. Surprisingly, this rather restrictive condition is always satisfied when the function f is strongly convex on $\operatorname{dom} \varphi$. A proof of this claim is presented in Lemma 6.2.12 for an even more general setting.

Remark 4.3.5. The proofs of Theorem 4.3.2 and 4.3.3 do not use any particular properties of the semismooth Newton steps s^k, hence the semismooth Newton method for computing s^k could be replaced by other choices. In particular, the Newton system in step **S3** could be replaced by a regularized version of it, see, e.g., (4.3.17).

4.3.2. Fast local convergence

The semismooth Newton steps achieve locally q-superlinear convergence under suitable conditions. We now will prove that, under appropriate assumptions, Algorithm 2 turns into a semismooth Newton method after finitely many iterations and thus achieves locally an at least q-superlinear rate of convergence.

Assumption 4.3.6. *Let the sequences $(x^k)_k$ and $(\Lambda_k)_k$ be generated by Algorithm 2 and suppose that $x^* \in \mathbb{R}^n$ and $\Lambda_* \in \mathbb{S}^n_{++}$ are accumulation points of $(x^k)_k$ and $(\Lambda_k)_k$, respectively. Let us consider the following conditions:*

(D.1) *There exists $k^* \in \mathbb{N}$, such that $\Lambda_k = \Lambda_*$ for all $k \geq k^*$.*

(D.2) *It holds $x^* \in \operatorname{int} \operatorname{dom} \varphi$.*

(D.3) *The proximity operator $\operatorname{prox}_{\varphi}^{\Lambda_*} : \mathbb{R}^n \to \mathbb{R}^n$ is semismooth at $u^* := x^* - \Lambda_*^{-1} \nabla f(x^*)$.*

(D.4) *There exist constants $\delta > 0$ and $C > 0$ such that for all $x \in B_\delta(x^*)$, every matrix $M \in \partial F^{\Lambda_*}(x)$ is nonsingular with $\|M^{-1}\| \leq C$.*

If, in addition, the accumulation point x^* is a stationary point of $(\mathcal{P})$, then we assume:

(D.5) *The accumulation point x^* is a strict local minimum and an isolated stationary point of the problem $(\mathcal{P})$.*

Remark 4.3.7 (CD-regularity). Let us also mention another alternative invertibility assumption. The mapping F^{Λ_*} is called *CD-regular* at the accumulation point x^* if every element $M \in \partial F^{\Lambda_*}(x^*)$ is nonsingular. In this case, if F^{Λ_*} is CD-regular at x^*, then it can be shown that condition (D.4) has to be satisfied. This well-known fact follows from the upper semicontinuity and local boundedness of Clarke's subdifferential $\partial F^{\Lambda_*} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, see [199, Proposition 3.1] or [238, Proposition 2.12]. Hence, assumption (D.4) can be substituted by the stronger CD-regularity of F^{Λ_*}. Let us further note that if the limit point x^* is a stationary point of $(\mathcal{P})$ and if F^{Λ_*} is semismooth and CD-regular at x^*, then Pang and Qi [182, Proposition 3] showed that x^* is an isolated solution of the nonsmooth equation $F^{\Lambda_*}(x) = 0$ and thus, an isolated stationary point of the problem $(\mathcal{P})$.

Clearly, since we have already shown that every accumulation point of a sequence of iterates $(x^k)_k$ generated by Algorithm 2 is a stationary point of the initial problem $(\mathcal{P})$ under suitable conditions, assumption (D.5) is well-defined and applicable in our situation.

Remark 4.3.8. In the following, we briefly assess and discuss the different conditions in Assumption 4.3.6.

- The conditions (D.3)–(D.4) are standard assumptions for local convergence of semismooth Newton methods, [199, 182, 197, 238].
- While the choice of the parameter matrices Λ_k did not influence our global convergence theory, it certainly can affect the rate of convergence of the semismooth Newton method. In particular, to maintain local q-superlinear convergence, we have to assume that the matrices Λ_k do not change too wildly and are kept fixed after an appropriate

number of iterates. From a numerical perspective, this assumption seems not to be too restrictive since a fixed parameter matrix can be used whenever the algorithm reaches a certain level of tolerance.

- On the other hand, condition (D.2) clearly limits the applicability of our local convergence theory to convex composite problems that are locally Lipschitz continuous in a neighborhood of the predestined solution x^*. Specifically, this excludes optimization problems with additional convex constraints where the solution lies at the boundary of the feasible set. However, this condition is mandatory to guarantee feasibility of the Newton steps $x^k + s^k$ in a neighborhood of x^* and thus, to show transition to fast local convergence. Let us again note, that the requirement

  $$x^k + s^k \in \operatorname{dom} \varphi$$

  in step **S3** of Algorithm 2 is mainly a result of our specific globalization technique. For instance, if the proximal gradient method is substituted by another globally convergent algorithm that is not sensitive to the feasibility of the generated iterates, then condition (D.2) is not necessary. In particular, Patrinos et al. [187] proposed a *forward-backward envelope*-based and globally convergent approach for *strongly convex problems* that overcomes this feasibility issue and can accordingly be combined with the semismooth Newton method. Moreover, in chapter 6, we introduce a *D-gap function* for *generalized variational inequalities* that can be shown to act as a merit function for the optimality condition (4.2.12). Based on the D-gap function, we will then construct a simple descent method that is also well-defined for infeasible inputs and hence can be used to replace the proximal gradient method.

- Assumption (D.5) will be used to prove that the whole sequence $(x^k)_k$ converges to x^*.

- While most of the latter assumptions have a rather natural intuition, the conditions (D.4) and (D.5) seem to be more abstract and may not be easily verified in practise. In the subsequent chapter, we will discuss these conditions in some more detail. In particular, we will show that for a certain class of functions φ a second order sufficient optimality condition can be formulated that, together with the strict complementarity condition, will imply the assumptions (D.4) and (D.5).

We continue with a brief illustrative example.

Example 4.3.9 (Discussion of the ℓ_1-case). Before presenting the main convergence result, let us exemplarily examine the assumptions (D.1)–(D.5) for an ordinary ℓ_1-optimization problem of the form

$$\min_x \; f(x) + \varphi(x), \quad \varphi(x) = \mu\|x\|_1, \quad \mu > 0. \tag{4.3.6}$$

Here, since the ℓ_1-norm is real valued, condition (D.2) is immediately satisfied. Now, suppose that the parameter matrices are chosen via

$$\Lambda_k := \Lambda_* := \lambda^{-1} I$$

for some fixed $\lambda > 0$. Then, clearly, assumption (D.1) is fulfilled. Moreover, in this case and as in Example 4.2.16, the proximity operator $\mathrm{prox}_{\varphi}^{\lambda^{-1}I}$ is a piecewise continuously differentiable function and thus semismooth at all $x \in \mathbb{R}^n$. This readily establishes assumption (D.3). In [157], Milzarek and Ulbrich analyzed the convergence of Algorithm 2 for ℓ_1-problems of the type (4.3.6). In particular, let the sequence $(x^k)_k$ be generated by Algorithm 2 and let x^* be an accumulation of $(x^k)_k$. Then, setting $\mathcal{A}^* := \{i : x_i^* = 0\}$, it was shown in [157, Lemma 4.6 and 4.7] that the second-order type condition

$$h^\top \nabla^2 f(x^*) h > 0, \quad \forall\, h \in \mathbb{R}^n \text{ with } h_{\mathcal{A}^*} = 0, \tag{4.3.7}$$

guarantees uniformly bounded invertibility of all elements $M \in \partial F^{\Lambda_*}(x^*)$ and, additionally, the point x^* is also a strict local solution and an isolated stationary point of the ℓ_1-problem (4.3.6); see also [97, Section 3.4] and [156, Lemma 4.3.2] for related results. Consequently, in the ℓ_1-setting, the assumptions (D.1)–(D.5) are satisfied, whenever the second order condition (4.3.7) holds at x^*. As already mentioned, for more general problems, the verification of the assumptions (D.4)–(D.5) is more involved and will be investigated in the next chapter.

We now present our result on the local convergence of Algorithm 2.

Theorem 4.3.10. *Let the assumptions* (A.1)–(A.3) *hold and let the sequences $(x^k)_k$, $(\Lambda_k)_k$ be generated by Algorithm 2. Furthermore, let $x^* \in \mathbb{R}^n$ and $\Lambda_* \in \mathbb{S}^n_{++}$ be accumulation points of the sequences $(x^k)_k$ and $(\Lambda_k)_k$ satisfying the conditions* (D.1)–(D.5) *and suppose that the sequence of iterates $(x^k)_k$ remains in a bounded set $\Omega_0 \subseteq \Omega$. Then, it holds:*

(i) *The whole sequence $(x^k)_k$ converges to the isolated local minimum x^*.*

(ii) *There exists $\hat{k} > 0$ such that x^k results from a semismooth Newton step, i.e., $k \in \mathcal{K}_N$, for all $k \geq \hat{k}$. In particular, $(x^k)_k$ converges q-superlinearly to x^*.*

(iii) *If, in addition, the proximity operator $\mathrm{prox}_{\varphi}^{\Lambda_*}$ is α-order semismooth at u^* for some $\alpha \in (0, 1]$ and the Hessian $\nabla^2 f(x)$ is Lipschitz continuous near x^*, then the order of convergence is $1 + \alpha$.*

Proof. Let us start with the first part. Therefore, let $x^* \in \mathbb{R}^n$ be an accumulation point of the sequence $(x^k)_k$ with the stated properties. Then, by Theorem 4.3.2, x^* is a stationary point and assumption (D.5) yields that x^* is an isolated local minimum of problem ($\mathcal{P}$) and also an isolated stationary point. Since every accumulation point of $(x^k)_k$ is stationary, x^* is an isolated accumulation point of $(x^k)_k$.

Now, consider an arbitrary subsequence $(x^k)_K$ that converges to the isolated accumulation point x^*. If we can show $(\|x^{k+1} - x^k\|)_K \to 0$ then a well-known result of Moré and Sorensen [160, Lemma 4.10] implies the convergence of the whole sequence $(x^k)_k$ to x^*. Now, by assumption (D.4), the matrix $M(x) \in \partial F^{\Lambda_*}(x)$ is uniformly boundedly invertible in a neighborhood of x^*. Hence, by using condition (D.1), there exist $k_1 \geq k^*$ and $C > 0$ such that $\|M(x^k)^{-1}\| \leq C$ for all $k \in K$, $k \geq k_1$. This estimate immediately leads to

$$\|x^{k+1} - x^k\| \leq \max\{C, 1\} \|F^{\Lambda_k}(x^k)\| \to \max\{C, 1\} \|F^{\Lambda_*}(x^*)\| = 0, \quad (K \ni k \to \infty).$$

As intended, [160, Lemma 4.10] now yields that the entire sequence $(x^k)_k$ converges to x^*, which concludes the proof of part (i). At this point, let us also note that the stationarity of x^* already implies $x^* \in \mathrm{dom}\ \varphi \subset \Omega$.

Next, we prove the second statement. From part (i) we know that $x^k \to x^*$. As before, using assumptions (D.2), (D.4), and (D.5), there exist $\delta_1 > 0$, and $C > 0$ such that:

- $\|M(x)^{-1}\| \le C$ for all $M(x) \in \partial F^{\Lambda_*}(x)$ and $x \in \bar{B}_{\delta_1}(x^*)$.
- x^* is the unique stationary point of ψ on $\bar{B}_{\delta_1}(x^*)$.
- $\psi(x) > \psi(x^*)$ for all $x \in \bar{B}_{\delta_1}(x^*) \setminus \{x^*\}$.
- $x \in \mathrm{int\ dom}\ \varphi$ for all $x \in \bar{B}_{\delta_1}(x^*)$.

Due to assumption (A.2), the function F^{Λ_*} is Lipschitz continuous on $\bar{B}_{\delta_1}(x^*)$ with a Lipschitz constant $L_1 > 0$. (We refer to the proof of Theorem 4.3.2 for details). Since the gradient ∇f is bounded on $\bar{B}_{\delta_1}(x^*)$, Theorem 2.2.1 implies that ψ is also Lipschitz continuous on $\bar{B}_{\delta_1}(x^*)$ with a constant $L_2 > 0$. Moreover, invoking (A.3), Theorem 2.6.5, and condition (D.3), we can infer that F^{Λ_*} is semismooth at x^*.

For all $x \in \bar{B}_{\delta_1}(x^*)$, the Newton step $s = -M(x)^{-1}F^{\Lambda_*}(x)$ is well defined and due to the semismoothness of F^{Λ_*} and the bound on $\|M(x)^{-1}\|$, it holds for $x^+ = x + s$:

$$\begin{aligned} \|x^+ - x^*\| &= \|M(x)^{-1}[F^{\Lambda_*}(x^*) + M(x)(x - x^*) - F^{\Lambda_*}(x)]\| \\ &\le C\|F^{\Lambda_*}(x^*) + M(x)(x - x^*) - F^{\Lambda_*}(x)\| = o(\|x - x^*\|) \quad (\|x - x^*\| \to 0). \end{aligned} \tag{4.3.8}$$

Now let

$$\gamma_f := \min\left\{\eta, \frac{c_\theta}{\sqrt{n}C_\theta(1 + \gamma_{\mathcal{F}})}\right\}, \quad \gamma_s := \min\left\{\frac{\gamma_f}{L_1 C + \gamma_f}, \frac{\alpha_1^2}{(L_2 C + \alpha_1)^2}, \right\}.$$

Then, $0 < \gamma_f \le \eta < 1$ and $0 < \gamma_s < 1$ and there exists $0 < \delta \le \delta_1$ such that

$$\|x^+ - x^*\| \le \gamma_s \|x - x^*\|, \quad \forall\ x \in \bar{B}_\delta(x^*).$$

This shows for all $x \in \bar{B}_\delta(x^*)$:

$$\begin{aligned} \|x - x^*\| &\le \|x^+ - x^*\| + \|s\| \\ &= \|x^+ - x^*\| + \|M(x)^{-1}F^{\Lambda_*}(x)\| \le \gamma_s\|x - x^*\| + C\|F^{\Lambda_*}(x)\|. \end{aligned}$$

Hence, it holds

$$\|x - x^*\| \le \frac{C}{1 - \gamma_s}\|F^{\Lambda_*}(x)\|, \quad \forall\ x \in \bar{B}_\delta(x^*).$$

Furthermore, using $F^{\Lambda_*}(x^*) = 0$ and the definition of γ_s, we obtain

$$\begin{aligned} \|F^{\Lambda_*}(x^+)\| &\le L_1\|x^+ - x^*\| \le L_1\gamma_s\|x - x^*\| \\ &\le \frac{L_1 C\gamma_s}{1 - \gamma_s}\|F^{\Lambda_*}(x)\| \le \gamma_f\|F^{\Lambda_*}(x)\| \le \eta\|F^{\Lambda_*}(x)\|. \end{aligned}$$

Since $x^k \to x^*$, there exists $k_1 \geq k^*$ such that $x^k \in \bar{B}_\delta(x^*)$ for all $k \geq k_1$, and hence, with the semismooth Newton step $s^k = -M(x^k)^{-1}F^{\Lambda_k}(x^k) = -M(x^k)^{-1}F^{\Lambda_*}(x^k)$ and $x^{k,+} = x^k + s^k$, it holds:

$$\|x^{k,+} - x^*\| \leq \gamma_s \|x^k - x^*\|, \tag{4.3.9}$$

$$\|x^k - x^*\| \leq \frac{C}{1-\gamma_s}\|F^{\Lambda_*}(x^k)\|, \quad \|x^{k,+} - x^*\| \leq \frac{C}{1-\gamma_s}\|F^{\Lambda_*}(x^{k,+})\|, \tag{4.3.10}$$

$$\|F^{\Lambda_*}(x^k)\| = \|F^{\Lambda_*}(x^k) - F^{\Lambda_*}(x^*)\| \leq L_1\|x^k - x^*\|, \tag{4.3.11}$$

$$\|F^{\Lambda_*}(x^{k,+})\| \leq \gamma_f \|F^{\Lambda_*}(x^k)\|, \tag{4.3.12}$$

$$|\psi(x^{k,+}) - \psi(x^*)| \leq L_2\|x^{k,+} - x^*\|. \tag{4.3.13}$$

Let k be any index with $k \geq k_1 \geq k^*$ and

$$\|F^{\Lambda_*}(x^k)\| < \min_{0 \leq \ell < k} \|F^{\Lambda_\ell}(x^\ell)\|. \tag{4.3.14}$$

Since the algorithm does not terminate finitely and $0 < \|F^{\Lambda_k}(x^k)\| \to 0$, there exist infinitely many such indices k. We now show that if k_2 is the smallest such index, then $k \in \mathcal{K}_N$ for all $k \geq k_2 + 1$.

Let k satisfy (4.3.14). Then we have $x^{k,+} \in B_{\gamma_s\delta}(x^*) \subset B_\delta(x^*)$ and $\Lambda_{k+1} = \Lambda_*$. Further, for all $q \in \mathcal{F}_k$ and the corresponding $r \leq k$ with $q = \theta(x^r) \equiv \theta(x^r, \Lambda_r)$, it holds:

$$\begin{aligned}
&\max_j\,[q_j - \theta_j(x^{k,+})] - \gamma_{\mathcal{F}}\|\theta(x^{k,+})\|_\infty \\
&\quad \geq \|q\|_\infty - (1+\gamma_{\mathcal{F}})\|\theta(x^{k,+})\|_\infty = \|\theta(x^r)\|_\infty - (1+\gamma_{\mathcal{F}})\|\theta(x^{k,+}, \Lambda_*)\|_\infty \\
&\quad \geq c_\theta\|F^{\Lambda_r}(x^r)\|_\infty - C_\theta(1+\gamma_{\mathcal{F}})\|F^{\Lambda_*}(x^{k,+})\|_\infty \geq \frac{c_\theta}{\sqrt{n}}\|F^{\Lambda_r}(x^r)\| - C_\theta(1+\gamma_{\mathcal{F}})\|F^{\Lambda_*}(x^{k,+})\| \\
&\quad \geq \frac{c_\theta}{\sqrt{n}}\|F^{\Lambda_*}(x^k)\| - C_\theta(1+\gamma_{\mathcal{F}})\|F^{\Lambda_*}(x^{k,+})\| \geq \left(\frac{c_\theta}{\sqrt{n}} - C_\theta(1+\gamma_{\mathcal{F}})\gamma_f\right)\|F^{\Lambda_*}(x^k)\| \geq 0
\end{aligned}$$

by (4.3.12) and the definition of γ_f. Hence, $x^{k,+}$ is acceptable for $\mathcal{F}_k$. Also, there holds

$$\|F^{\Lambda_*}(x^{k,+})\| \leq \gamma_f\|F^{\Lambda_*}(x^k)\| \leq \eta\|F^{\Lambda_*}(x^k)\| = \eta \min_{0 \leq \ell \leq k}\|F^{\Lambda_\ell}(x^\ell)\| \leq \eta\rho_k < \rho_k. \tag{4.3.15}$$

Thus, the first condition in (4.2.18) is satisfied by $x^{k,+}$ (replacing x^{k+1}). Next, we show that also the second condition in (4.2.18) is satisfied. Since x^* is the unique local minimum on $\bar{B}_\delta(x^*) \supset \{x^k, x^{k,+}\}$, there holds $\psi(x^k) > \psi(x^*)$ and $\psi(x^{k,+}) > \psi(x^*)$. If we have $\psi(x^{k,+}) \leq \psi(x^k)$, the second condition of (4.2.18) is satisfied by $x^{k,+}$ replacing x^{k+1}. If we have $\psi(x^{k,+}) > \psi(x^k)$, the following holds

$$\begin{aligned}
|\psi(x^{k,+}) - \psi(x^k)| &\leq |\psi(x^{k,+}) - \psi(x^*)| \leq L_2\|x^{k,+} - x^*\| \\
&\leq L_2\sqrt{\gamma_s}\|x^k - x^*\|^{\frac{1}{2}}\|x^{k,+} - x^*\|^{\frac{1}{2}} \leq \frac{L_2 C\sqrt{\gamma_s}}{1-\gamma_s}\sqrt{\|F^{\Lambda_*}(x^k)\|\|F^{\Lambda_*}(x^{k,+})\|}
\end{aligned}$$

$$\leq \frac{L_2 C\sqrt{\gamma_s}}{1-\sqrt{\gamma_s}}\sqrt{\|F^{\Lambda_*}(x^k)\|\|F^{\Lambda_*}(x^{k,+})\|} \leq \alpha_1\sqrt{\|F^{\Lambda_*}(x^k)\|\|F^{\Lambda_*}(x^{k,+})\|},$$

where we used the Lipschitz continuity of ψ, the inequalities (4.3.9) and (4.3.10), and the definition of γ_s. Hence, we have shown that for all k satisfying (4.3.14), the semismooth Newton iterate satisfies all requirements such that it is chosen as new iterate. Thus, we have $x^{k+1} = x^{k,+} = x^k + s^k$ and $k \in \mathcal{K}_N$. Furthermore, (4.3.15) shows that

$$\|F^{\Lambda_*}(x^{k+1})\| < \min_{0\leq\ell\leq k} \|F^{\Lambda_\ell}(x^\ell)\|$$

and consequently x^{k+1} satifies again (4.3.14) with k replaced by $k+1$. Hence, inductively, we see $\{k : k > k_2\} \subset \mathcal{K}_N$ and thus we can choose $\hat{k} = k_2 + 1$. The superlinear convergence follows from (4.3.8).

We verify the third and last part. If the proximity operator $\mathrm{prox}_\varphi^{\Lambda_*}$ is α-order semismooth at u^* and $\nabla^2 f$ is Lipschitz continuous near x^*, then F^{Λ_*} is α-order semismooth at x^* and thus

$$\|F^{\Lambda_*}(x^*) + M(x)(x - x^*) - F^{\Lambda_*}(x)\| = O(\|x - x^*\|^{1+\alpha}), \quad \|x - x^*\| \to 0.$$

Hence, the asserted order of convergence follows from (4.3.8). □

Remark 4.3.11. Reconsidering the proofs of Theorem 4.3.2 and Theorem 4.3.10, we can see that boundedness of the sequence $(x^k)_k$ is only required in a special case of the proof of global convergence to guarantee boundedness of the sequence $(\|F^{\Lambda_k}(x^k)\|)_{k\in\mathcal{K}_P}$. Alternatively, the assumption

„ *The sequence $(\psi(x^k))_k$ remains in a compact set* “

does also ensure boundedness of the terms $\|F^{\Lambda_k}(x^k)\|$, $k \in \mathcal{K}_P$. This follows easily from the arguments used in sub-case 1 in the proof of Theorem 4.3.2. A second and more elegant variant can be achieved by slightly modifying Algorithm 2 and by replacing the first growth condition (4.2.18) by

$$\|F^{\Lambda_{k+1}}(x^{k+1})\| \leq \eta\rho_k \text{ and } \psi(x^{k+1}) \leq \psi(x^k) + \alpha_1\sqrt{\zeta_k \cdot \|F^{\Lambda_{k+1}}(x^{k+1})\|},$$

where $\zeta_k := \|F^{\Lambda_{\ell_N(k)}}(x^{\ell_N(k)})\|$ and $\ell_N(k) := \max(\{0\} \cup (\mathcal{K}_N \cap \{1, ..., k\}))$ denotes the index of the last accepted Newton iteration. Clearly, if $k \in \mathcal{K}_N$, then this alternative condition reduces to the old filter growth condition (4.2.18). It can be readily shown that this adjusted version of Algorithm 2 converges globally and locally without any additional boundedness assumptions.

In the convex setting and similar to our global convergence analysis, the growth conditions (4.2.18) and (4.2.19) are again not necessary to establish fast local convergence of Algorithm 2. Moreover, assumption (D.4) can be slightly weakened.

Theorem 4.3.12. *Let the assumptions* (A.1)–(A.3) *and* (C.1) *be satisfied and let the sequences $(x^k)_k$ and $(\Lambda_k)_k$ be generated by Algorithm 2. Suppose that problem $(\mathcal{P})$ possesses an optimal solution, the growth conditions* (4.2.18) *and* (4.2.19) *are not used in Algorithm*

2 and that the sequence $(x^k)_k$ remains in a compact set. Let $x^* \in \mathbb{R}^n$ and $\Lambda_* \in \mathbb{S}^n_{++}$ be accumulation points of the sequences $(x^k)_k$ and $(\Lambda_k)_k$ satisfying the conditions (D.1)–(D.3), *and* (D.5)*. Moreover, suppose that the following invertibility assumption is satisfied:*

- *There exist constants $k_* \in \mathbb{N}$ and $C > 0$ such that for all $k \geq k_*$ the generalized derivative $M_k := M(x^k) \in \partial F^{\Lambda_k}(x^k)$ is invertible with $\|M_k^{-1}\| \leq C$.*

Then, it holds:

(i) *The whole sequence $(x^k)_k$ converges to the isolated local minimum x^*.*

(ii) *The algorithm eventually turns into a pure semismooth Newton method and the sequence $(x^k)_k$ converges locally q-superlinearly to x^*.*

Proof. The first part of Theorem 4.3.12 can be established by using Theorem 4.3.3 and by mimicking the proof of Theorem 4.3.10 (i). Furthermore, the assumptions (A.2), (D.1)–(D.2), and the bounded invertibility of the Newton matrices $M(x^k)$ imply that there exist $k_0 := \max\{k^*, k_*\}$, $C > 0$, and $\delta > 0$ such that:

- $\|M_k^{-1}\| \leq C$ for all $k \geq k_0$.
- $x^k \in \bar{B}_\delta(x^*)$ for all $k \geq k_0$ and $\bar{B}_\delta(x^*) \subset \operatorname{int} \operatorname{dom} \varphi$.
- F^{Λ_*} is Lipschitz continuous on $\bar{B}_\delta(x^*)$.

Additionally and as in Theorem 4.3.10, it can be shown that F^{Λ_*} is semismooth at x^*. Thus, for all $k \geq k_0$, the Newton step $s^k = -M_k^{-1} F^{\Lambda_*}(x^k)$ is well-defined and there holds for $x^{k,+} = x^k + s^k$:

$$\begin{aligned} \|x^{k,+} - x^*\| &= \|M_k^{-1}[F^{\Lambda_*}(x^*) + M_k(x^k - x^*) - F^{\Lambda_*}(x^k)]\| \\ &\leq C\|F^{\Lambda_*}(x^*) + M_k(x^k - x^*) - F^{\Lambda_*}(x^k)\| = o(\|x^k - x^*\|), \quad k \to \infty. \end{aligned} \tag{4.3.16}$$

Clearly, at this point, we can reuse the arguments of the proof of Theorem 4.3.10 to show that the Newton step $x^{k,+}$ is acceptable for the current filter and thus, is chosen as new iterate for all $k \geq \hat{k}$ and sufficiently large $\hat{k} \geq k_0$. Let us note that, in this situation, we only have to verify the filter acceptance criterion. In particular, the second estimate in (4.3.10) is not needed and hence, we can work with weaker invertibility assumptions. The q-superlinear convergence then follows from (4.3.16). □

Finally, for robustifying the semismooth Newton steps, we are also interested in the more general case where, in the Newton system, the generalized derivative $M(x)$ is replaced by a regularized version M_ρ,

$$M_\rho(x) := M(x) + \rho(\|F^\Lambda(x)\|) \cdot \mathcal{R}(x), \quad M(x) \in \partial F^\Lambda(x), \quad \Lambda \in \mathbb{S}^n_{++}. \tag{4.3.17}$$

Here, the function $\rho : \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be continuous and monotonically increasing with $\rho(0) = 0$ and $\mathcal{R} : \mathbb{R}^n \to \mathbb{S}^n_+$ is a matrix-valued mapping that may depend on x and that is supposed to be (globally) uniformly bounded, i.e., there exists $C_\mathcal{R} \geq 0$ such that

$$\|\mathcal{R}(x)\| \leq C_\mathcal{R}, \quad \forall\, x \in \mathbb{R}^n.$$

In our numerical comparison in chapter 7, we will primarily work with regularizations of the form $\rho(t) = ct^p$ with $p \in (0,1]$, $c > 0$ and $\mathcal{R}(x) \equiv I$. Clearly, by Remark 4.3.5, we see that this adaption does not affect the global convergence of Algorithm 2. Moreover, by using a continuity argument and the well-known Banach perturbation lemma, the matrices $M_\rho(x)$ remain uniformly boundedly invertible in a certain neighborhood of x^* whenever assumption (D.4) holds. Furthermore, for the regularized Newton iterate $x_\rho^+ = x - M_\rho(x)^{-1}F^{\Lambda_*}(x)$ it follows

$$\begin{aligned}\|x_\rho^+ - x^*\| &\le \|M_\rho(x)^{-1}\|\|F^{\Lambda_*}(x^*) + M_\rho(x)(x - x^*) - F^{\Lambda_*}(x)]\| \\ &\le \|M_\rho(x)^{-1}\| \cdot \left[\|F^{\Lambda_*}(x^*) + M(x)(x - x^*) - F^{\Lambda_*}(x)\| + C_{\mathcal{R}}\rho(\|F^{\Lambda_*}(x)\|) \cdot \|x - x^*\|\right] \\ &= o(\|x - x^*\|), \quad \|x - x^*\| \to 0.\end{aligned}$$

Consequently, the q-superlinear rate of convergence, established in Theorem 4.3.10 and 4.3.12, is also not affected, if we use regularized derivatives of the form (4.3.17) in our algorithm.

5. Second order theory and decomposability

In the following sections, we present and discuss second order conditions for the problem

$$(\mathcal{P}_c) \qquad \min_{x\in\mathbb{R}^n} \; \psi_c(x) = f(x) + \phi(F(x)),$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $F : \mathbb{R}^n \to \mathbb{R}^m$ are supposed to be twice continuously differentiable and $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous function, as usual. Our overall goal is to replace the isolated stationarity and bounded invertibility assumptions (D.4) and (D.5), which were necessary to establish fast local convergence of the semismooth Newton method, by suitable second order conditions. On the other hand, as a second motivation, this methodology will also allow us to get a deeper insight on the local structure of stationary points and optimal solutions of $(\mathcal{P})$ and on their influence on the convergence behavior.

Clearly, problem $(\mathcal{P}_c)$ is a more general variant of problem $(\mathcal{P})$, since the composition $\phi \circ F$ need not be convex. However, if we set $F \equiv I$, then problem $(\mathcal{P}_c)$ obviously reduces to the minimization problem discussed in the last sections. Optimization problems of the form $(\mathcal{P}_c)$ are called *convex composite problems* and are an important and well-studied class of optimization problems. For instance, constrained nonlinear programs can be modeled via convex composite problems. Other applications comprise convex inclusions, minimax optimization problems, and exact penalization of general constrained problems, see, e.g., [28, 31], [81, Chapter 14], and [134, 206]. Moreover, as we have already seen, ϕ can also act as a nonsmooth, convex regularization term, like the ℓ_1-norm, the nuclear norm, or other structure-inducing regularizations.

During the last decades, various algorithms have been proposed to solve the convex composite problem $(\mathcal{P}_c)$. Of course, there is a huge amount of highly specialized methods, which have been developed for specific choices of ϕ. However, a large number of general algorithms can also be applied to the abstract class of problems $(\mathcal{P}_c)$ or to certain subclasses. These algorithms concentrate on general and proximal descent methods [28, 134], trust-region methods [263, 264], and Gauss-Newton methods [30, 135, 255] that work with a linearization of $F(x)$ to build a sequence of simpler subproblems of $(\mathcal{P}_c)$.

As we will see in the subsequent sections, in order to derive second order properties for our initial problem $(\mathcal{P})$, it is advantageous to consider the more general problem $(\mathcal{P}_c)$ first. Our analysis is primarily based on the second order theory that is presented in the monograph [27] of Bonnans and Shapiro for possibly infinite dimensional and constrained optimization problems of the form

$$(\mathcal{P}_K) \qquad \min_{x} \; f(x) \quad \text{s.t.} \quad F(x) \in K,$$

where K is a convex, closed set and the functions f and F are typically twice continuously differentiable. More specifically, based on the concepts of *(parabolic) second order tangent sets* [55], *second order regularity* [22, 23, 24] and related results in sensitivity and pertubation analysis of optimization problems [22, 23, 27], the authors Bonnans, Cominetti, Shapiro, and others have introduced, developed and collected profound, theoretical tools and frameworks that enable a detailed study of second order properties and conditions for problem $(\mathcal{P}_K)$ on a very abstract level. Furthermore, by setting $K \equiv \text{epi } \phi$ or $\phi \equiv \iota_K$, it is not hard to see that the convex composite problem $(\mathcal{P}_c)$ can also be interpreted as a general, constrained optimization problem $(\mathcal{P}_K)$ and vice versa. Using this straightforward connection, Bonnans and Shapiro transfered their theory and results developed for $(\mathcal{P}_K)$ to the convex composite setting, see, e.g., the sections 3.3.4 and 3.4.1 in [27], section 5 in [24] and also [217]. In the next sections we will sketch the second order theory presented in [27] and give a summary of the most important definitions, steps, and the different theoretical components. We will also add proofs whenever they facilitate the understanding of the overall concept and the underlying structure and ideas (or, of course, when a specific proof seems not to be available). Since we are interested in establishing stationarity and invertibility conditions for problem $(\mathcal{P})$, we will mainly focus on the "translated" second order results and terminology for convex composite problems.

Outline and motivation

In the following paragraph, we give a short roadmap of the main steps of this section and we make some introductory remarks.

Apparently, the task of stating second order conditions for our initial problem

$$(\mathcal{P}) \qquad \min_{x\in\mathbb{R}^n} \; f(x) + \varphi(x),$$

is notably complicated by the possible nonsmoothness and nonlinearity of the function φ. As we will see, in order to derive tight, *"no gap"* second order conditions, we will have to consider and incorporate certain terms that describe the curvature induced by the nonsmooth function φ. Here, the terminology "no gap" means that the only difference between sufficient and necessary second order conditions is between a strict and a non strict inequality. No gap second order conditions are desirable and often very beneficial, since they allow associating the sufficient second order condition with a natural quadratic growth condition.

In subsection 5.2, we will present a pair of no gap second order conditions that has been established by Bonnans and Shapiro in [27] for the general problems $(\mathcal{P}_c)$ and $(\mathcal{P}_K)$. Specifically, their analysis shows that the mentioned curvature term can be characterized by the convex conjugate of a second order directional epiderivative of φ. However, since it is hard to analyze stationarity and nonsingularity properties under these highly abstract conditions, we want to introduce a class of functions, that will allow us to use certain structural properties and that will simplify the second order conditions. In particular, following and inspired by the results in [217], we will assume that the function φ is *decomposable* and can be written as a composition of a convex, proper, lower semicontinuous and positively homogeneous function $\varphi_d : \mathbb{R}^m \to (-\infty, +\infty]$ and a twice continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^m$ in a certain neighborhood of a fixed point. Hence, we are also strongly interested in deriving second order conditions for optimization problems of the form

$$(\mathcal{P}_d) \qquad \min_{x \in \mathbb{R}^n} \; f(x) + \varphi_d(F(x)).$$

Clearly, this motivates a discussion of second order conditions of the more general convex composite problems $(\mathcal{P}_c)$.

In subsection 5.3, we will see that the class of decomposable functions is rather rich and includes ℓ_1-minimization and group sparse problems, and also semidefinite programs, low rank structured problems or even general nonlinear problems. Moreover, the complicated curvature term will have an easy representation and we can show that a suitable, corresponding second order sufficient condition implies isolated stationarity of a stationary point of $(\mathcal{P})$ and $(\mathcal{P}_d)$. Under the strict complementarity condition and using the so-called $\mathcal{VU}$-concept, [130, 103], we show that it is also possible to give a complete characterization of the second order conditions of $(\mathcal{P})$ in terms of the (generalized) derivative of $\text{prox}_\varphi^\Lambda$ at a certain point. This deep connection will then lead to new and general nonsingularity results, which are presented in subsection 5.4.

5.1. A first second order sufficient condition and isolated stationarity

We start with a discussion of the corresponding first order necessary conditions for $(\mathcal{P}_c)$. In the next sections, we will assume that the following properties are always satisfied:

- The functions $f : \mathbb{R}^n \to \mathbb{R}$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable.
- The mapping $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ is convex, proper, and lower semicontinuous.

5.1.1. First order necessary conditions

Let $\bar{x} \in F^{-1}(\text{dom } \phi)$ be a local minimum of problem $(\mathcal{P}_c)$, then the necessary first order optimality conditions for $(\mathcal{P}_c)$ take the following form

$$(5.1.1) \qquad (\psi_c)^{\downarrow}_{-}(\bar{x}; h) \geq 0, \quad \forall \; h \in \mathbb{R}^n.$$

Moreover, if *Robinson's constraint qualification*

$$(5.1.2) \qquad 0 \in \text{int}\{F(\bar{x}) + DF(\bar{x})\mathbb{R}^n - \text{dom } \phi\}$$

is satisfied at $\bar{x}$, then Lemma 2.5.6 implies that the composite function $\phi \circ F$ is directionally epidifferentiable and it holds

$$(\phi \circ F)^{\downarrow}(\bar{x}; h) = \phi^{\downarrow}(F(\bar{x}); DF(\bar{x})h).$$

Consequently, by Corollary 2.5.7, ψ_c is also directionally epidifferentiable at $\bar{x}$ and we obtain

$$\psi_c^{\downarrow}(\bar{x}; h) = \nabla f(\bar{x})^\top h + \phi^{\downarrow}(F(\bar{x}); DF(\bar{x})h).$$

Due to (5.1.1) and the convexity of ϕ, the functions $\Upsilon : \mathbb{R}^n \to [-\infty, +\infty]$, $\Upsilon(y) := \psi_c^{\downarrow}(\bar{x}; y)$ and $\Pi : \mathbb{R}^m \to [-\infty, +\infty]$, $\Pi(y) := \phi^{\downarrow}(F(\bar{x}); y)$ are convex, proper, lower semicontinuous, and positively homogeneous functions. Moreover, using Lemma 2.5.11 (i), it also follows that Υ and Π are subdifferentiable at 0. Now, let $y \in \operatorname{dom} \phi$ be arbitrary, then by applying (2.5.1), we have

$$\Pi(y - F(\bar{x})) \leq \phi(y) - \phi(F(\bar{x})) < +\infty,$$

which establishes $(\operatorname{dom} \phi) - F(\bar{x}) \subseteq \operatorname{dom} \Pi$. Hence, the regularity condition (i) in Lemma 2.5.15 is satisfied and we can infer

$$\partial\Upsilon(0) = \nabla f(\bar{x}) + DF(\bar{x})^\top \partial\Pi(0).$$

Thus, using $\partial\Pi(0) = \partial\phi(F(\bar{x}))$ (see, e.g., Example 2.5.17) and $\psi_c^{\downarrow}(\bar{x}; 0) = 0$, we see that the optimality condition (5.1.1) implies the condition

$$0 \in \partial\Upsilon(0) = \nabla f(\bar{x}) + DF(\bar{x})^\top \partial\phi(F(\bar{x})).$$

Clearly, the latter condition can be reformulated in the following way: there exists $\lambda \in \mathbb{R}^m$ such that

$$\nabla f(\bar{x}) + DF(\bar{x})^\top \lambda = 0, \quad \lambda \in \partial\phi(F(\bar{x})). \tag{5.1.3}$$

These results motivate the following definition and theorem.

Definition 5.1.1. *Let $\bar{x} \in F^{-1}(\operatorname{dom} \phi)$ be given. A point $\lambda \in \mathbb{R}^m$ that (together with $\bar{x}$) satisfies the conditions* (5.1.3) *is called* Lagrange multiplier. *The set of all possible Lagrange multipliers is denoted by $\mathcal{M}(\bar{x})$, i.e.,*

$$\mathcal{M}(\bar{x}) := \{\lambda \in \mathbb{R}^m : \nabla f(\bar{x}) + DF(\bar{x})^\top \lambda = 0, \, \lambda \in \partial\phi(F(\bar{x}))\}. \tag{5.1.4}$$

The point $\bar{x}$ is called a stationary point *of problem $(\mathcal{P}_c)$ if and only if $\mathcal{M}(\bar{x}) \neq \emptyset$.*

Theorem 5.1.2 (First order necessary conditions). *Let $\bar{x} \in F^{-1}(\operatorname{dom} \varphi)$ be a local minimum of problem $(\mathcal{P}_c)$ and suppose that Robinson's constraint qualification holds at $\bar{x}$. Then, there exists $\lambda \in \mathbb{R}^m$ such that*

$$\nabla f(\bar{x}) + DF(\bar{x})^\top \lambda = 0, \quad \lambda \in \partial\phi(F(\bar{x})).$$

In particular, $\bar{x}$ is a stationary point of $(\mathcal{P}_c)$.

Remark 5.1.3. Suppose that $\bar{x}$ is stationary point of $(\mathcal{P}_c)$. Then, our preceding discussion shows, that the function Π is subdifferentiable at 0 and, due to Lemma 2.5.15, we have

$$0 \in \nabla f(\bar{x}) + DF(\bar{x})^\top \partial\Pi(0) \subseteq \nabla f(\bar{x}) + \partial(\Pi \circ DF(\bar{x}))(0).$$

Consequently, it holds

$$0 \leq \nabla f(\bar{x})^\top (h - 0) + \Pi(DF(\bar{x})h) - \Pi(0) = \nabla f(\bar{x})^\top h + \phi^{\downarrow}(F(\bar{x}); DF(\bar{x})h), \quad \forall \, h \in \mathbb{R}^n.$$

Hence, in this case, Remark 2.5.8 implies that our initial, first order condition

$$(\psi_c)^{\downarrow}_{-}(\bar{x}; h) \geq 0, \quad \forall\, h \in \mathbb{R}^n,$$

is also fulfilled (even if Robinson's constraint qualification does not hold at $\bar{x}$). In summary, we see that the first order necessary conditions (5.1.3) are generally stronger than condition (5.1.1). We like to mention that this natural "gap" also appears when discussing and deriving KKT conditions in nonlinear programming.

In the subsequent analysis, we will also need the next, quite standard property.

Lemma 5.1.4 (cf. [27, Proposition 4.43]). *Assume that $\bar{x} \in F^{-1}(\text{dom } \phi)$ is a stationary point of $(\mathcal{P}_c)$ and that Robinson's constraint qualification is satisfied at $\bar{x}$. Then, $\mathcal{M}(\bar{x})$ is a nonempty, convex, and compact set and, additionally, the sets $\mathcal{M}(x)$ are uniformly bounded for all x in a neighborhood of $\bar{x}$.*

5.1.2. Second order conditions and the strict constraint qualification

First order sufficient conditions that guarantee optimality of a stationary point often fail to hold in practice. This observation can be traced back to the fact, that the directional epiderivative $(\psi_c)^{\downarrow}_{-}(\bar{x}; \cdot)$ does not provide any information about optimality of a stationary point $\bar{x}$ along directions $h \in \mathbb{R}^n$ that satisfy the condition

$$(\psi_c)^{\downarrow}_{-}(\bar{x}; h) \leq 0. \tag{5.1.5}$$

A very common approach to circumvent this lack of information is to pose certain, appropriate second order conditions and to incorporate second order information. In the following, we introduce the so-called *critical cone* of ψ_c that exactly consists of those directions h fulfilling (5.1.5).

Definition 5.1.5. *Let $\bar{x} \in F^{-1}(\text{dom } \phi)$ be given. The* critical cone *of ψ_c at $\bar{x}$ is defined by*

$$\mathcal{C}(\bar{x}) := \{h \in \mathbb{R}^n : (\psi_c)^{\downarrow}_{-}(\bar{x}; h) \leq 0\}. \tag{5.1.6}$$

Additionally, if $\bar{x}$ is a stationary point of problem $(\mathcal{P}_c)$ and if Robinson's constraint qualification (5.1.2) *is satisfied at $\bar{x}$, then the critical cone $\mathcal{C}(\bar{x})$ can be equivalently represented as follows:*

$$\mathcal{C}(\bar{x}) = \{h \in \mathbb{R}^n : (\psi_c)^{\downarrow}_{-}(\bar{x}; h) = 0\} = \{h \in \mathbb{R}^n : DF(\bar{x})h \in N_{\partial\phi(F(\bar{x}))}(\lambda)\}, \tag{5.1.7}$$

where $\lambda \in \mathcal{M}(\bar{x})$ is an arbitrary Lagrange multiplier.

Proof. Let us briefly verify the alternative representation of the critical cone in the case that $\bar{x}$ is a stationary point of $(\mathcal{P}_c)$ and Robinson's constraint qualification holds at $\bar{x}$. The formula

$$\mathcal{C}(\bar{x}) = \{h \in \mathbb{R}^n : (\psi_c)^{\downarrow}_{-}(\bar{x}; h) = 0\}$$

follows directly from definition (5.1.6) and from Remark 5.1.3. (Let us note that Robinson's constraint qualification need not necessarily hold at $\bar{x}$ for this implication). Now, let $h \in \mathcal{C}(\bar{x})$

be arbitrary. Then, due to Remark 2.5.8, it holds

$$\begin{aligned} 0 &\geq \nabla f(\bar{x})^\top h + \phi^\downarrow(F(\bar{x}); DF(\bar{x})h) \\ &= -(DF(\bar{x})^\top \lambda)^\top h + \sigma_{\partial\phi(F(\bar{x}))}(DF(\bar{x})h) = \sup_{v \in \partial\phi(F(\bar{x}))} \langle v - \lambda, DF(\bar{x})h \rangle. \end{aligned}$$

Hence, it follows $\langle v - \lambda, DF(\bar{x})h \rangle \leq 0$ for all $v \in \partial\phi(F(\bar{x}))$ and $DF(\bar{x})h \in N_{\partial\phi(F(\bar{x}))}(\lambda)$. On the other hand, if $h \in \mathbb{R}^n$ satisfies $DF(\bar{x})h \in N_{\partial\phi(F(\bar{x}))}(\lambda)$, then the above discussion immediately implies

$$\nabla f(\bar{x})^\top h + \phi^\downarrow(F(\bar{x}); DF(\bar{x})h) \leq 0.$$

Using Corollary 2.5.7, this establishes $h \in \mathcal{C}(\bar{x})$. □

Now, we are able to present a first second order condition that was introduced and studied by Burke et al. in [29, 31].

Theorem 5.1.6 (Second order sufficient conditions). *Let $\bar{x}$ be a stationary point of problem $(\mathcal{P}_c)$ and suppose that the second order sufficient condition*

$$\sup_{\lambda \in \mathcal{M}(\bar{x})} \left\{ h^\top \nabla^2 f(\bar{x}) h + \langle \lambda, D^2F(\bar{x})[h,h] \rangle \right\} > 0, \quad \forall\ h \in \mathcal{C}(\bar{x}) \setminus \{0\} \tag{5.1.8}$$

is satisfied. Then, there exists $\alpha > 0$ such that for all x in a neighborhood of $\bar{x}$ it holds

$$f(x) + \phi(F(x)) \geq f(\bar{x}) + \phi(F(\bar{x})) + \alpha \|x - \bar{x}\|^2,$$

and hence, $\bar{x}$ is a (strict) locally optimal solution of $(\mathcal{P}_c)$.

Proof. The proof is exactly as in [31, Theorem 4.2]. □

In the case of cone constrained optimization, i.e., $\varphi \equiv \iota_K$, where $K \subseteq \mathbb{R}^m$ is a convex, closed cone, it was already observed by Robinson [203] that condition (5.1.8) does not ensure that $\bar{x}$ is an *isolated*, local minimum of the optimization problem $(\mathcal{P}_c)$, see, e.g., Example (2.5) in [203]. In the following we will discuss several conditions that together with the second order sufficient conditions (5.1.8) will guarantee that a stationary point is an isolated stationary point of problem $(\mathcal{P}_c)$.

The next result combines Lemma 4.44 and Proposition 4.47 in [27] and states a corresponding "translated" version for the convex composite setting. For various related formulations and examples we like to refer to [216].

Lemma 5.1.7 (Strict constraint qualification). *Let $\bar{x}$ be a stationary point of problem $(\mathcal{P}_c)$ and let $\bar{\lambda} \in \mathcal{M}(\bar{x})$ be a corresponding Lagrange multiplier. Suppose that $\bar{\lambda}$ satisfies the following* strict constraint qualification

$$0 \in \mathrm{int}\{F(\bar{x}) + DF(\bar{x})\mathbb{R}^n - \bar{\Phi}\}, \tag{5.1.9}$$

where $\bar{\Phi} := \{y \in \mathbb{R}^m : \langle \bar{\lambda}, y - F(\bar{x}) \rangle = \phi(y) - \phi(F(\bar{x}))\} \subseteq \mathrm{dom}\ \phi$. Then:

(i) *The Lagrange multiplier $\bar{\lambda}$ is unique, i.e., $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$.*

(ii) *The multifunction $\mathcal{M} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is upper Lipschitzian at the stationary point $\bar{x}$.*

Proof. The proof of this Lemma is based on the (rather easy) observation that the set $\bar{\Phi}$ and the condition (5.1.9) yield the right extension of the strict constraint qualification given in Definition 4.46 in [27]. Alternatively, the statements (i) and (ii) can be shown directly by mimicking the proof of Proposition 4.47 in [27] and by appropriately using the structure of the set $\bar{\Phi}$. We will not go into details here. □

5.1.3. Constraint nondegeneracy and the strict complementarity condition

In the following, we present the concepts of *constraint nondegeneracy* and *strict complementarity*, which will play an essential role in our subsequent analysis. Constraint nondegeneracy was originally introduced by Robinson [204, 205] to study sensitivity properties of nonlinear programs. Robinson also showed that in the case of nonlinear programming, the nondegeneracy condition reduces to the well-known *Linear Independence Constraint Qualification* (LICQ). In [27], constraint nondegeneracy was used in a reduction approach for cone-reducible problems to establish quantitative stability results (see section 4.6 in [27]). A related version of the nondegeneracy condition was also discussed in the paper [26], where it was connected to transversality. Here, we will work with the more general formulation that was studied by Shapiro in [218].

Over the last years, the concept of constraint nondegeneracy has been used in different fields of optimization to investigate and discuss second order conditions and second order information. For instance, in [226, 43, 119], based on the nondegeneracy condition, a second order theory was developed for semidefinite programs and low-rank structured problems that can be applied to establish fast local convergence of corresponding semismooth Newton-type methods. Further examples and applications can also be found in [26, 27, 217] and [25].

Definition 5.1.8 (Nondegeneracy and strict complementarity). *Let $x \in F^{-1}(\text{dom}\ \phi)$ be a feasible point of problem $(\mathcal{P}_c)$. We say that the* nondegeneracy condition *is fulfilled at x, if*

$$DF(x)\mathbb{R}^n + \text{lin}\ N_{\partial\phi(F(x))}(\lambda) = \mathbb{R}^m, \tag{5.1.10}$$

where $\lambda \in \partial\phi(F(x))$ is an arbitrary subgradient. Moreover, we say that the strict complementarity condition *is satisfied at x, if there exists $\lambda \in \mathcal{M}(x)$ such that*

$$\lambda \in \text{ri}\ \partial\phi(F(x)). \tag{5.1.11}$$

Remark 5.1.9. Let us note that the nondegeneracy condition can only hold at some $x \in F^{-1}(\text{dom}\ \phi)$ if ϕ is subdifferentiable at $F(x)$. Otherwise, the lineality space lin $N_{\partial\phi(F(x))}(\lambda)$ is empty and condition (5.1.10) cannot be satisfied by definition.

In the next paragraph, we want to derive several equivalent representations of the nondegeneracy condition (5.1.10). Therefore, let $x \in F^{-1}(\text{dom}\ \varphi)$ be a feasible point and assume that ϕ is subdifferentiable at $F(x)$. Moreover, let $\lambda \in \partial\phi(F(x))$ be an arbitrary subgradient and let us define the following sets

$$\mathcal{U}_1 := \text{lin } N_{\partial\phi(F(x))}(\lambda), \quad \mathcal{U}_2 := \text{lin } \phi^{\downarrow}(F(x);\cdot), \quad \mathcal{U}_3 := [\text{aff } \partial\phi(F(x)) - \lambda]^{\perp}.$$

Then, it holds $\mathcal{U}_1 = \mathcal{U}_2 = \mathcal{U}_3$. Clearly, this shows that the subspaces $\mathcal{U}_1$ and $\mathcal{U}_3$ do not depend on the specific choice of the subgradient λ. In particular, for all $\lambda_1, \lambda_2 \in \partial\phi(F(x))$ we have

$$\text{lin } N_{\partial\phi(F(x))}(\lambda_1) = \text{lin } N_{\partial\phi(F(x))}(\lambda_2). \tag{5.1.12}$$

Let us briefly prove the equivalence of the three different subspaces $\mathcal{U}_1$–$\mathcal{U}_3$. From the definition of the normal cone, it follows for any $h \in \text{lin } N_{\partial\phi(F(x))}(\lambda)$:

$$\langle v - \lambda, h\rangle = 0, \quad \forall\ v \in \partial\phi(F(x)).$$

Obviously, using Lemma 2.5.11 (ii), this implies

$$\langle \lambda, h\rangle = \phi^{\downarrow}(F(x);h). \tag{5.1.13}$$

Furthermore, since the lineality space of the normal cone $N_{\partial\phi(F(x))}(\lambda)$ is a linear subspace, we also have $-h \in \text{lin } N_{\partial\phi(F(x))}(\lambda)$. Consequently, equation (5.1.13) also holds for $-h$ and we immediately obtain

$$\phi^{\downarrow}(F(x);h) + \phi^{\downarrow}(F(x);-h) = 0.$$

On the other hand, if $h \in \text{lin } \phi^{\downarrow}(F(x);\cdot)$ is given, then again by applying Lemma 2.5.11 (ii), we can infer

$$\langle \lambda - v, h\rangle \le 0 \quad \text{and} \quad \langle v - \lambda, h\rangle \le 0$$

for all $v \in \partial\phi(F(x))$. This shows $h \in \text{lin } N_{\partial\phi(F(x))}(\lambda)$ and establishes $\mathcal{U}_1 = \mathcal{U}_2$. To finish the proof, let $h \in \mathcal{U}_2^{\perp}$ be arbitrary. Then, using (5.1.13), it holds

$$\langle h + \lambda, d\rangle = \phi^{\downarrow}(F(x);d), \quad \forall\ d \in \text{lin } \phi^{\downarrow}(F(x);\cdot).$$

However, by Lemma 2.5.11 (ii) and Lemma 2.1.9, this is equivalent to $h + \lambda \in \text{aff } \partial\phi(F(\bar{x}))$. Since $\mathcal{U}_2$ is a closed, linear subspace, it follows $\mathcal{U}_2 = [\mathcal{U}_2^{\perp}]^{\perp} = \mathcal{U}_3$, as desired.

Next, we want to present and verify a connection between the nondegeneracy condition (5.1.10) and Robinson's constraint qualification. Let us set

$$\Xi := \{(h,t) \in \mathbb{R}^m \times \mathbb{R} : h \in \text{lin } \phi^{\downarrow}(F(x);\cdot),\, t = \phi^{\downarrow}(F(x);h)\}.$$

and let $(h,t) \in \Xi$ be arbitrary. From the definition of the set Ξ, it directly follows $(h,t) \in \text{epi } \phi^{\downarrow}(F(x);\cdot)$ and we have

$$\phi^{\downarrow}(F(x);-h) = -\phi^{\downarrow}(F(x);h) = -t.$$

Together, these observations imply

$$(h,t) \in \text{epi } \phi^{\downarrow}(F(x);\cdot) \cap -[\text{epi } \phi^{\downarrow}(F(x);\cdot)] = \text{lin epi } \phi^{\downarrow}(F(x);\cdot).$$

Now, suppose that $(h,t) \in \text{lin epi } \phi^{\downarrow}(F(x);\cdot)$ is a given vector, then it holds

$$\phi^{\downarrow}(F(x);h) + \phi^{\downarrow}(F(x);-h) \leq t + (-t) = 0.$$

Using the subadditivity of the directional epiderivative, we readily establish $h \in \text{lin } \phi^{\downarrow}(F(x);\cdot)$ and

$$t \geq \phi^{\downarrow}(F(x);h) = -\phi^{\downarrow}(F(x);-h) \geq -(-t) = t.$$

Hence, we have shown that the two sets Ξ and $\text{lin epi } \phi^{\downarrow}(F(x);\cdot)$ coincide. Furthermore, due to Lemma 2.5.3, we obtain the following, useful connection to the tangent cone of the epigraph epi ϕ:

$$\Xi = \text{lin epi } \phi^{\downarrow}(F(x);\cdot) = \text{lin } T_{\text{epi } \phi}(F(x),\phi(F(x))).$$

As a consequence, the nondegeneracy condition (5.1.10) can be equivalently rewritten as

$$\begin{pmatrix} DF(x)\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - \text{lin } T_{\text{epi } \phi}(F(x),\phi(F(x))) = \begin{pmatrix} \mathbb{R}^m \\ \mathbb{R} \end{pmatrix}$$

and, due to $\text{lin epi } \phi^{\downarrow}(F(x);\cdot) \subset \text{epi } \phi^{\downarrow}(F(x);\cdot)$, the following condition must be automatically satisfied at x

$$\begin{pmatrix} DF(x)\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - T_{\text{epi } \phi}(F(x),\phi(F(x))) = \begin{pmatrix} \mathbb{R}^m \\ \mathbb{R} \end{pmatrix}.$$

However, by applying Proposition 2.97 and Corollary 2.98 of [27], the latter condition is equivalent to

$$0 \in \text{int} \left\{ \begin{pmatrix} F(x) \\ \varphi(F(x)) \end{pmatrix} + \begin{pmatrix} DF(x)\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - \text{epi } \varphi \right\},$$

which in fact is just another equivalent reformulation of Robinson's constraint qualification (5.1.2). In summary, our latter computations have shown that the nondegeneracy condition implies Robinson's constraint qualification. In particular, if the nondegeneracy condition holds at a local solution $\bar{x}$ of $(\mathcal{P}_c)$, then $\mathcal{M}(\bar{x})$ is nonempty and $\bar{x}$ is a stationary point of problem $(\mathcal{P}_c)$.

We conclude this discussion with a brief example. Let $K \subset \mathbb{R}^n$ be a convex, closed, and nonempty set and let us define $\phi := \iota_K$. Furthermore, let $x \in F^{-1}(K)$ and $\lambda \in \partial\phi(F(x)) = N_K(F(x))$ be arbitrary. Then, the nondegeneracy condition can be simplified as follows. By (2.1.2), we have

$$N_{\partial\phi(F(x))}(\lambda) = N_{N_K(F(x))}(\lambda) = \{y \in T_K(F(x)) : \langle \lambda, y \rangle = 0\} = T_K(F(x)) \cap \{\lambda\}^{\perp}.$$

Moreover, due to $0 \in N_K(F(x))$ and (5.1.12), we obtain

$$\text{lin } N_{\partial\phi(F(x))}(\lambda) = \text{lin } N_{\partial\phi(F(x))}(0) = \text{lin } T_K(F(x)).$$

Thus, in this situation, the nondegeneracy condition reduces to

$$DF(x)\mathbb{R}^n - \text{lin } T_K(F(x)) = \mathbb{R}^m.$$

Let us note that this is exactly the condition that was introduced and analyzed by Bonnans and Shapiro in [26, 27, 218].

The following result of Lemaréchal and Sagastizábal provides a helpful, alternative characterization of the strict complementarity condition.

Lemma 5.1.10. *Let $S \subset \mathbb{R}^n$ be a convex, nonempty, and closed set and let $\lambda \in S$ be arbitrary. Then, it holds*

$$\lambda \in \text{ri } S \quad \Longleftrightarrow \quad N_S(\lambda) \textit{ is a subspace.}$$

Proof. A proof of this result can be found in [131, Proposition 2.2]. □

The next result is analogue to [27, Proposition 4.75] and [218, Theorem 2.1] and completes our discussion of the nondegeneracy and strict complementarity condition.

Lemma 5.1.11. *Suppose that $\bar{x}$ is a local solution or a stationary point of $(\mathcal{P}_c)$. Then, the following holds:*

(i) *If the nondegeneracy condition holds at $\bar{x}$, then the set $\mathcal{M}(\bar{x})$ reduces to a singleton.*

(ii) *Conversely, suppose that $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$ and that the strict complementarity conditions is satisfied at $\bar{x}$. Then, the nondegeneracy condition holds at $\bar{x}$.*

Proof. The proof is a mere "translation" of the proofs given in [27, 218]. Nonetheless, for the sake of completeness, we want to present a proof of this statement.

At first, let $\bar{x}$ be a local solution of $(\mathcal{P}_c)$ and suppose that the nondegeneracy condition is satisfied. Then, we have seen that Robinson's constraint qualification must hold at $\bar{x}$ and, consequently, $\bar{x}$ is also a stationary point of $(\mathcal{P}_c)$. In particular, we have $\mathcal{M}(\bar{x}) \neq \emptyset$. Of course, the argumentation is identical if $\bar{x}$ is already a stationary point. Now, let us assume that there exists $\bar{\lambda}, \lambda \in \mathcal{M}(\bar{x})$, $\bar{\lambda} \neq \lambda$. It follows

$$DF(\bar{x})^\top(\lambda - \bar{\lambda}) = 0, \quad \lambda - \bar{\lambda} \in \partial\phi(F(\bar{x})) - \bar{\lambda} \subset \text{aff } \partial\phi(F(\bar{x})) - \bar{\lambda}$$

and we infer $\lambda - \bar{\lambda} \in [DF(\bar{x})\mathbb{R}^n]^\perp \cap [\mathcal{U}_3]^\perp$. However, by taking the orthogonal complement, the nondegeneracy condition is equivalent to

$$[DF(\bar{x})\mathbb{R}^n]^\perp \cap [\text{lin } N_{\partial\phi(F(\bar{x}))}(\bar{\lambda})]^\perp = \{0\}, \tag{5.1.14}$$

see, e.g., [11, Proposition 6.26]. This clearly implies $\lambda = \bar{\lambda}$. Next, to prove the second part, let us suppose that the nondegeneracy condition is not satisfied. Then, due to (5.1.14), there exists $v \neq 0$ such that $v \in \ker DF(\bar{x})^\top$ and $v \in [\text{lin } N_{\partial\phi(F(\bar{x}))}(\bar{\lambda})]^\perp = \text{aff } \partial\phi(F(\bar{x})) - \bar{\lambda}$. Obviously, this establishes

$$DF(\bar{x})^\top(tv) = 0 \quad \text{and} \quad \bar{\lambda} + tv \in \text{aff } \partial\phi(F(\bar{x})), \tag{5.1.15}$$

for all $t \in \mathbb{R}$. Now, the second part of equation (5.1.15) and the strict complementarity condition imply $\bar{\lambda} + tv \in \partial\phi(F(\bar{x}))$ for all $t > 0$ sufficiently small. Next, by combining $\bar{\lambda} \in \mathcal{M}(\bar{x})$ and (5.1.15), we obtain

$$\nabla f(\bar{x}) + DF(\bar{x})^\top(\bar{\lambda} + tv) = 0$$

and, consequently, it holds $\bar{\lambda} + tv \in \mathcal{M}(\bar{x})$ for all $t > 0$ sufficiently small. Clearly, this is a contradiction to the assumption $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$. □

5.1.4. Isolated stationarity

The following theorem establishes several conditions under which isolated stationarity of an arbitrary stationary point of problem $(\mathcal{P}_c)$ can be guaranteed. This result (together with Corollary 5.1.14) can be traced back to Robinson [203, Theorem 2.3]. Similar results can also be found in section 4.4.4 in [27].

Theorem 5.1.12 (Isolated stationarity). *Let $\bar{x} \in F^{-1}(\mathrm{dom}\ \phi)$ be a stationary point of problem $(\mathcal{P}_c)$. Suppose that the second order sufficient conditions* (5.1.8) *hold at $\bar{x}$ and that one of the following conditions is satisfied:*

(i) *The nondegeneracy condition* (5.1.10) *holds at $\bar{x}$.*

(ii) *$DF(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is onto.*

(iii) *The multiplier $\bar{\lambda} \in \mathcal{M}(\bar{x})$ fulfills the strict constraint qualification* (5.1.9).

Then, $\bar{x}$ is an isolated stationary point of $(\mathcal{P}_c)$.

Proof. We slightly adjust the proof given in [203, Theorem 2.3] and prove this statement by contradiction. (We also want to refer to the proofs of [27, Theorem 4.51 and Proposition 4.52], where similar techniques were used). Let $(x^k)_k$ be a sequence of stationary points of problem $(\mathcal{P}_c)$ that converges to $\bar{x}$ and let $(\lambda^k)_k$ be a corresponding sequence of multipliers, i.e., $\lambda^k \in \mathcal{M}(x^k)$ for all $k \in \mathbb{N}$. Since each of the conditions (i)–(iii) implies that Robinson's constraint qualification is satisfied at $\bar{x}$, Lemma 5.1.4 ensures boundedness of the sequence $(\lambda^k)_k$. Furthermore, due to Lemma 5.1.7 (i) and 5.1.11 (i), the set of Lagrange multipliers also reduces to a singleton $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$.

Now, there exists a subsequence $(\lambda^k)_{k\in K_1}$ of $(\lambda^k)_k$ that converges to some limit $\tilde{\lambda} \in \mathbb{R}^m$. The continuity of ∇f, F, DF, and of the proximity operator yields

$$\begin{aligned} \nabla f(\bar{x}) + DF(\bar{x})^\top\tilde{\lambda} &= \lim_{K_1 \ni k \to \infty} \nabla f(x^k) + DF(x^k)^\top\lambda^k = 0, \\ F(\bar{x}) - \mathrm{prox}_\phi^I(F(\bar{x}) + \tilde{\lambda}) &= \lim_{K_1 \ni k \to \infty} F(x^k) - \mathrm{prox}_\phi^I(F(x^k) + \lambda^k) = 0. \end{aligned}$$

Obviously, this shows $\tilde{\lambda} \in \mathcal{M}(\bar{x})$ and, since $\mathcal{M}(\bar{x})$ is a singleton, we also have $\tilde{\lambda} = \bar{\lambda}$. Next, using $\lambda^k \in \partial\phi(F(x^k))$ and a first order Taylor expansion, we obtain

$$\phi(F(x^k)) - \phi(F(\bar{x})) \le \langle \lambda^k, F(x^k) - F(\bar{x})\rangle = \langle \lambda^k, DF(\bar{x})(x^k - \bar{x})\rangle + o(\|x^k - \bar{x}\|).$$

Let us define $t_k := \|x^k - \bar{x}\|$ and $h^k := (x^k - \bar{x})/t_k$. Then, by passing to another subsequence $K \subseteq K_1$ if necessary, we can assume that $(h^k)_{k\in K}$ converges to some $h \in \mathbb{R}^n$. It follows

$$\frac{F(x^k) - F(\bar{x})}{t_k} = \frac{F(\bar{x} + t_k h^k) - F(\bar{x})}{t_k} = DF(\bar{x})h^k + \frac{o(t_k)}{t_k} \xrightarrow[K \ni k \to \infty]{} DF(\bar{x})h$$

and

$$\begin{aligned} \phi^{\downarrow}(F(\bar{x}); DF(\bar{x})h) &\le \liminf_{K \ni k \to \infty} \frac{\phi(F(x^k)) - \phi(F(\bar{x}))}{t_k} \\ &\le \liminf_{K \ni k \to \infty} \langle \lambda^k, DF(\bar{x})h^k \rangle + \frac{o(t_k)}{t_k} = \langle \bar{\lambda}, DF(\bar{x})h \rangle. \end{aligned}$$

Adding $\nabla f(\bar{x})^\top h$ on both side of the latter inequality and applying Corollary 2.5.7, we establish

$$\psi_c^{\downarrow}(\bar{x}; h) = \nabla f(\bar{x})^\top h + \phi^{\downarrow}(F(\bar{x}); DF(\bar{x})h) \le \langle \nabla f(\bar{x}) + DF(\bar{x})^\top \bar{\lambda}, h \rangle = 0.$$

Together with the stationarity of $\bar{x}$, this implies $h \in \mathcal{C}(\bar{x}) \setminus \{0\}$. Next, a Taylor expansion of ∇f, F, and DF at $\bar{x}$ and Lemma 2.5.14 yield

$$\begin{aligned} 0 &= -\nabla f(x^k) - DF(x^k)^\top \lambda^k \\ &= -\nabla f(\bar{x}) - DF(\bar{x})^\top \lambda^k - t_k \left(\nabla^2 f(\bar{x})h^k + \sum_{i=1}^m \lambda_i^k \cdot \nabla^2 F_i(\bar{x})h^k - \frac{o(t_k)}{t_k} \right) \\ &= -DF(\bar{x})^\top(\lambda^k - \bar{\lambda}) - t_k \left(\nabla^2 f(\bar{x})h^k + \sum_{i=1}^m \lambda_i^k \cdot \nabla^2 F_i(\bar{x})h^k - \frac{o(t_k)}{t_k} \right) \end{aligned} \tag{5.1.16}$$

and

$$\partial\phi^*(\lambda^k) \ni F(x^k) = F(\bar{x}) + t_k DF(\bar{x})h^k + t_k^2 D^2F(\bar{x})[h^k, h^k] + o(t_k^2).$$

Furthermore, by the monotonicity of the subdifferential operator $\partial\phi^*$ (combine, e.g., [11, Theorem 20.40] and Lemma 2.5.14), we have

$$0 \le \langle F(x^k) - F(\bar{x}), \lambda^k - \bar{\lambda} \rangle = \langle \lambda^k - \bar{\lambda}, t_k DF(\bar{x})h^k + t_k^2 D^2F(\bar{x})[h^k, h^k] \rangle + o(t_k^2).$$

By multiplying (5.1.16) with $t_k(h^k)^\top$ and adding the latter inequality, we get

$$-(h^k)^\top \nabla^2 f(\bar{x})h^k - \langle \bar{\lambda}, D^2F(\bar{x})[h^k, h^k] \rangle + \frac{o(t_k)}{t_k} + \frac{o(t_k^2)}{t_k^2} \ge 0.$$

Taking the limit $K \ni k \to \infty$, we finally obtain

$$h^\top \nabla^2 f(\bar{x})h + \langle \bar{\lambda}, D^2F(\bar{x})[h, h] \rangle \le 0.$$

However, due to $h \in \mathcal{C}(\bar{x}) \setminus \{0\}$ and $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$, this contradicts the second order sufficient conditions (5.1.8). □

Remark 5.1.13. Let us note, that this proof also works under slightly weaker assumptions. Specifically, if Robinson's constraint qualification is satisfied at $\bar{x}$, then the set $\mathcal{M}(x)$ is uniformly bounded in a neighborhood of $\bar{x}$ and the sequence $(\lambda^k)_k$, which was constructed in the proof of Theorem 5.1.12, will stay in a bounded set. Of course, if $\mathcal{M}(\bar{x})$ is not a singleton, a stronger form of the second order sufficient conditions has to be used, in order to establish a contradiction at the end of the proof.

The latter observations motivate the following corollary, which obviously requires no proof.

Corollary 5.1.14. *Let $\bar{x} \in F^{-1}(\text{dom } \phi)$ be a stationary point of $(\mathcal{P}_c)$. Suppose that the* strong second order sufficient conditions *hold at $\bar{x}$*

$$h^\top \nabla^2 f(\bar{x}) h + \langle \lambda, D^2 F(\bar{x})[h,h] \rangle > 0, \quad \forall\ h \in \mathcal{C}(\bar{x}) \setminus \{0\}, \quad \forall\ \lambda \in \mathcal{M}(\bar{x})$$

and that Robinson's constraint qualification is satisfied at $\bar{x}$. Then, $\bar{x}$ is an isolated stationary point of $(\mathcal{P}_c)$

Remark 5.1.15. Let us briefly reconsider our initial problem

$$\min_x \ f(x) + \varphi(x) = \psi(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous mapping. Clearly, in this situation, we have $F \equiv I$ and, consequently, condition (ii) in Theorem 5.1.12 is satisfied. Let $\bar{x} \in \text{dom } \varphi$ be an arbitrary stationary point, then the corresponding second order sufficient conditions (5.1.8) reduce to the condition

$$h^\top \nabla^2 f(\bar{x}) h > 0, \quad \forall\ h \in \mathcal{C}(\bar{x}) = \{h \in \mathbb{R}^n : \psi^\downarrow(\bar{x}; h) = 0\}.$$

Thus, if the Hessian of f at $\bar{x}$ is positive definite on the critical cone $\mathcal{C}(\bar{x})$, then $\bar{x}$ is a strict local minimum and an isolated stationary point of problem $(\mathcal{P})$. In the following sections, we will discuss and present situations where isolated stationarity can be obtained under weaker assumptions. In particular, we have ignored any possible second order information of the nonsmooth function φ so far.

5.2. No gap second order conditions

We will now introduce and present a pair of so-called no gap second order conditions for the convex composite problem $(\mathcal{P}_c)$. In contrast to the previous discussions, these conditions also take the possible curvature of the nonsmooth and nonlinear function ϕ into account. Since the formulation and derivation of these general second order conditions relies on a number of not yet mentioned concepts, such as, e.g., second order directional derivatives of ϕ or outer second order regularity, we start with some preliminary definitions.

This subsection is primarily based on the sections 3.2.1, 3.3.4, and 3.4.1 in [27] and summarizes the most important results of Bonnans and Shapiro. For more details and information on second order conditions for composite functions we refer to [24, 27]. To improve the overall comprehensibility of this subsection and of the abstract second order theory, we decided to recreate some of the proofs given in [27] and to "translate" them to the convex composite setting. The complete proofs can be found in the Appendix.

5.2.1. Second order directional (epi-)derivatives and second order tangent sets

At first, we will give a brief overview on second order directional derivatives and their relation to second order tangent sets. Let us note that the following definitions are formulated for general functions ϱ, g, and G that are not necessarily connected to our initial problem or to problem $(\mathcal{P}_c)$.

Definition 5.2.1. *Let $\varrho : \mathbb{R}^n \to (-\infty, +\infty]$ and $x \in \operatorname{dom} \varrho$ be given and suppose that ϱ is directionally differentiable at x in the direction $h \in \mathbb{R}^n$ such that $\varrho'(x;h)$ is finite. Then, the* lower *and* upper (parabolic) second order directional derivatives *of ϱ at x are defined by:*

$$\varrho''_-(x;h,w) := \liminf_{t\downarrow 0} \frac{\varrho(x+th+\frac{1}{2}t^2w)-\varrho(x)-t\varrho'(x;h)}{\frac{1}{2}t^2},$$
$$\varrho''_+(x;h,w) := \limsup_{t\downarrow 0} \frac{\varrho(x+th+\frac{1}{2}t^2w)-\varrho(x)-t\varrho'(x;h)}{\frac{1}{2}t^2}.$$

We say that ϱ is twice (parabolically) directionally differentiable *at x, in the direction h, if the upper and lower second order directional derivatives coincide for all $w \in \mathbb{R}^n$. In that case, the term $\varrho''(x;h,w)$, $w \in \mathbb{R}^n$, will be used to denote the common values.*

Apparently, if the function ϱ is twice continuously differentiable at x, then it is also twice directionally differentiable at x and a second order Taylor expansion immediately yields

$$\varrho''(x;h,w) = \nabla\varrho(x)^\top w + h^\top\nabla^2\varrho(x)h.$$

Next, we introduce an analogous terminology for second order directional epiderivatives.

Definition 5.2.2. *Let $\varrho : \mathbb{R}^n \to (-\infty, +\infty]$, $x \in \operatorname{dom} \varrho$, and $h \in \mathbb{R}^n$ be given and suppose that the directional epiderivatives $\varrho^\downarrow_-(x;h)$ and $\varrho^\downarrow_+(x;h)$ are finite. Then, the* lower *and* upper (parabolic) second order directional epiderivatives *of ϱ at x are defined as follows:*

$$\varrho^{\downarrow\downarrow}_-(x;h,w) := \liminf_{t\downarrow 0,\, \tilde{w}\to w} \frac{\varrho(x+th+\frac{1}{2}t^2\tilde{w})-\varrho(x)-t\varrho^\downarrow_-(x;h)}{\frac{1}{2}t^2},$$
$$\varrho^{\downarrow\downarrow}_+(x;h,w) := \sup_{(t_k)_k\in\mathcal{N}_0} \liminf_{k\to\infty,\, \tilde{w}\to w} \frac{\varrho(x+t_kh+\frac{1}{2}t_k^2\tilde{w})-\varrho(x)-t_k\varrho^\downarrow_+(x;h)}{\frac{1}{2}t_k^2}.$$

In addition, if ϱ is directionally epidifferentiable at x, in the direction h, and the lower and upper second order directional epiderivatives coincide for all $w \in \mathbb{R}^n$, then ϱ is said to be twice (parabolically) directionally epidifferentiable *at x, in the direction h. In this case, we will use the term $\varrho^{\downarrow\downarrow}(x;h,w)$, $w \in \mathbb{R}^n$, to denote the coinciding derivatives.*

Again, if ϱ is Lipschitz continuous in a neighborhood of x and directionally differentiable at x, then for all $h, w \in \mathbb{R}^n$, the second order epiderivatives $\varrho^{\downarrow\downarrow}_-(x;h,w)$ and $\varrho^{\downarrow\downarrow}_+(x;h,w)$ reduce to the more conventional lower and upper second order directional derivatives. Clearly, our definition of second order epidifferentiability can be equivalently rephrased as epi-convergence of the second order difference quotients

$$\Delta_t^2\, \varrho(x;h)(w) := \frac{\varrho(x+th+\frac{1}{2}t^2w) - \varrho(x) - t\varrho^{\downarrow}(x;h)}{\frac{1}{2}t^2},$$

for all $w \in \mathbb{R}^n$. A thorough treatment and further properties of second order (epi-)derivatives can be found in [15, 16, 206], [208, Section 13.J], and [27, Chapter 3]. As in Lemma 2.5.3, the second order epiderivatives can be connected to so-called *second order tangent sets.*

Definition 5.2.3 (cf. [27, Definition 3.28]). *Let $S \subset \mathbb{R}^n$ be nonempty and let $x \in S$ be given. The sets*

$$\begin{aligned} T_S^{i,2}(x,h) &:= \left\{w \in \mathbb{R}^n : \operatorname{dist}(x+th+\tfrac{1}{2}t^2w, S) = o(t^2),\, t \geq 0\right\}, \\ T_S^2(x,h) &:= \left\{w \in \mathbb{R}^n : \exists\ t_k \downarrow 0 \textit{ such that } \operatorname{dist}(x+t_kh+\tfrac{1}{2}t_k^2w, S) = o(t_k^2)\right\} \end{aligned}$$

are called the inner *and* outer second order tangent sets *to the set S at the point x in the direction $h \in \mathbb{R}^n$, respectively.*

The inner and outer second order tangent sets are closed and it holds $T_S^{i,2}(x,h) \subset T_S^2(x,h)$ for all $h \in \mathbb{R}^n$. Moreover, if the set S is convex, then the distance function $\operatorname{dist}(\cdot, S)$ is convex and, in that case, the inner second order tangent set $T_S^{i,2}(x,h)$ is also a convex set. The outer second order tangent set $T_S^2(x,h)$ can be nonconvex, even if S is a convex set.

Lemma 5.2.4 (cf. [27, Proposition 3.41]). *Let $\varrho : \mathbb{R}^n \to (-\infty, +\infty]$ and $x \in \operatorname{dom}\, \varrho$ be given. For $h \in \mathbb{R}^n$, suppose that $\varrho_-^{\downarrow}(x;h)$ and $\varrho_+^{\downarrow}(x;h)$ are finite. Then, we have*

$$T^2_{\text{epi }\varrho}[(x,\varrho(x)),(h,\varrho_-^{\downarrow}(x;h))] = \text{epi } \varrho_-^{\downarrow\downarrow}(x;h,\cdot),$$

$$T^{i,2}_{\text{epi }\varrho}[(x,\varrho(x)),(h,\varrho_+^{\downarrow}(x;h))] = \text{epi } \varrho_+^{\downarrow\downarrow}(x;h,\cdot).$$

Lemma 5.2.4 immediately implies that the upper second order directional epiderivative $\varrho_+^{\downarrow\downarrow}(x;h,\cdot)$ is convex, when the function ϱ is convex and $\varrho^{\downarrow}(x;h)$ is finite. We conclude this preparatory subsection with a chain rule for second order epiderivatives.

Lemma 5.2.5 (cf. [27, Proposition 3.42]). *Let $G : \mathbb{R}^n \to \mathbb{R}^m$ be a twice continuously differentiable function and let $\varrho : \mathbb{R}^m \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. Moreover, suppose that Robinson's constraint qualification*

$$0 \in \operatorname{int}\{G(x) + DG(x)\mathbb{R}^n - \operatorname{dom}\, \varrho\}$$

is satisfied at $x \in G^{-1}(\operatorname{dom}\, \varrho)$ and that $\varrho^{\downarrow}(G(x); DG(x)h)$ is finite. Then, it holds

$$(\varrho \circ G)_-^{\downarrow\downarrow}(x;h,w) = \varrho_-^{\downarrow\downarrow}(G(x); DG(x)h, DG(x)w + D^2G(x)[h,h]),$$

$$(\varrho \circ G)_+^{\downarrow\downarrow}(x;h,w) = \varrho_+^{\downarrow\downarrow}(G(x); DG(x)h, DG(x)w + D^2G(x)[h,h]).$$

5.2.2. Outer second order regularity and second order conditions

The concept of outer second order regularity was introduced and developed in [22, 24, 27]. It is an essential analytical tool and one of the central ideas that allows to close the gap between second order sufficient and necessary conditions. The following definition is taken from [27, Definition 3.94].

Definition 5.2.6 (Outer second order regularity). *Let $\varrho : \mathbb{R}^n \to (-\infty, +\infty]$ be an extended real valued function and let $x \in \mathrm{dom}\ \varrho$ be arbitrary. We say that ϱ is* outer second order regular *at the point x, in a direction $h \in \mathbb{R}^n$, if $\varrho^{\downarrow}_{-}(x, h)$ is finite and if for any sequences $(t_k)_k$, $t_k \downarrow 0$, and $(w^k, \tau^k)_k \subset \mathbb{R}^{n+1}$ that satisfy $t_k(w^k, \tau^k) \to 0$ and*

$$\varrho(x + t_k h + \tfrac{1}{2} t_k^2 w^k) \leq \varrho(x) + t_k \varrho^{\downarrow}_{-}(x; h) + \tfrac{1}{2} t_k^2 \tau^k \tag{5.2.1}$$

there exists a sequence $(\tilde{w}^k, \tilde{\tau}^k)_k \subset \mathbb{R}^{n+1}$ such that, for all $k \in \mathbb{N}$, it holds $\varrho^{\downarrow\downarrow}_{-}(x; h, \tilde{w}^k) \leq \tilde{\tau}^k$ and

$$\tilde{w}^k - w^k \to 0, \quad \tilde{\tau}^k - \tau^k \to 0, \quad \text{as } k \to \infty.$$

The function ϱ is said to be outer second order regular *at x on a set $\mathcal{H} \subset \mathbb{R}^n$, if $\varrho^{\downarrow}_{-}(x; h)$ is finite for all $h \in \mathcal{H}$ and if ϱ is outer second order regular at x, in all directions $h \in \mathcal{H}$.*

Let us note that condition (5.2.1) is equivalent to

$$\begin{pmatrix} x \\ \varrho(x) \end{pmatrix} + t_k \begin{pmatrix} h \\ \varrho^{\downarrow}_{-}(x; h) \end{pmatrix} + \frac{1}{2} t_k^2 \begin{pmatrix} w^k \\ \tau^k \end{pmatrix} \in \mathrm{epi}\ \varrho. \tag{5.2.2}$$

Thus, outer second order regularity implies that for any parabolic curve of the form (5.2.2), which is entirely contained in the epigraph epi ϱ and which is tangential to the direction $(h, \varrho^{\downarrow}_{-}(x; h)) \in T_{\mathrm{epi}\ \varrho}(x, \varrho(x))$, the term (w^k, τ^k) will eventually approach the outer second order tangent set $T^2_{\mathrm{epi}\ \varrho}[(x, \varrho(x)), (h, \varrho^{\downarrow}_{-}(x; h))]$ as $k \to \infty$. Next, we present a chain rule-type result for outer second order regular functions.

Lemma 5.2.7. *Let $g : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable and let $\varrho : \mathbb{R}^m \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function. Furthermore, suppose that Robinson's constraint qualification is satisfied at $x \in G^{-1}(\mathrm{dom}\ \varrho)$ and ϱ is outer second order regular at $G(x)$ in the direction $DG(x)h$. Then, the composite function $\psi : \mathbb{R}^n \to (-\infty, +\infty]$, $\psi(x) := g(x) + \varrho(G(x))$, is outer second order regular at x in the direction h.*

Proof. A proof can be found in [27, Proposition 3.88 and 3.96]. □

We are now ready to state the no gap second order conditions for the convex composite problem $(\mathcal{P}_c)$. Recall that in our initial setting the functions $f : \mathbb{R}^n \to \mathbb{R}$, $F : \mathbb{R}^n \to \mathbb{R}^m$ are supposed to be twice continuously differentiable and $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ is assumed to be convex, proper, and lower semicontinuous.

This theorem summarizes and combines the Theorems 3.45, 3.83, 3.86, 3.108, and 3.109 in [27], see also [24, Theorem 5.2] for an analogue formulation. As already mentioned, a complete proof of Theorem 5.2.8 is provided in the appendix.

Theorem 5.2.8 (No gap second order conditions). *Let $\bar{x} \in F^{-1}(\mathrm{dom}\,\phi)$ be given and suppose that Robinson's constraint qualification is fulfilled at $\bar{x}$. Then, the following statements do hold:*

(i) (Second order necessary conditions). *Additionally, let $\bar{x}$ be a locally optimal solution of $(\mathcal{P}_c)$. Then, for any $h \in \mathcal{C}(\bar{x})$ and any convex function $\zeta(\cdot) \geq \phi_-^{\downarrow\downarrow}(F(\bar{x}), DF(\bar{x})h, \cdot)$ the following inequality is satisfied:*

$$\max_{\lambda \in \mathcal{M}(\bar{x})} \left\{ h^\top \nabla^2 f(\bar{x}) h + \langle \lambda, D^2 F(\bar{x})[h,h] \rangle - \zeta^*(\lambda) \right\} \geq 0. \tag{5.2.3}$$

(ii) (Second order sufficient conditions). *Let $\bar{x}$ be a stationary point of problem $(\mathcal{P}_c)$. Suppose that for every $h \in \mathcal{C}(\bar{x})$, the function ϕ is outer second order regular at $F(\bar{x})$ in the direction $DF(\bar{x})h$ and that it holds*

$$\max_{\lambda \in \mathcal{M}(\bar{x})} \left\{ h^\top \nabla^2 f(\bar{x}) h + \langle \lambda, D^2 F(\bar{x})[h,h] \rangle - \xi_{\phi,h}^*(\lambda) \right\} > 0, \tag{5.2.4}$$

for all $h \in \mathcal{C}(\bar{x}) \setminus \{0\}$, where $\xi_{\phi,h}(\cdot) := \phi_-^{\downarrow\downarrow}(F(\bar{x}), DF(\bar{x})h, \cdot)$. Then, for some $\alpha > 0$ and all x in a neighborhood of $\bar{x}$ it follows

$$f(x) + \phi(F(x)) \geq f(\bar{x}) + \phi(F(\bar{x})) + \alpha \|x - \bar{x}\|^2, \tag{5.2.5}$$

and hence, $\bar{x}$ is a locally optimal solution of $(\mathcal{P}_c)$. Moreover, if the function $\xi_{\phi,h}$ is convex for every $h \in \mathcal{C}(\bar{x})$, then the second order conditions (5.2.4) *are necessary and sufficient for the quadratic growth condition* (5.2.5) *and there is no gap between the second order necessary and sufficient conditions.*

In the second order necessary conditions we can always choose $\zeta(\cdot) = \phi_+^{\downarrow\downarrow}(F(\bar{x}); DF(\bar{x})h, \cdot)$ as an upper and convex estimate of the function $\xi_{\phi,h}$. Moreover, if ϕ is twice directionally epidifferentiable at $F(\bar{x})$, in all directions $DF(\bar{x})h$, $h \in \mathcal{C}(\bar{x})$, then the epiderivative $\xi_{\phi,h}$ is convex and there occurs no gap between the second order conditions.

Remark 5.2.9. Let $\bar{x}$ be a stationary point of problem $(\mathcal{P}_c)$. Then, ϕ is subdifferentiable at $F(\bar{x})$ and for any $t > 0$ and $h, w \in \mathbb{R}^n$ it holds

$$\phi(F(\bar{x}) + tDF(\bar{x})h + \tfrac{1}{2}t^2 w) - \phi(F(\bar{x})) \geq \langle \lambda, tDF(\bar{x})h + \tfrac{1}{2}t^2 w \rangle, \quad \forall\, \lambda \in \partial\phi(F(\bar{x})).$$

Now, choosing $\lambda \in \mathcal{M}(\bar{x}) \subset \partial\phi(F(\bar{x}))$ and $h \in \mathcal{C}(\bar{x})$, the latter inequality yields

$$\begin{aligned} &\phi_-^{\downarrow\downarrow}(F(\bar{x}); DF(\bar{x})h, w) \\ &\qquad \geq \liminf_{t \downarrow 0,\, \tilde{w} \to w} \frac{\frac{1}{2}t^2 \langle \lambda, \tilde{w} \rangle - t \cdot (\nabla f(\bar{x})^\top h + \phi_-^{\downarrow}(F(\bar{x}); DF(\bar{x})h))}{\frac{1}{2}t^2} = \langle \lambda, w \rangle, \end{aligned}$$

where we used $DF(\bar{x})^\top \lambda = -\nabla f(\bar{x})$ and the definition of the critical cone $\mathcal{C}(\bar{x})$. Hence, this immediately implies

$$\xi_{\phi,h}^*(\lambda) = \sup_{w \in \mathbb{R}^n} \langle \lambda, w \rangle - \phi_-^{\downarrow\downarrow}(F(\bar{x}); DF(\bar{x})h, w) \leq 0.$$

Furthermore, since the convex function ζ in (5.2.3) satisfies $\zeta(\cdot) \geq \xi_{\phi,h}(\cdot)$, we can apply [11, Proposition 13.14] and obtain the following, final estimate

$$\zeta^*(\lambda) \leq \xi^*_{\phi,h}(\lambda) \leq 0, \quad \forall\, \lambda \in \mathcal{M}(\bar{x}), \quad \forall\, h \in \mathcal{C}(\bar{x}).$$

This shows that the second order sufficient condition (5.2.4) is generally weaker than the second order condition (5.1.8), which was discussed in the previous subsection. Moreover, the second order necessary conditions indicate that the stronger conditions (5.1.8) cannot be expected to hold at a local minimum of $(\mathcal{P}_c)$ without any further assumptions on the nonsmooth function ϕ.

Remark 5.2.10. A careful study of the proof of Theorem 5.2.8 shows that the second order optimality conditions can also be stated in terms of the lower second order directional epiderivative $(\psi_c)_-^{\downarrow\downarrow}$. In particular, it is possible to derive the following pair of second order conditions:

Let $\bar{x}$ be a local solution of problem $(\mathcal{P}_c)$, then it holds

$$\inf_w \, (\psi_c)_-^{\downarrow\downarrow}(\bar{x}; h, w) \geq 0, \quad \forall\, h \in \mathcal{C}(\bar{x}).$$

On the contrary, assume that $\bar{x}$ satisfies the first order necessary condition $(\psi_c)_-^{\downarrow}(\bar{x}; h) \geq 0$ for all $h \in \mathbb{R}^n$ and suppose that ψ_c is outer second order regular at $\bar{x}$ in all directions $h \in \mathcal{C}(\bar{x}) = \{h : (\psi_c)_-^{\downarrow}(\bar{x}; h) \leq 0\}$. Then, the second order growth condition holds at $\bar{x}$ if and only if the following second order sufficient conditions are satisfied

$$\inf_w \, (\psi_c)_-^{\downarrow\downarrow}(\bar{x}; h, w) > 0, \quad \forall\, h \in \mathcal{C}(\bar{x}) \setminus \{0\}.$$

Clearly, these alternative conditions have the advantage that Robinson's constraint qualification is not needed explicitly. Second order conditions of this form were already studied and introduced by Ben-Tal and Zowe [15, 16] by using (parabolic) second order directional derivatives. We also refer to [27, Proposition 3.105] for more details.

Remark 5.2.11. Let us briefly discuss the corresponding pair of no gap second order conditions for our initial problem $(\mathcal{P})$. Again, as in Remark 5.1.15, we set $\phi \equiv \varphi$ and $F \equiv I$. Obviously, in this situation, the nondegeneracy condition (5.1.10) is satisfied at any point $x \in \operatorname{dom} \varphi$ and, due to Lemma 4.1.2 (ii), the set of Lagrange multipliers reduces to the singleton $\mathcal{M}(\bar{x}) = \{-\nabla f(\bar{x})\}$ when $\bar{x} \in \operatorname{dom} \varphi$ is a stationary point. Now, let $\bar{x} \in \operatorname{dom} \varphi$ be a local solution of the initial problem $(\mathcal{P})$. Then, by (5.2.3), for any $h \in \mathcal{C}(\bar{x})$ and any convex function $\zeta(\cdot) \geq \varphi_-^{\downarrow\downarrow}(\bar{x}; h, \cdot)$, we have

$$h^\top \nabla^2 f(\bar{x}) h - \zeta^*(-\nabla f(\bar{x})) \geq 0.$$

On the other hand, let $\bar{x}$ be a stationary point of problem $(\mathcal{P})$. Let us assume that φ is outer second order regular at $\bar{x}$ on $\mathcal{C}(\bar{x})$ and that for all $h \in \mathcal{C}(\bar{x}) \setminus \{0\}$ the second order sufficient condition

$$h^\top \nabla^2 f(\bar{x}) h - \xi^*_{\varphi,h}(-\nabla f(\bar{x})) > 0$$

is satisfied with $\xi_{\varphi,h}(\cdot) := \varphi_-^{\downarrow\downarrow}(\bar{x}; h, \cdot)$. Then, by Theorem 5.2.8 (ii), $\bar{x}$ is a (strict) locally optimal solution of $(\mathcal{P})$. As we have seen in Remark 5.2.9, this second order sufficient condition is certainly weaker than the condition presented in Remark 5.1.15. Unfortunately, due to the presence of the term $\xi^*_{\varphi,h}$, we cannot infer that $\bar{x}$ is also an isolated stationary point of the problem $(\mathcal{P})$. In the next section, we introduce several classes of nonsmooth functions φ that will allow us to resolve this disadvantage.

Finally, we present a specific property of the upper second order directional epiderivative that will be needed for the second order analysis of the proximity operator in the next subsection. This result is based on [23, Corollary 4.1] or [27, Proposition 3.48].

Lemma 5.2.12. *Let $\varrho : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex and proper function and let $x \in \operatorname{dom} \varrho$ and $\lambda \in \mathbb{R}^n$ be given. Suppose that there exists a convex, nonempty set $\Omega \subset \operatorname{dom} \varrho^\downarrow(x; \cdot)$, such that $\varrho^\downarrow(x; h)$ is finite for every $h \in \Omega$. Then, the function*

$$\Xi_\lambda : \mathbb{R}^n \to [-\infty, +\infty], \quad \Xi_\lambda(h) := -\sup_{w \in \mathbb{R}^n} \langle \lambda, w \rangle - \varrho^{\downarrow\downarrow}_+(x; h, w)$$

is convex on Ω.

Proof. First, let us note that the upper second order directional epiderivative $\varrho^{\downarrow\downarrow}_+(x; h, w)$ is well-defined for all $h \in \Omega$ and $w \in \mathbb{R}^n$. We rewrite the function Ξ_λ as follows

$$\begin{aligned} \Xi_\lambda(h) &= -\sup_{w \in \mathbb{R}^n} \langle \lambda, w \rangle - \varrho^{\downarrow\downarrow}_+(x; h, w) \\ &= -\sup_{(w,\gamma) \in \mathbb{R}^n \times \mathbb{R}} \{\langle \lambda, w \rangle - \gamma : (w, \gamma) \in \operatorname{epi} \varrho^{\downarrow\downarrow}_+(x; h, \cdot)\} = -\sigma_{\operatorname{epi} \varrho^{\downarrow\downarrow}_+(x;h,\cdot)}(\lambda, -1). \end{aligned}$$

Due to Lemma 5.2.4, the epigraph $\operatorname{epi} \varrho^{\downarrow\downarrow}_+(x; h, \cdot)$ coincides with the inner second order tangent set $T^{i,2}_{\operatorname{epi} \varrho}[(x, \varrho(x)), (h, \varrho^\downarrow(x; h))]$. However, in this situation, [27, Proposition 3.48] is applicable, which establishes the convexity of Ξ_λ on Ω. □

Obviously, if the function ϱ is subdifferentiable at x, then the set Ω in Lemma 5.2.12 can be chosen as $\Omega = \operatorname{dom} \varrho^\downarrow(x; \cdot)$.

5.3. Decomposable functions

In the following section, we will introduce and discuss an important class of nonsmooth and not necessarily convex functions for which the curvature term in the second order conditions (5.2.3) and (5.2.4) has an easy representation. As anticipated in Remark 5.2.11, this will allow us to formulate no gap second order conditions that additionally guarantee isolated stationarity of a stationary point of problem $(\mathcal{P})$ and that combine our different theoretical results we have developed so far.

The concept of *decomposable functions* was proposed by Shapiro in [217] and is strongly related to the notions of *amenable functions*, see, e.g., [193, 194] or [208, Chapter 10.F], and of *C^ℓ-cone reducible sets* in nonlinear, constrained optimization. We will see that the class of decomposable functions is quite rich and a large number of nonsmooth optimization problems can be treated within the framework of decomposable function. In particular, we will show that ℓ_1- and group sparse problems possess a decomposable structure. Moreover, decomposability is also applicable to structural more advanced problems such as semidefinite programming or nuclear norm regularized optimization problems.

Before stating the main definition, let us give a short summary of our current position. In Remark 5.1.15 and 5.2.11, we have seen that the choice $F \equiv I$ leads to two basic second order sufficient conditions that both seem to be too strong and too weak to characterize or ensure isolated stationarity. As already mentioned, the overall idea is now to rewrite φ as a suitable composite function $\varphi \equiv \varphi_d \circ F$ and to shift the second order discussion to the corresponding convex composite problem $(\mathcal{P}_d)$. In this respect, if $\bar{x}$ is a stationary point of problem $(\mathcal{P})$, then, by comparing Theorem 5.2.8 and Theorem 5.1.12, the decomposition pair (φ_d, F) should be chosen such that:

- The function φ_d is outer second order regular and twice epidifferentiable at $F(\bar{x})$, in all directions $DF(\bar{x})h$, $h \in \mathcal{C}(\bar{x})$.
- The curvature term $\xi^*_{\varphi_d,h}(\lambda)$ vanishes for all multipliers $\lambda \in \mathcal{M}(\bar{x})$ and $h \in \mathcal{C}(\bar{x})$.
- The mapping $DF(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is onto or the nondegeneracy condition is satisfied at $\bar{x}$. (Or other regularity conditions hold at $\bar{x}$, see Corollary 5.1.14).

In this case, there will be no gap between the second order necessary and sufficient conditions and Theorem 5.1.12 is applicable. In the following, we will systematically verify and derive these motivational observations and show that the concept of decomposability adequately unites our different demands and results.

In the next paragraph, based on [217, Definition 1.1], we give a definition of decomposable functions. Since the framework in [217] is tailored to the real valued setting, we have to adjust the definitions given in [217] to cope with the possible extended valuedness. Moreover, we also extend the definition in [217] and introduce *fully decomposable functions*, which will play an essential role in our further analysis. Thereafter, we discuss the main properties of decomposable functions, present various illustrating examples and provide several calculation rules. We conclude this section with one of our main results. In particular, by combining diverse theoretical frameworks, such as the $\mathcal{VU}$-theory, second order sensitivity analysis of the Moreau envelope, and the strict complementarity condition, we will establish that the curvature term $\xi^*_{\varphi,h}(-\nabla f(\bar{x}))$ for the initial function φ, which already appeared in Remark 5.2.11, does not depend on a specific decomposition pair and can be expressed via a Moore-Penrose inverse of the Frechét derivative of the proximity operator. This is one of the most fundamental steps to show that, for the class of decomposable functions, the second order sufficient condition also ensures nonsingularity of any element of the Clarke subdifferential $\partial F^\Lambda(\bar{x})$ of a stationary point of $(\mathcal{P})$.

5.3.1. Decomposability

Let us start with a detailed definition of decomposable functions.

Definition 5.3.1 (Decomposable functions). *A function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is called* C^ℓ-decomposable*, $\ell \in \mathbb{N}$, at a point $\bar{x} \in \operatorname{dom} \varphi$, if there exists an open neighborhood U of $\bar{x}$*

such that

$$\varphi(x) = \varphi(\bar{x}) + \varphi_d(F(x)), \quad \forall\, x \in U, \tag{5.3.1}$$

and the functions φ_d and F satisfy:

(i) *$F : \mathbb{R}^n \to \mathbb{R}^m$ is ℓ-times continuously differentiable on U and it holds $F(\bar{x}) = 0$.*

(ii) *The mapping $\varphi_d : \mathbb{R}^m \to (-\infty, +\infty]$ is convex, proper, lower semicontinuous, and positively homogeneous.*

(iii) *Robinson's constraint qualification holds at $\bar{x}$:*

$$0 \in \operatorname{int}\{F(\bar{x}) + DF(\bar{x})\mathbb{R}^n - \operatorname{dom}\, \varphi_d\} = \operatorname{int}\{DF(\bar{x})\mathbb{R}^n - \operatorname{dom}\, \varphi_d\}. \tag{5.3.2}$$

*We say that φ is C^ℓ-*fully decomposable *at $\bar{x}$ if φ is C^ℓ-decomposable at $\bar{x}$ and if, in addition, the nondegeneracy condition*

$$DF(\bar{x})\mathbb{R}^n + \operatorname{lin}\, N_{\partial\varphi_d(0)}(\lambda) = \mathbb{R}^m$$

is satisfied at $\bar{x}$ for an arbitrary subgradient $\lambda \in \partial\varphi_d(0)$.

If φ is decomposable at $\bar{x}$, then the functions $\varphi_d : \mathbb{R}^m \to (-\infty, +\infty]$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ are called a *decomposition pair* of φ. Of course, a decomposition of φ as in (5.3.1) does not need to be unique. Thus, in general, the function φ can have many different decomposition pairs. Let us also mention that, in the fully decomposable case, since the nondegeneracy condition implies Robinson's constraint qualification (5.3.2), assumption (iii) in Definition 5.3.1 is superfluous. We continue with two important remarks.

Remark 5.3.2. Due to Lemma 2.5.13, the function φ_d is subdifferentiable at 0. Hence, the nondegeneracy condition in Definition 5.3.1 is always well-defined. Moreover, since φ_d is convex, proper, and positively homogeneous, the set $DF(\bar{x})\mathbb{R}^n - \operatorname{dom}\, \varphi_d$ is a convex, nonempty cone and, consequently, Robinson's constraint qualification (5.3.2) is equivalent to the condition

$$DF(\bar{x})\mathbb{R}^n - \operatorname{dom}\, \varphi_d = \mathbb{R}^m.$$

Remark 5.3.3 (Stationarity and decomposable optimization problems). So far, we have only considered the trivial decomposition $F \equiv I$. For more general decompositions as in Definition 5.3.1, we have to be more careful when speaking of local solutions and stationary points. If the function φ can be rewritten as a composition $\varphi \equiv \varphi_d \circ F$, then it is clear that every local solution of the initial problem

$$\min_x \; f(x) + \varphi(x) \tag{5.3.3}$$

is also a local minimum of the convex composite problem

$$\min_x \; f(x) + \varphi_d(F(x)) + \bar{c}, \tag{5.3.4}$$

for some constant $\bar{c} \in \mathbb{R}$, and vice versa. However, this analogy does not need to be true for the different notions of stationarity, i.e., a stationary point of the initial problem (5.3.3) is not necessarily a stationary point of the composite problem (5.3.4) in the sense of Definition 5.1.1. Our earlier discussion in Remark 5.1.3 already showed that, in general, we can only expect the following implication

$$\mathcal{M}(\bar{x}) \neq \emptyset \quad \Longrightarrow \quad F^{\Lambda}(\bar{x}) = 0,$$

where $\mathcal{M}(\bar{x}) = \{\lambda \in \partial\varphi_d(F(\bar{x})) : \nabla f(\bar{x}) + DF(\bar{x})^\top \lambda = 0\}$ is the set of corresponding Lagrange multipliers of (5.3.4) and $\Lambda \in \mathbb{S}^n_{++}$ is an arbitrary parameter matrix. Now, if the function φ is C^ℓ-decomposable at some point $\bar{x} \in \mathrm{dom}\ \varphi$ with decomposition pair (φ_d, F), then Robinson's constraint qualification guarantees that these two different stationarity concepts are actually equivalent, i.e., we have

$$\mathcal{M}(\bar{x}) \neq \emptyset \quad \Longleftrightarrow \quad F^{\Lambda}(\bar{x}) = 0.$$

Moreover, due to the stability property of Robinson's constraint qualification (see subsection 2.1.4), the latter equivalence is also satisfied for any other stationary point that lies in a certain neighborhood of $\bar{x}$. Consequently, $\bar{x}$ is an isolated stationary point of our initial problem if and only if $\bar{x}$ is an isolated stationary point of the composite problem (5.3.4). Therefore, in the context of decomposable functions, there is no difference between these two stationarity concepts. Thus, the overall idea of decomposability and decomposable optimization problems can be summarized as follows.

Let us suppose that the function φ is C^ℓ-decomposable at a local minimum or a stationary point $\bar{x} \in \mathrm{dom}\ \varphi$ of (5.3.3). Then, the second order analysis of $\bar{x}$ and of the initial problem (5.3.3) can be completely transferred to the composite problem (5.3.4). Since the latter problem is a general convex composite problem of the form $(\mathcal{P}_c)$, our abstract second order framework and theory of the sections 5.1 and 5.2 can be applied. Furthermore, due to the demonstrated, local equivalence of the problems (5.3.3) and (5.3.4), any optimality and stationarity result that is obtained for the composite setting, can be passed to the original problem (5.3.3).

In the following, we want to briefly assess the class of decomposable functions with respect to its generality and compare it to the so-called class of amenable functions, which was introduced by Poliquin and Rockafellar in [193, 194].

Definition 5.3.4 (cf. [193, Definition 1.1 and 1.2]). *A function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is* amenable *at $\bar{x} \in \mathrm{dom}\ \varphi$ if there is an open neighborhood U of $\bar{x}$ such that*

$$\varphi(x) = \varphi_a(F(x)), \quad \forall\ x \in U,$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is a continuously differentiable function and $\varphi_a : \mathbb{R}^m \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous mapping and the following basic constraint qualification is satisfied:

$$\text{there is no } y \neq 0 \text{ in } N_{\mathrm{dom}\ \varphi_a}(F(\bar{x})) \text{ with } DF(\bar{x})^\top y = 0.$$

The function φ *is called* fully amenable *if such a representation exists and, additionally, the function* F *is twice continuously differentiable and* φ_a *is piecewise linear-quadratic.*

By [27, Proposition 2.97 and Corollary 2.98], the basic constraint qualification is just an equivalent reformulation of Robinson's constraint qualification. Thus, any C^1-decomposable function is amenable. On the other hand, an amenable function is C^1-decomposable if and only if the function φ_a is additionally positively homogeneous and it holds $F(\bar{x}) = 0$. Moreover, it is also clear that a C^2-decomposable function does not need to be fully amenable and vice versa. Finally, in [217], it was shown that a C^2-fully decomposable function is also *partly smooth* in the sense of Lewis [133]. For a more detailed overview and discussion of these different decomposition concepts we refer to Hare [100, 101] and [217]. For instance, a comprehensive visualization of the connection between various classes of functions, including amenable, decomposable and partly smooth functions, can be found in [101, Figure 1].

Let us also mention that in [100, 101] decomposability was defined and studied without explicitly assuming Robinson's constraint qualification. Here, we decided to add this extra regularity condition, since, from a practical point of view, it does not seem to be too restrictive and it notably simplifies the calculus and the overall analysis. In particular, it can be shown that many results for (fully) amenable functions, such as sum and chain rules for second order epiderivatives, do also hold for (fully) decomposable functions and that basic techniques and proofs can be expanded to the class of decomposable functions.

5.3.2. Properties of decomposable functions and decomposable problems

In this subsection, we gradually derive and collect basic properties of decomposable functions. More specifically, we will show that any decomposition pair of a decomposable function φ fulfills the structural requirements that were stated at the beginning of this section.

Therefore, let us suppose that the function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is C^ℓ-decomposable at some point $x \in \operatorname{dom} \varphi$ with corresponding decomposition pair (φ_d, F) and $\ell \in \mathbb{N}$. Then, it follows $\varphi_d(0) = 0$, $\partial\varphi_d(0) \neq \emptyset$, and, due to the lower semicontinuity and positive homogeneity of the mapping φ_d, we have

$$\begin{aligned} \varphi_d(h) \leq \liminf_{\tilde{h} \to h} \; \varphi_d(\tilde{h}) &= \liminf_{t \downarrow 0,\, \tilde{h} \to h} \; \frac{\varphi_d(0 + t\tilde{h}) - \varphi_d(0)}{t} \\ &\leq \liminf_{t \downarrow 0} \; \frac{\varphi_d(0 + th) - \varphi_d(0)}{t} \leq \limsup_{t \downarrow 0} \; \frac{\varphi_d(0 + th) - \varphi_d(0)}{t} = \varphi_d(h) \end{aligned}$$

for all $h \in \mathbb{R}^m$. Using Remark 2.5.4, this shows that φ_d is directionally differentiable and directionally epidifferentiable at 0, in all directions $h \in \mathbb{R}^m$, and both derivatives coincide, i.e.,

$$\varphi_d^{\downarrow}(0; h) = \varphi_d'(0; h) = \varphi_d(h), \quad \forall\, h \in \mathbb{R}^m. \tag{5.3.5}$$

Now, let $h \in \text{dom } \varphi_d$, $w \in \mathbb{R}^m$, and $t > 0$ be arbitrary, then we obtain

$$\frac{\varphi_d(0 + th + \frac{1}{2}t^2 w) - \varphi_d(0) - t\varphi_d^{\downarrow}(0; h)}{\frac{1}{2}t^2} = \frac{\varphi_d(h + \frac{1}{2}tw) - \varphi_d(h)}{\frac{1}{2}t}.$$

Since φ_d is convex and $\varphi_d(h)$ is finite, the limit of the difference quotient on the right side of the latter equality exists, as $t \downarrow 0$, and is given by the directional derivative $\varphi_d'(h; w)$; see, e.g., Lemma 2.5.5. Hence, the function φ_d is twice directionally differentiable at 0 in the direction h. Using the epidifferentiability of φ_d and analogue arguments, we also see that φ_d is twice directionally epidifferentiable at 0 in the direction h. The second order derivatives are given by

$$\varphi_d^{\downarrow\downarrow}(0; h, w) = \varphi_d^{\downarrow}(h; w), \quad \varphi_d''(0; h, w) = \varphi_d'(h; w), \quad \forall\ w \in \mathbb{R}^m. \tag{5.3.6}$$

Next, let us consider arbitrary sequences $(t_k)_k$, $t_k \downarrow 0$, $(w^k)_k$, $(\tau^k)_k$ with $t_k(w^k, \tau^k) \to 0$, as $k \to \infty$, and let us suppose that the following inequality

$$\varphi_d(0 + t_k h + \tfrac{1}{2}t_k^2 w^k) \leq \varphi_d(0) + t_k \varphi_d^{\downarrow}(0; h) + \tfrac{1}{2}t_k^2 \tau^k = t_k \varphi_d(h) + \tfrac{1}{2}t_k^2 \tau^k,$$

is satisfied for all $k \in \mathbb{N}$. Then, the convexity of φ_d, our last computations, and the positive homogeneity of the directional epiderivative $\varphi_d^{\downarrow}(h; \cdot)$ imply

$$\tau^k \geq \frac{\varphi_d(h + \frac{1}{2}t_k w^k) - \varphi_d(h)}{\frac{1}{2}t_k} \geq \frac{\varphi_d^{\downarrow}(h; \frac{1}{2}t_k w^k)}{\frac{1}{2}t_k} = \varphi_d^{\downarrow}(h; w^k) = \varphi_d^{\downarrow\downarrow}(0; h, w^k).$$

Thus, φ_d is outer second order regular at 0 in all directions $h \in \text{dom } \varphi_d$. Using Robinson's constraint qualification we are able to transfer this properties to the initial function φ.

Lemma 5.3.5. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be C^2-decomposable at some point $x \in \text{dom } \varphi$ with decomposition pair (φ_d, F). Then, φ_d is twice directionally epidifferentiable and outer second order regular at $F(x)$ on* dom φ_d *and φ is twice directionally epidifferentiable and outer second order regular at x in all directions $h \in \mathbb{R}^n$ with $DF(x)h \in \text{dom } \varphi_d$.*

Proof. The first statement just summarizes our preceding calculations. Now, let $h \in \mathbb{R}^n$ with $DF(x)h \in \text{dom } \varphi_d$ be arbitrary. The C^2-decomposability implies that the function $F : \mathbb{R}^n \to \mathbb{R}^n$ is twice continuously differentiable and Robinson's constraint qualification

$$0 \in \text{int}\{F(x) + DF(x)\mathbb{R}^n - \text{dom } \varphi_d\}$$

is satisfied at x. Moreover, since φ_d is proper and we have $DF(x)h \in \text{dom } \varphi_d$, it follows from (5.3.5)

$$\varphi_d^{\downarrow}(F(x); DF(x)h) = \varphi_d^{\downarrow}(0; DF(x)h) = \varphi_d(DF(x)h) \in \mathbb{R}.$$

Consequently, Lemma 5.2.5 is applicable and, due to (5.3.6), we obtain

$$\begin{aligned}\varphi_-^{\downarrow\downarrow}(x; h, w) &= (\varphi_d \circ F)_-^{\downarrow\downarrow}(x; h, w) \\ &= \varphi_d^{\downarrow}(DF(x)h; DF(x)w + D^2F(x)[h, h]) = (\varphi_d \circ F)_+^{\downarrow\downarrow}(x; h, w) = \varphi_+^{\downarrow\downarrow}(x; h, w)\end{aligned}$$

for all $w \in \mathbb{R}^n$. Hence, φ is twice directionally epidifferentiable at x in the direction h. The outer second order regularity of φ follows from the regularity properties of φ_d and Lemma 5.2.7. □

In the following, we return to the discussion of our initial optimization problem $(\mathcal{P})$. Let us suppose, that $\bar{x} \in \operatorname{dom} \varphi$ is a stationary point of problem $(\mathcal{P})$ and let the function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be C^2-decomposable at $\bar{x}$ with decomposition pair (φ_d, F). Then, due to Robinson's constraint qualification or Remark 5.3.3, $\bar{x}$ is also a stationary point of the composite problem

$$\min_x \; f(x) + \varphi_d(F(x)) + \varphi(\bar{x}). \tag{5.3.7}$$

Since this problem is of the form $(\mathcal{P}_c)$ the theory developed in the sections 5.1 and 5.2 is applicable. In particular, Remark 5.2.9 implies

$$\xi^*_{\varphi_d,h}(\lambda) = \sup_w \langle \lambda, w \rangle - (\varphi_d)^{\downarrow\downarrow}_{-}(F(\bar{x}); DF(\bar{x})h, w) \leq 0,$$

for all $\lambda \in \mathcal{M}(\bar{x})$ and $h \in \mathcal{C}(\bar{x})$. In our specific situation, we have

$$\begin{aligned}\psi^{\downarrow}_{-}(\bar{x}; h) = \nabla f(\bar{x})^\top h + \varphi^{\downarrow}_{-}(\bar{x}; h) &= \nabla f(\bar{x})^\top h + \varphi^{\downarrow}_{d}(F(\bar{x}); DF(\bar{x})h) \\ &= \nabla f(\bar{x})^\top h + \varphi_d(DF(\bar{x})h),\end{aligned}$$

for all $h \in \mathbb{R}^n$, and the sets $\mathcal{M}(\bar{x})$ and $\mathcal{C}(\bar{x})$ are given by

$$\mathcal{C}(\bar{x}) = \{h : \nabla f(\bar{x})^\top h + \varphi_d(DF(\bar{x})h) = 0\}, \; \mathcal{M}(\bar{x}) = \{\lambda \in \partial\varphi_d(0) : \nabla f(\bar{x}) + DF(\bar{x})^\top \lambda = 0\}.$$

Moreover, using (5.3.6) and Lemma 2.5.5, it follows

$$(\varphi_d)^{\downarrow\downarrow}_{-}(F(\bar{x}); DF(\bar{x})h, 0) = \varphi^{\downarrow}_{d}(DF(\bar{x})h; 0) \leq \varphi'_d(DF(\bar{x})h; 0) = 0. \tag{5.3.8}$$

Hence, the decomposability of φ and the structure of φ_d imply that the additional curvature term in (5.2.3) and (5.2.4) vanishes, i.e.,

$$\xi^*_{\varphi_d,h}(\lambda) = 0, \quad \forall\, \lambda \in \mathcal{M}(\bar{x}), \quad \forall\, h \in \mathcal{C}(\bar{x}).$$

Together with our previous results, this observation allows to formulate no gap second order conditions for decomposable problems.

Theorem 5.3.6 (Second order conditions for decomposable problems). *Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. Let $\bar{x} \in \operatorname{dom} \varphi$ be given and assume that φ is C^2-decomposable at $\bar{x}$ with corresponding decomposition pair (φ_d, F). Then, the following statements do hold:*

(i) (Second order necessary conditions). *Suppose that $\bar{x}$ is a locally optimal solution of the initial problem $(\mathcal{P})$. Then, for any $h \in \mathcal{C}(\bar{x})$ the following inequality is satisfied:*

$$\max_{\lambda \in \mathcal{M}(\bar{x})} \left\{ h^\top \nabla^2 f(\bar{x}) h + \langle \lambda, D^2F(\bar{x})[h,h] \rangle \right\} \geq 0. \tag{5.3.9}$$

(ii) (Second order sufficient conditions). *Let $\bar{x}$ be a stationary point of the initial optimization problem $(\mathcal{P})$. Then, the quadratic growth condition,*

$$f(x) + \varphi(x) \geq f(\bar{x}) + \varphi(\bar{x}) + \alpha \|x - \bar{x}\|^2, \tag{5.3.10}$$

holds for some $\alpha > 0$ and all x in a neighborhood of $\bar{x}$ if and only if the following second order sufficient condition is fulfilled,

$$\max_{\lambda \in \mathcal{M}(\bar{x})} \left\{ h^\top \nabla^2 f(\bar{x}) h + \langle \lambda, D^2 F(\bar{x})[h, h] \rangle \right\} > 0, \quad \forall \, h \in \mathcal{C}(\bar{x}) \setminus \{0\}. \tag{5.3.11}$$

Therefore, in the latter case, $\bar{x}$ is also a (strict) locally optimal solution of the problem $(\mathcal{P})$. Moreover, if the function φ is C^2-fully decomposable at $\bar{x}$, then the second order sufficient condition (5.3.11) *additionally implies that $\bar{x}$ is an isolated stationary point of the initial problem $(\mathcal{P})$.*

Proof. The decomposability of φ implies that any (strict) local solution or (isolated) stationary point of problem $(\mathcal{P})$ is also a (strict) local solution or (isolated) stationary point of the composite problem (5.3.7) and vice versa. Since φ_d is twice epidifferentiable and outer second order regular at $F(\bar{x})$, in all directions $DF(\bar{x})h$, $h \in \mathcal{C}(\bar{x})$, and it holds

$$\xi^*_{\varphi_d, h}(\lambda) = 0, \quad \forall \, \lambda \in \mathcal{M}(\bar{x}), \quad \forall \, h \in \mathcal{C}(\bar{x}),$$

the second order conditions in Theorem 5.2.8 clearly reduce to the present conditions (i) and (ii). Now, suppose that φ is additionally C^2-fully decomposable at $\bar{x}$. Then, by Lemma 5.1.11, the set of Lagrange multipliers reduces to a singleton, i.e., $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$ and Theorem 5.1.12 is applicable. This readily shows that $\bar{x}$ is an isolated stationary point of (5.3.7) and of the initial problem $(\mathcal{P})$. □

Remark 5.3.7. Let us note that the discussion of the nondegeneracy condition in section 5.1 yields

$$\text{lin } N_{\partial \varphi_d(0)}(\bar{\lambda}) \subset \{y \in \mathbb{R}^m : \langle \bar{\lambda}, y \rangle = \varphi_d(y)\} = N_{\partial \varphi_d(0)}(\bar{\lambda}),$$

see, e.g., (5.1.12). On the other hand, in the decomposable setting, the set $\bar{\Phi}$, which was used to define the strict constraint qualification in Theorem 5.1.12, can be characterized as follows

$$\bar{\Phi} = \{y \in \mathbb{R}^m : \langle \bar{\lambda}, y - F(\bar{x}) \rangle = \varphi_d(y) - \varphi_d(F(\bar{x}))\} = \{y \in \mathbb{R}^m : \langle \bar{\lambda}, y \rangle = \varphi_d(y)\}.$$

Thus, if φ is C^2-fully decomposable at $\bar{x}$, then the strict constraint qualification is also satisfied at the (unique) multiplier $\bar{\lambda}$ and the mapping $\mathcal{M} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is upper Lipschitzian.

Remark 5.3.8. In Theorem 5.3.6, the function φ does not necessarily need to be convex. We have added this extra condition, since it is one of the basic and natural assumptions of our initial setting $(\mathcal{P})$. However, let us note that in the nonconvex case we cannot work with first order optimality conditions that are based on the proximity operator of φ. Thus, Lemma 4.1.2 is not applicable and stationarity of a feasible point $\bar{x} \in \text{dom } \varphi$ can only be characterized via

$$\psi^{\downarrow}_{-}(\bar{x}; d) = \nabla f(\bar{x})^\top d + \varphi^{\downarrow}_{-}(\bar{x}; d) \geq 0, \quad \forall \, d \in \mathbb{R}^n.$$

Since a more thorough treatment of the fully nonconvex case relies on an extension of Clarke's subdifferential for real extended valued functions and on certain regularity concepts, we will not go into detail here.

We conclude this subsection with an important computational result. In particular, we will show that the curvature term $-\xi^*_{\varphi,h}$ of the original function φ has an explicit representation and that, in the decomposable case, the second order conditions in Remark 5.2.11 and Theorem 5.3.6 actually coincide. Once more, this beautifully illustrates the fact that our second order results for the initial problem $(\mathcal{P})$ are independent of the respective decomposition pair (φ_d, F). Again, for convenience, we will assume that the standard conditions for problem $(\mathcal{P})$ are satisfied.

Lemma 5.3.9. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function and suppose that $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous mapping. Moreover, let $\bar{x} \in \operatorname{dom} \varphi$ be a stationary point of problem $(\mathcal{P})$ and let φ be C^2-decomposable at $\bar{x}$ with decomposition pair (φ_d, F). Then, for all $h \in \mathcal{C}(\bar{x})$, it follows*

$$-\xi^*_{\varphi,h}(-\nabla f(\bar{x})) = \max_{\lambda \in \mathcal{M}(\bar{x})} \langle \lambda, D^2F(\bar{x})[h,h] \rangle. \tag{5.3.12}$$

In addition, if φ is C^2-fully decomposable at $\bar{x}$, then the supremum that defines $\xi^*_{\varphi,h}(-\nabla f(\bar{x}))$ is attained at some $\hat{w} \in \mathbb{R}^n$, i.e., it holds

$$-\xi^*_{\varphi,h}(-\nabla f(\bar{x})) = \nabla f(\bar{x})^\top \hat{w} + \varphi^{\downarrow\downarrow}(\bar{x}; h, \hat{w}).$$

Proof. By Lemma 5.3.5, we know that φ is twice epidifferentiable at $\bar{x}$ in all directions $h \in \mathbb{R}^n$ with $DF(\bar{x})h \in \operatorname{dom} \varphi_d$. Now, let $h \in \mathcal{C}(\bar{x})$ be arbitrary, then it follows

$$\varphi_d(DF(\bar{x})h) = \varphi_d^{\downarrow}(F(\bar{x}); DF(\bar{x})h) = -\nabla f(\bar{x})^\top h \in \mathbb{R}.$$

Consequently, due to Lemma 5.2.5 and (5.3.6), we have

$$\begin{aligned} \varphi^{\downarrow\downarrow}(\bar{x}; h, w) = (\varphi_d \circ F)^{\downarrow\downarrow}(\bar{x}; h, w) &= \varphi_d^{\downarrow\downarrow}(F(\bar{x}); DF(\bar{x})h, DF(\bar{x})w + D^2F(\bar{x})[h,h]) \\ &= \varphi_d^{\downarrow}(DF(\bar{x})h; DF(\bar{x})w + D^2F(\bar{x})[h,h]). \end{aligned}$$

Next, Remark 5.2.9 and (5.3.8) imply

$$\varphi_d^{\downarrow}(DF(\bar{x})h; 0) = \varphi_d^{\downarrow\downarrow}(F(\bar{x}); DF(\bar{x})h, 0) = 0,$$

i.e., φ_d is subdifferentiable at $DF(\bar{x})h$. Now, let us set $\bar{w} := D^2F(\bar{x})[h,h]$, then the convex conjugate $-\xi^*_{\varphi,h}$ can be computed via

$$\begin{aligned} -\xi^*_{\varphi,h}(-\nabla f(\bar{x})) &= -\sup_{w} \, \langle w, -\nabla f(\bar{x}) \rangle - \varphi^{\downarrow\downarrow}(\bar{x}; h, w) \\ &= \inf_{w} \, \langle w, \nabla f(\bar{x}) \rangle + \varphi_d^{\downarrow}(DF(\bar{x})h; DF(\bar{x})w + \bar{w}). \end{aligned} \tag{5.3.13}$$

Here, since the directional epiderivative $\Pi(y) := \varphi_d^{\downarrow}(DF(\bar{x})h; y + \bar{w})$ is a convex, proper, and lower semicontinuous function, the latter problem can be dualized by applying the Fenchel-

Rockafellar duality framework for convex optimization problems. More specifically, by setting $\varrho(y) := \langle y, \nabla f(\bar{x})\rangle$, the dual problem of (5.3.13) is formally given by

$$\max_{v} \; -\varrho^*(DF(\bar{x})^\top v) - \Pi^*(-v),$$

see, e.g., [11, Chapter 15 and Definition 15.19]. Furthermore, due to Lemma 2.5.11 (ii), the convex conjugates ϱ^* and Π^* can be expressed as follows:

- $\varrho^*(DF(\bar{x})^\top v) = \sup_{y} \langle y, DF(\bar{x})^\top v - \nabla f(\bar{x})\rangle = \iota_{\{\nabla f(\bar{x})\}}(DF(\bar{x})^\top v)$.
- $\Pi^*(-v) = \sup_{y} \; -\langle y, v\rangle - \sigma_{\partial\varphi_d(DF(\bar{x})h)}(y + \bar{w}) = \langle \bar{w}, v\rangle + \sigma^*_{\partial\varphi_d(DF(\bar{x})h)}(-v)$
  $= \langle \bar{w}, v\rangle + \iota_{\partial\varphi_d(DF(\bar{x})h)}(-v)$.

Furthermore, by combing $\varphi_d(DF(\bar{x})h) = \sigma_{\partial\varphi_d(0)}(DF(\bar{x})h)$ and Example 2.5.17, it holds

$$\partial\varphi_d(DF(\bar{x})h) = \partial\sigma_{\partial\varphi_d(0)}(DF(\bar{x})h) = \{v \in \partial\varphi_d(0) : \varphi_d(DF(\bar{x})h) = \langle v, DF(\bar{x})h\rangle\}.$$

Thus, the dual problem can be rewritten as the following constrained problem

$$\max_{v} \; -\langle \bar{w}, v\rangle \quad \text{s. t.} \quad \begin{cases} -v \in \partial\varphi_d(0), \\ \nabla f(\bar{x}) - DF(\bar{x})^\top v = 0, \\ \nabla f(\bar{x})^\top h + \varphi_d(DF(\bar{x})h) = 0. \end{cases}$$

Moreover, since the first two constraints are equivalent to $-v \in \mathcal{M}(\bar{x})$ and the third condition is satisfied by any $h \in \mathcal{C}(\bar{x})$, we finally obtain

$$\max_{v} \; -\varrho^*(DF(\bar{x})^\top v) - \Pi^*(-v) = \max_{\lambda \in \mathcal{M}(\bar{x})} \langle \lambda, D^2F(\bar{x})[h, h]\rangle. \tag{5.3.14}$$

To finish the proof of the first part, it remains to be shown that there is no duality gap between the primal problem (5.3.13) and the dual problem (5.3.14). In particular, by [11, Theorem 15.23 and Proposition 15.24], this is the case when the following regularity condition is satisfied:

$$0 \in \operatorname{int}\{DF(\bar{x})\mathbb{R}^n - \operatorname{dom} \Pi\} = \operatorname{int}\{\bar{w} + DF(\bar{x})\mathbb{R}^n - \operatorname{dom} \varphi_d^\downarrow(DF(\bar{x})h; \cdot)\}. \tag{5.3.15}$$

We want to verify condition (5.3.15) by using Robinson's constraint qualification. First, since φ_d is convex and positively homogeneous, it follows

$$\varphi_d^\downarrow(DF(\bar{x})h; y) \le \varphi_d(DF(\bar{x})h + y) - \varphi_d(DF(\bar{x})h) \le \varphi_d(y).$$

and hence, we have $\operatorname{dom} \varphi_d \subset \operatorname{dom} \varphi_d^\downarrow(DF(\bar{x})h; \cdot)$. Now, Robinson's constraint qualification and Remark 5.3.2 imply

$$\mathbb{R}^m = DF(\bar{x})\mathbb{R}^n - \operatorname{dom} \varphi_d \subset DF(\bar{x})\mathbb{R}^n - \operatorname{dom} \varphi_d^\downarrow(DF(\bar{x})h; \cdot).$$

This easily establishes (5.3.15) and shows that problem (5.3.13) and (5.3.14) coincide and have the same optimal value. Next, let us suppose that the mapping φ is C^2-fully decomposable at $\bar{x}$. Then, the set of Lagrange multipliers reduces to a singleton $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$ and, obviously, the point $\bar{v} := -\bar{\lambda}$ is a solution of the dual problem (5.3.14). According to [11, Corollary 19.2], the (possibly empty) set of solutions of the primal problem (5.3.13) is given by

$$\partial\varrho^*(DF(\bar{x})^\top \bar{v}) \cap DF(\bar{x})^{-1}[\partial\Pi^*(-\bar{v})].$$

Moreover, by using the formulae for ϱ^*, Π^*, and $\partial\varphi_d(DF(\bar{x})h)$, Example 2.5.16, and the stationarity of $\bar{x}$, it follows:

- $\partial\varrho^*(DF(\bar{x})^\top \bar{v}) = N_{\{\nabla f(\bar{x})\}}(DF(\bar{x})^\top \bar{v}) = \{z : \langle z, \nabla f(\bar{x}) - DF(\bar{x})^\top \bar{v}\rangle \leq 0\} = \mathbb{R}^n$.
- $\partial\Pi^*(-\bar{v}) = \bar{w} + N_{\partial\varphi_d(DF(\bar{x})h)}(\bar{\lambda}) \supset \bar{w} + N_{\partial\varphi_d(0)}(\bar{\lambda})$.

The nondegeneracy condition implies that there exist $\hat{w} \in \mathbb{R}^n$, $-\hat{y} \in \text{lin } N_{\partial\varphi_d(0)}(\bar{\lambda})$ such that $\bar{w} = DF(\bar{x})\hat{w} - \hat{y}$. Thus, due to

$$DF(\bar{x})\hat{w} \in \bar{w} + \text{lin } N_{\partial\varphi_d(0)}(\bar{\lambda}) \subset \partial\Pi^*(-\bar{v}),$$

and [11, Corollary 19.2], the point $\hat{w}$ is a solution of the primal problem (5.3.13) and we can conclude the proof of the second part. □

Remark 5.3.10. Although the convexity of φ is not really necessary for the proof of the previous Lemma, it has an interesting, structural consequence. Particularly, in this situation, Lemma 5.2.12 is applicable and we can infer that the function

$$h \mapsto -\xi^*_{\varphi,h}(-\nabla f(\bar{x})) = \max_{\lambda \in \mathcal{M}(\bar{x})} \langle \lambda, D^2F(\bar{x})[h,h]\rangle$$

is convex on the critical cone $\mathcal{C}(\bar{x}) \subset \text{dom } \varphi^\downarrow(\bar{x};\cdot) = \{h : DF(\bar{x})h \in \text{dom } \varphi_d\}$. This clearly demonstrates that the curvature term $-\xi^*_{\varphi,h}$ captures basic second order properties of the nonsmooth function φ.

Remark 5.3.11. Let us note that a similar duality argument as in Lemma 5.3.9 is also used in the proof of the abstract second order conditions in Theorem 5.2.8. For more details we refer to the Appendix.

Computational results of the chain rule-type (5.3.12) have already been studied and established for different classes of functions and decomposition concepts. In particular, in [206, Theorem 4.5 and 4.7], Rockafellar showed that the curvature term in Lemma 5.3.9 is connected to another second order epigraphical framework, the so-called *second order subderivative*, and that similar chain rule-type formulae do exist for the class of fully amenable functions. Moreover, based on the results in [206], Rockafellar et al. have developed an extensive second order calculus for fully amenable functions, see, e.g., [193, 194] and [208, Theorem 13.14 and 13.67]. Extensions to the general infinite dimensional setting were studied by Cominetti in [56]. Finally, Mifflin and Sagastizábal, [150, 152], analyzed functions with a so-called primal-dual gradient (pdg)-structure. Under a special index set based regularity assumption they provide a profound calculus for functions with pdg-structure that resemble the result of Lemma 5.3.9. Since decomposable and pdg-structured functions are related, see, e.g., Hare [100, 101], Lemma 5.3.9 may turn out to be a special case of the pdg-theory. However, since the connection between decomposable and pdg-structured functions and a corresponding application of the results in [150, 152] are not apparent, we leave a further, more detailed investigation to future work.

5.3.3. Examples and calculus

In the following, we present various examples of decomposable functions and optimization problems that illustrate the generality and broad applicability of this concept. Let us note that a large part of our examples is motivated by similar examples addressed by Shapiro in [217, Section 2]. Moreover, let us also mention that, in the decomposable setting, the nondegeneracy condition has a much easier representation. In particular, in section 5.1, we have seen that the set lin $N_{\partial\varphi_d(0)}(\lambda)$, $\lambda \in \partial\varphi_d(0)$, is equivalent to the subspace lin $\varphi_d^{\downarrow}(0;\cdot)$. Consequently, due to (5.3.5), it holds

$$\text{lin } N_{\partial\varphi_d(0)}(\lambda) = \text{lin } \varphi_d^{\downarrow}(0;\cdot) = \text{lin } \varphi_d.$$

We start with a discussion of several nonsmooth optimization problems and their corresponding second order conditions.

Example 5.3.12 (ℓ_1-regularized minimization). Let us consider the ℓ_1-optimization problem

$$\min_{x\in\mathbb{R}^n} \ \psi(x) = f(x) + \mu\|x\|_1, \tag{5.3.16}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and $\mu > 0$ is a regularization parameter. In the following, we want to show that the weighted ℓ_1-norm $\varphi(x) = \mu\|x\|_1$ is C^∞-fully decomposable at any point $\bar{x} \in \mathbb{R}^n$. Therefore, let $\bar{x} \in \mathbb{R}^n$ be arbitrary and let us define the index sets $\mathcal{I}(\bar{x}) := \{i : \bar{x}_i \neq 0\}$ and $\mathcal{A}(\bar{x}) := \{i : \bar{x}_i = 0\}$. Then, the directional derivative of φ at $\bar{x}$ is given by

$$\begin{aligned}\varphi'(\bar{x};h) &= \sum_{i\in\mathcal{A}(\bar{x})} \lim_{t\downarrow 0} \frac{\mu|0+th_i| - 0}{t} + \sum_{i\in\mathcal{I}(\bar{x})} \lim_{t\downarrow 0} \frac{\mu|\bar{x}_i+th_i| - \mu|\bar{x}_i|}{t} \\ &= \mu\|h_{\mathcal{A}(\bar{x})}\|_1 + \mu\cdot\text{sign}(\bar{x})^\top h.\end{aligned}$$

Moreover, for all x in a sufficiently small neighborhood $U \subset \mathbb{R}^n$ of $\bar{x}$, it follows $\mathcal{I}(\bar{x}) \subseteq \mathcal{I}(x)$ and $\text{sign}(x_{\mathcal{I}(\bar{x})}) = \text{sign}(\bar{x}_{\mathcal{I}(\bar{x})})$. This implies

$$\mu\|x\|_1 = \mu\|x_{\mathcal{A}(\bar{x})}\|_1 + \mu\cdot\text{sign}(\bar{x})^\top x = \mu\|\bar{x}\|_1 + \varphi'(\bar{x}; x - \bar{x}) \tag{5.3.17}$$

for all $x \in U$. Consequently, let us consider the two functions $\varphi_d : \mathbb{R}^n \to \mathbb{R}$, $\varphi_d(y) := \varphi'(\bar{x}; y)$ and $F : \mathbb{R}^n \to \mathbb{R}^n$, $F(x) := x - \bar{x}$. Then, the pair (φ_d, F) satisfies the following properties:

- The function F is of class $C^\infty(\mathbb{R}^n)$ and it holds $F(\bar{x}) = 0$.
- φ_d is a convex, real valued, Lipschitz continuous, and positively homogeneous function.

- The derivative mapping $DF(\bar{x}) = I$ is obviously onto.

Thus, by (5.3.17) and Definition 5.3.1, (φ_d, F) is a decomposition pair of φ and we have verified that the ℓ_1-norm is C^∞-fully decomposable at any point $\bar{x} \in \mathbb{R}^n$.

Now, let us discuss the corresponding optimality conditions for problem (5.3.16) and let us suppose that $\bar{x} \in \mathbb{R}^n$ is an arbitrary, given stationary point, i.e., it holds $F^\Lambda(\bar{x}) = 0$ for some $\Lambda \in \mathbb{S}^n_{++}$. Then, the first order optimality conditions (5.1.3) take the following form

$$\begin{aligned} & \nabla f(\bar{x}) + \lambda = 0, \quad \lambda \in \partial\varphi_d(0) \\ \iff \quad & \nabla f(\bar{x})_{\mathcal{I}(\bar{x})} + \mu \cdot \operatorname{sign}(\bar{x}_{\mathcal{I}(\bar{x})}) = 0, \quad \nabla f(\bar{x})_{\mathcal{A}(\bar{x})} \in [-\mu, \mu]^{\bar{m}}, \quad \bar{m} := |\mathcal{A}(\bar{x})|, \end{aligned}$$

where we used $\partial|\cdot|(0) = [-1, 1]$. Furthermore, the critical cone $\mathcal{C}(\bar{x})$ associated with problem (5.3.16) is given by

$$\mathcal{C}(\bar{x}) = \{h : \nabla f(\bar{x})^\top h + \varphi_d(h) = 0\} = \{h : \langle \nabla f(\bar{x})_{\mathcal{A}(\bar{x})}, h_{\mathcal{A}(\bar{x})} \rangle + \mu \|h_{\mathcal{A}(\bar{x})}\|_1 = 0\}.$$

Due to

$$\nabla f(\bar{x})_i \cdot h_i + \mu |h_i| \geq (\mu - |\nabla f(\bar{x})_i|) \cdot |h_i| \geq 0, \quad \forall\, i \in \mathcal{A}(\bar{x}),$$

we can further simplify the critical cone and obtain the following final expression

$$\mathcal{C}(\bar{x}) = \{h \in \mathbb{R}^n : h_i = 0, \forall\, i \in \mathcal{A}_0(\bar{x}), \, h_i \in \mathbb{R}_- \cdot \nabla f(\bar{x})_i, \forall\, i \in \mathcal{A}_\pm(\bar{x})\},$$

where $\mathcal{A}_0(\bar{x}) := \{i \in \mathcal{A}(\bar{x}) : |\nabla f(\bar{x})_i| < \mu\}$ and $\mathcal{A}_\pm(\bar{x}) := \{i \in \mathcal{A}(\bar{x}) : |\nabla f(\bar{x})_i| = \mu\}$. Hence, by using the full decomposability of φ, $D^2F(\bar{x}) \equiv 0$, and Theorem 5.3.6, the second order sufficient conditions reduce to

$$h^\top \nabla^2 f(\bar{x}) h > 0, \quad \forall\, h \in \mathcal{C}(\bar{x}) \setminus \{0\}. \tag{5.3.18}$$

Moreover, any stationary point $\bar{x} \in \mathbb{R}^n$ that satisfies the second order conditions (5.3.18) is a (strict) locally optimal solution and, additionally, an isolated stationary point of the problem (5.3.16). On the other hand, any local solution $\bar{x}$ of the ℓ_1-problem (5.3.16) also has to fulfill the following second order necessary conditions

$$h^\top \nabla^2 f(\bar{x}) h \geq 0, \quad \forall\, h \in \mathcal{C}(\bar{x}).$$

Let us note that second order conditions of this form were already studied by Casas, Herzog, and Wachsmuth, [40], in an infinite dimensional, optimal control setting. In [98, 254, 157], the *strong second order sufficient condition*

$$h^\top \nabla^2 f(\bar{x}) h > 0, \quad \forall\, h \in \operatorname{aff}\, \mathcal{C}(\bar{x}) \setminus \{0\},$$

was used to analyze local convergence properties of ℓ_1-minimization algorithms. In particular, Milzarek and Ulbrich, [157], showed that the strong second order conditions guarantee isolated stationarity and invertibility of the generalized derivatives of $F^\Lambda(\bar{x})$. Similar invertibility results were also established in [97, 225]. Finally, let us mention, that the strong second order sufficient condition is equivalent to

$$\lambda_{\min}(\nabla^2 f(\bar{x})_{[\mathcal{E}\mathcal{E}]}) > 0, \quad \mathcal{E} := \mathcal{A}_{\pm}(\bar{x}) \cup \mathcal{I}(\bar{x}).$$

Example 5.3.13 (Group sparse problems). We consider the following optimization problem with a group sparse penalty term

$$\min_x \ \psi(x) = f(x) + \sum_{i=1}^{s} \ \omega_i \|x_{g_i}\|_2. \tag{5.3.19}$$

Here, the index sets g_i, $i = 1, ..., s$, form a disjoint partitioning of the set $\{1, ..., n\}$ and the parameters ω_i, $i = 1, ..., s$, are supposed to be positive. Again, we want to show that the nonsmooth function $\varphi(x) := \sum_{i=1}^{s} \omega_i \|x_{g_i}\|_2$ is fully decomposable at any point $\bar{x} \in \mathbb{R}^n$. As in the last example, let $\bar{x}$ be arbitrary and let us define the index sets $\mathcal{I}(\bar{x}) := \{i : \bar{x}_{g_i} \neq 0\}$, $\mathcal{A}(\bar{x}) := \{i : \bar{x}_{g_i} = 0\}$. Moreover, let us set

$$\mathcal{G}_{\mathcal{A}} := \bigcup_{i \in \mathcal{A}(\bar{x})} g_i, \quad m := |\mathcal{G}_{\mathcal{A}}| = \sum_{i \in \mathcal{A}(\bar{x})} |g_i|, \quad \bar{m} := |\mathcal{A}(\bar{x})|,$$

and suppose that $i_1, ..., i_m$ and $j_1, ..., j_{\bar{m}}$ denote the different elements in $\mathcal{G}_{\mathcal{A}}$ and $\mathcal{A}(\bar{x})$, respectively. For our further analysis, we will also need the following (one-to-one) relabeling of the active groups g_i, $i \in \mathcal{A}(\bar{x})$,

$$q_\ell \subset \{1, ..., m\}, \quad q_\ell = \{k \in \{1, ..., m\} : i_k \in g_{j_\ell}\}, \quad \ell \in \{1, ..., \bar{m}\}.$$

Now, we can define the decomposition functions $\varphi_d : \mathbb{R}^{m+1} \to \mathbb{R}$ and $F : \mathbb{R}^n \to \mathbb{R}^{m+1}$ via

$$\varphi_d(t, y) := t + \sum_{\ell=1}^{\bar{m}} \omega_{j_\ell} \|y_{q_\ell}\|_2, \quad F(x) := \begin{pmatrix} \sum_{i \in \mathcal{I}(\bar{x})} \omega_i(\|x_{g_i}\|_2 - \|\bar{x}_{g_i}\|_2) \\ I_{[\mathcal{G}_{\mathcal{A}}\cdot]} \cdot x \end{pmatrix}. \tag{5.3.20}$$

Clearly, we have $\varphi(x) = \varphi(\bar{x}) + \varphi_d(F(x))$ for all $x \in \mathbb{R}^n$ and, as in the ℓ_1-norm example, it follows $\mathcal{I}(\bar{x}) \subset \mathcal{I}(x)$ for all x in a certain neighborhood U of $\bar{x}$. Consequently, the pair (φ_d, F) satisfies the following properties:

- The function F is of class $C^\infty(U)$ and it holds $F(\bar{x}) = 0$.
- φ_d is a convex, real valued, Lipschitz continuous, and positively homogeneous function.
- The derivative mapping $DF(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^{m+1}$ is given by

$$DF(\bar{x}) = \begin{pmatrix} \nabla F_1(\bar{x})^\top \\ I_{[\mathcal{G}_{\mathcal{A}}\cdot]} \end{pmatrix}, \quad [\nabla F_1(\bar{x})]_{g_i} = \begin{cases} \omega_i \frac{\bar{x}_{g_i}}{\|\bar{x}_{g_i}\|_2} & \text{if } i \in \mathcal{I}(\bar{x}), \\ 0 & \text{otherwise.} \end{cases}$$

  It is easy to see that the rows of $DF(\bar{x})$ are pairwise orthogonal, i.e., $DF(\bar{x})$ has full row rank and is onto.

Hence, (φ_d, F) is a decomposition pair of φ and the group sparse penalty term is C^∞-fully decomposable at any point $\bar{x} \in \mathbb{R}^n$. We want to point out that, besides the somewhat technical relabeling of the active groups, this construction is straightforward. Since the function

φ is only nonsmooth at groups with value zero, the overall idea is to split the functional φ into its smooth and nonsmooth parts. This immediately leads to a decomposition of the form (5.3.20).

Now, let us consider an arbitrary stationary point $\bar{x}$ of problem (5.3.19). The corresponding first order optimality conditions (5.1.3) take the following form

$$\begin{aligned} & \nabla f(\bar{x}) + \nabla F_1(\bar{x})\gamma + I_{[\mathcal{G}_{\mathcal{A}}\cdot]}^\top \lambda = 0, \quad (\gamma, \lambda) \in \partial\varphi_d(0,0) \\ \iff \quad & \nabla f(\bar{x})_{g_i} + \omega_i \frac{\bar{x}_{g_i}}{\|\bar{x}_{g_i}\|} = 0, \ \forall\ i \in \mathcal{I}(\bar{x}), \quad \nabla f(\bar{x})_{g_i} \in \bar{B}_{\omega_i}(0), \ \forall\ i \in \mathcal{A}(\bar{x}), \end{aligned}$$

where we used $\partial\|\cdot\|_2(0) = \bar{B}_1(0)$ and $\partial\varphi_d(0,0) = \{1\} \times \prod_{i\in\mathcal{A}(\bar{x})} \bar{B}_{\omega_i}(0)$. The critical cone is given by

$$\begin{aligned} \mathcal{C}(\bar{x}) &= \{h : \nabla f(\bar{x})^\top h + \varphi_d(DF(\bar{x})h) = 0\} \\ &= \{h : \textstyle\sum_{i=1}^s \nabla f(\bar{x})_{g_i}^\top h_{g_i} + \sum_{i\in\mathcal{A}(\bar{x})} \omega_i \|h_{g_i}\|_2 + \nabla F_1(\bar{x})^\top h = 0\} \\ &= \{h : \nabla f(\bar{x})_{g_i}^\top h_{g_i} + \omega_i \|h_{g_i}\|_2 = 0, \ \forall\ i \in \mathcal{A}(\bar{x})\}. \end{aligned}$$

Next, let us introduce the index sets $\mathcal{A}_0(\bar{x}) := \{i \in \mathcal{A}(\bar{x}) : \|\nabla f(\bar{x})_{g_i}\|_2 < \omega_i\}$ and $\mathcal{A}_\pm(\bar{x}) := \{i \in \mathcal{A}(\bar{x}) : \|\nabla f(\bar{x})_{g_i}\|_2 = \omega_i\}$. Then, for all $i \in \mathcal{A}_0(\bar{x})$, it follows

$$0 = \nabla f(\bar{x})_{g_i}^\top h_{g_i} + \omega_i \|h_{g_i}\|_2 \geq (\omega_i - \|\nabla f(\bar{x})_{g_i}\|_2) \cdot \|h_{g_i}\|_2 \geq 0 \quad \iff \quad h_{g_i} = 0$$

and for $i \in \mathcal{A}_\pm(\bar{x})$, we obtain

$$\nabla f(\bar{x})_{g_i}^\top h_{g_i} + \|\nabla f(\bar{x})_{g_i}\| \|h_{g_i}\|_2 = 0 \quad \iff \quad h_{g_i} \in \mathbb{R}_-\cdot \nabla f(\bar{x})_{g_i}.$$

Thus, the critical cone can be represented as follows:

$$\mathcal{C}(\bar{x}) = \{h \in \mathbb{R}^n : h_{g_i} = 0, \forall\ i \in \mathcal{A}_0(\bar{x}),\ h_{g_i} \in \mathbb{R}_-\cdot \nabla f(\bar{x})_{g_i}, \forall\ i \in \mathcal{A}_\pm(\bar{x})\}.$$

Finally, for all $h \in \mathbb{R}^n$, we have $\langle(\gamma,\lambda), D^2F(\bar{x})[h,h]\rangle = h^\top \nabla^2 F_1(\bar{x}) h$. Here, the Hessian $\nabla^2 F_1(\bar{x})$ is a block-structured matrix that captures the curvature of the smooth part of φ; it is given by

$$\nabla^2 F_1(\bar{x})_{[g_i g_j]} = \begin{cases} \frac{\omega_i}{\|\bar{x}_{g_i}\|_2} I - \frac{\omega_i}{\|\bar{x}_{g_i}\|_2^3} \bar{x}_{g_i} \bar{x}_{g_i}^\top & \text{if } i = j \in \mathcal{I}(\bar{x}), \\ 0 & \text{otherwise.} \end{cases}$$

Consequently, by Theorem 5.3.6 and the full decomposability of φ, the second order sufficient conditions

$$h^\top \nabla^2 f(\bar{x}) h + h^\top \nabla^2 F_1(\bar{x}) h > 0, \quad \forall\ h \in \mathcal{C}(\bar{x}) \setminus \{0\}$$

guarantee that every stationary point of the group sparse problem (5.3.19) is also a (strict) local minimizer and an isolated stationary point. Moreover, any local solution of (5.3.19) must satisfy the corresponding second order necessary conditions (5.3.9). Let us remark that similar second order conditions were investigated by Casas, Herzog, Stadler and Wachsmuth, [106, 39], in an infinite dimensional, directionally sparse framework.

Example 5.3.14 (Total variation). We consider the total variation-regularized optimization problem

$$\min_x \; f(x) + \mu \sum_{i=1}^{m} \|D_{|i}x\|_2, \tag{5.3.21}$$

where $\mu > 0$ is a parameter and $D = (D_{|1}^\top, ..., D_{|m}^\top)^\top \in \mathbb{R}^{2m\times n}$, $D_{|i} \in \mathbb{R}^{2\times n}$, $i = 1, ..., m$, is a discrete gradient operator using forward differences and periodic or Neumann boundary conditions. Here, the subscript index "$|i$" is used to denote two dimensional objects of the form

$$x_{|i} := \begin{pmatrix} x_{2i-1} \\ x_{2i} \end{pmatrix} \in \mathbb{R}^2 \quad \text{and} \quad D_{|i} := \begin{pmatrix} D_{[2i-1,\cdot]} \\ D_{[2i,\cdot]} \end{pmatrix} \in \mathbb{R}^{2\times n},$$

and to simplify the notation. In the following, we will analyze the decomposability properties of this problem and of the total variation semi-norm $\varphi(x) := \mu \sum_{i=1}^m \|D_{|i}x\|_2$. Since this example has a similar structure as the group sparsity problem, we can reuse the basic constructions of Example 5.3.13. We define $\mathcal{A}(\bar{x}) := \{i : D_{|i}x = 0\}$, $\mathcal{I}(\bar{x}) := \{i : D_{|i}x \neq 0\}$, $\bar{m} := |\mathcal{A}(\bar{x})|$ and

$$\varphi_d(t, y) := t + \mu \sum_{i=1}^{\bar{m}} \|y_{|i}\|_2, \quad F(x) := \begin{pmatrix} \mu \sum_{i\in\mathcal{I}(\bar{x})}(\|D_{|i}x\|_2 - \|D_{|i}\bar{x}\|_2) \\ D_{|\mathcal{A}(\bar{x})} \cdot x \end{pmatrix}.$$

Again, it holds $\varphi(x) = \varphi(\bar{x}) + \varphi_d(F(x))$ for all $x \in \mathbb{R}^n$ and it follows $\mathcal{I}(\bar{x}) \subset \mathcal{I}(x)$ for all x in a small neighborhood U of $\bar{x}$. Accordingly, the pair (φ_d, F) has the following properties:

- F is of class $C^\infty(U)$ and it holds $F(\bar{x}) = 0$.
- φ_d is a convex, real valued, Lipschitz continuous, and positively homogeneous mapping.
- Due to $\operatorname{dom} \varphi_d = \mathbb{R}^{2\bar{m}+1}$, Robinson's constraint qualification is always satisfied. The derivative mapping $DF(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^{2\bar{m}+1}$ is given by

$$DF(\bar{x}) = \begin{pmatrix} \nabla F_1(\bar{x})^\top \\ D_{|\mathcal{A}(\bar{x})} \end{pmatrix}, \quad \nabla F_1(\bar{x}) = \sum_{i\in\mathcal{I}(\bar{x})} \frac{\mu}{\|D_{|i}\bar{x}\|_2} D_{|i}^\top D_{|i}\bar{x} = D^\top \Xi D\bar{x},$$

where $\Xi \in \mathbb{R}^{2m\times 2m}$ is a (block)-diagonal matrix and it holds $\Xi = \operatorname{blockdiag}(\Xi^1, ..., \Xi^m)$ with

$$\Xi^i \in \mathbb{R}^{2\times 2}, \qquad \Xi^i = \begin{cases} \frac{\mu}{\|D_{|i}\bar{x}\|_2} I & \text{if } i \in \mathcal{I}(\bar{x}), \\ 0 & \text{if } i \in \mathcal{A}(\bar{x}). \end{cases}$$

Thus, φ is C^∞-decomposable at any point $\bar{x}$ with decomposition pair (φ_d, F). Furthermore, if the nondegeneracy condition,

$$\mathbb{R} \times \mathbb{R}^{2\bar{m}} = DF(\bar{x})\mathbb{R}^n - \operatorname{lin} \varphi_d = \begin{pmatrix} \nabla F_1(\bar{x})^\top \\ D_{|\mathcal{A}(\bar{x})} \end{pmatrix} \mathbb{R}^n - \mathbb{R} \times \{0\} \quad \Longleftrightarrow \quad D_{|\mathcal{A}(\bar{x})}\mathbb{R}^n = \mathbb{R}^{2\bar{m}},$$

is fulfilled, i.e., if the matrix $D_{|\mathcal{A}(\bar{x})}$ is onto, then φ is C^∞-fully decomposable at $\bar{x}$. Let us note that in applications this additional requirement can be somewhat restrictive and

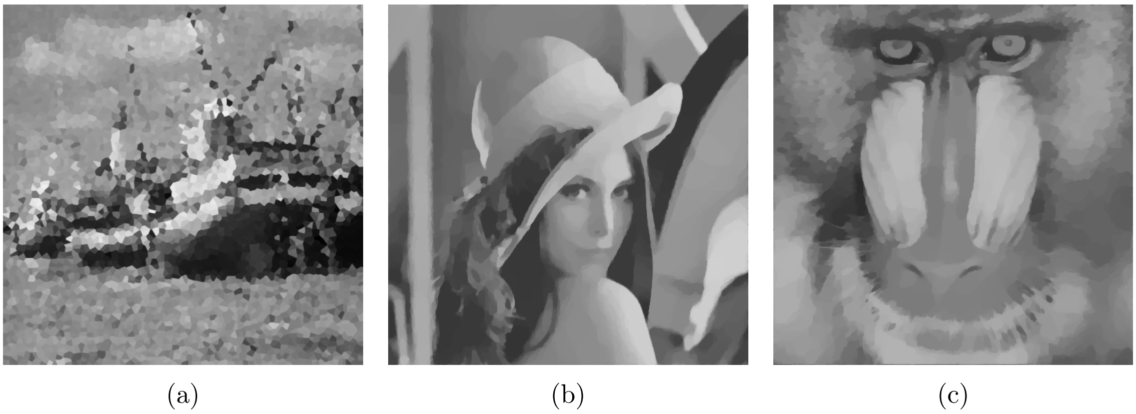


Figure 5.1.: Illustration of the failure of the nondegeneracy condition for total variation imaging. In subfigure (a), a pixelated version of the image `boat.png` is shown that does not satisfy the nondegeneracy condition. In subfigure (b) and (c) (approximate) solutions of the image reconstruction problem

$$\min_x \ f(x) + \mu \sum_{i=1}^{n^2} \|D_{|i}x\|_2, \quad f(x) = \frac{1}{2}\|x - b\|_2^2, \quad \mu = \frac{1}{4}$$

are presented. In both examples, we tried to restore noisy versions b of the images `lena.png` and `mandrill.tiff` by solving the latter minimization problem. (More specifically, we added Gaussian noise with zero mean and variance $\sigma = 0.1$). The respective reconstructions $\bar{x}$, which are shown in subfigure (b) and (c), do not fulfill the nondegeneracy condition. However, in this situation, since the function f is strongly convex, the second order conditions (5.3.22) are satisfied and, consequently, the shown images are strict local minima and isolated stationary points of the above nonsmooth optimization problem.

cannot be expected to be satisfied in general; an exemplary discussion of this problem and a corresponding illustration for image reconstruction problems is provided in Figure 5.1.

Now, let $\bar{x} \in \mathbb{R}^n$ be a stationary point of problem (5.3.21). The corresponding first order optimality conditions are given by

$$\begin{aligned} & \nabla f(\bar{x}) + \nabla F_1(\bar{x})\gamma + D_{|\mathcal{A}(\bar{x})}^\top \lambda = 0, \quad (\gamma, \lambda) \in \partial\varphi_d(0,0) \\ \iff \quad & \nabla f(\bar{x}) + D^\top \Xi D\bar{x} + D_{|\mathcal{A}(\bar{x})}^\top \lambda = 0, \quad \lambda_{|i} \in \bar{B}_\mu(0), \ \forall \ i = 1, ..., \bar{m}, \end{aligned}$$

where we used $\partial\varphi_d(0,0) = \{1\} \times \prod_{i=1}^{\bar{m}} \bar{B}_\mu(0)$. Next, let $i_1, ..., i_{\bar{m}}$ denote the elements of the set $\mathcal{A}(\bar{x})$. For $i \in \mathcal{A}(\bar{x})$, we define the indices k_i as follows

$$k_i := j \quad :\iff \quad \text{the index } i \text{ is the } i_j\text{-th element of } \mathcal{A}(\bar{x}).$$

Let $(1, \bar{\lambda}) \in \mathcal{M}(\bar{x})$ be an arbitrary, but fixed Lagrange multiplier. Then, it holds

$$\begin{aligned} 0 &= \nabla f(\bar{x})^\top h + \varphi_d(DF(\bar{x})h) = \nabla f(\bar{x})^\top h + \nabla F_1(\bar{x})^\top h + \mu \sum_{i \in \mathcal{A}(\bar{x})} \|D_{|i} h\|_2 \\ &= \mu \sum_{i \in \mathcal{A}(\bar{x})} \|D_{|i} h\|_2 - \langle D_{|\mathcal{A}(\bar{x})} h, \bar{\lambda} \rangle = \sum_{i \in \mathcal{A}(\bar{x})} \left\{ \mu \|D_{|i} h\|_2 - \langle D_{|i} h, \bar{\lambda}_{|k_i} \rangle \right\}. \end{aligned}$$

Due to $\bar{\lambda}_{|k_i} \in \bar{B}_\mu(0)$, $i \in \mathcal{A}(\bar{x})$, and by repeating the argumentation of Example 5.3.13, we obtain the following, final representation of the critical cone

$$\mathcal{C}(\bar{x}) = \bigcap_{(1,\lambda) \in \mathcal{M}(\bar{x})} \{h \in \mathbb{R}^n : D_{|i} h = 0, \, \forall \, i \in \mathcal{A}_0(\bar{x}, \lambda), \, D_{|i} h \in \mathbb{R}_- \cdot \lambda_{|k_i}, \, \forall \, i \in \mathcal{A}_\pm(\bar{x}, \lambda)\},$$

where $\mathcal{A}_0(\bar{x}, \lambda) := \{i \in \mathcal{A}(\bar{x}) : \|\lambda_{|k_i}\|_2 < \mu\}$ and $\mathcal{A}_\pm(\bar{x}, \lambda) := \{i \in \mathcal{A}(\bar{x}) : \|\lambda_{|k_i}\|_2 = \mu\}$. (Let us recall that the critical cone does not depend on the choice of the multiplier λ, see Definition 5.1.5). Finally, for all $h \in \mathbb{R}^n$, we have

$$\langle (\gamma, \lambda), D^2 F(\bar{x})[h, h] \rangle = h^\top \nabla^2 F_1(\bar{x}) h,$$

i.e., the curvature term in the second order conditions does not depend on any specific Lagrange multiplier $(\gamma, \lambda) \in \mathcal{M}(\bar{x})$. Moreover, the Hessian $\nabla^2 F_1(\bar{x})$ can be computed explicitly; it holds $\nabla^2 F_1(\bar{x}) = D^\top \hat{\Xi} D$,

$$\hat{\Xi} = \operatorname{blockdiag}(\hat{\Xi}^1, ..., \hat{\Xi}^m), \quad \hat{\Xi}^i = \begin{cases} \Xi^i - \frac{\mu}{\|D_{|i}\bar{x}\|_2^3} (D_{|i}\bar{x})(D_{|i}\bar{x})^\top & \text{if } i \in \mathcal{I}(\bar{x}), \\ 0 & \text{if } i \in \mathcal{A}(\bar{x}). \end{cases}$$

Hence, the second order necessary and sufficient conditions for the total variation problem (5.3.21) take the forms

$$h^\top \nabla^2 f(\bar{x}) h + (Dh)^\top \hat{\Xi} D h \geq 0, \quad \forall \, h \in \mathcal{C}(\bar{x}),$$

and

$$h^\top \nabla^2 f(\bar{x}) h + (Dh)^\top \hat{\Xi} D h > 0, \quad \forall \, h \in \mathcal{C}(\bar{x}) \setminus \{0\}, \tag{5.3.22}$$

respectively. In addition, if the matrix $D_{|\mathcal{A}(\bar{x})}$ is onto and if the second order conditions (5.3.22) are satisfied, then Theorem 5.3.6 (ii) implies that $\bar{x}$ is an isolated stationary point of problem (5.3.21). However, since this additional condition is likely to be violated in practice, the function φ does not need to be fully decomposable at the stationary point $\bar{x}$. Nonetheless, in this specific situation, since the second order sufficient conditions are independent of the choice of the multiplier $(\gamma, \lambda) = (1, \lambda) \in \mathcal{M}(\bar{x})$ and φ_d is real valued, we can apply Corollary 5.1.14. Consequently, the second order conditions (5.3.22) imply that $\bar{x}$ is an isolated stationary point of problem (5.3.21), no matter whether φ is fully decomposable or not.

Although the (possible) nonuniqueness of the multiplier λ has no influence on the second order conditions, it can cause certain numerical difficulties. This was already observed by Dong, Hintermüller, and Neri, [65], who studied a semismooth Newton method for a primal-dual interpretation of the KKT conditions (5.1.3) of the problem (5.3.21) in an ℓ_1-setting.

Example 5.3.15 (Nonlinear programming). In this example, we want to show that nonlinear optimization problems of the form

$$\min_x \ f(x) \quad \text{s. t.} \quad g(x) \le 0, \quad h(x) = 0, \tag{5.3.23}$$

are decomposable at any feasible point $\bar{x}$, where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, and $h : \mathbb{R}^n \to \mathbb{R}^p$ are supposed to be twice continuously differentiable. Let us set $\varphi(x) := \iota_{\mathbb{R}^m_-}(g(x)) + \iota_{\{0\}}(h(x))$, $\mathcal{I}(\bar{x}) := \{i : g_i(\bar{x}) < 0\}$, $\mathcal{A}(\bar{x}) := \{i : g_i(\bar{x}) = 0\}$, and $\bar{m} := |\mathcal{A}(\bar{x})|$. Then, the constrained problem (5.3.23) can be rewritten in our basic form

$$\min_x \ \psi(x) = f(x) + \varphi(x)$$

and, due to $\varphi(\bar{x}) = 0$, it holds

$$\varphi(x) = \varphi(\bar{x}) + \iota_{\mathbb{R}^{\bar{m}}_-}(g_{\mathcal{A}(\bar{x})}(x)) + \iota_{\{0\}}(h(x)),$$

for all x in a neighborhood U of $\bar{x}$. Now, let us define the decomposition functions $\varphi_d : \mathbb{R}^{\bar{m}} \times \mathbb{R}^p \to (-\infty, +\infty]$ and $F : \mathbb{R}^n \to \mathbb{R}^{\bar{m}} \times \mathbb{R}^p$ as follows

$$\varphi_d(y, z) := \iota_{\mathbb{R}^{\bar{m}}_-}(y) + \iota_{\{0\}}(z), \quad F(x) := \begin{pmatrix} g_{\mathcal{A}(\bar{x})}(x) \\ h(x) \end{pmatrix}.$$

The pair (φ_d, F) satisfies the following properties:

- The function F is twice continuously differentiable on $\mathbb{R}^n$ and it holds $F(\bar{x}) = 0$.
- φ_d is a convex, proper, lower semicontinuous, and positively homogeneous mapping.

Moreover, in this situation, Robinson's constraint qualification (5.3.2) is given by

$$\mathbb{R}^{\bar{m}} \times \mathbb{R}^p = DF(\bar{x})\mathbb{R}^n - \operatorname{dom} \varphi_d = \begin{pmatrix} \nabla g_{\mathcal{A}(\bar{x})}(\bar{x})^\top \\ \nabla h(\bar{x})^\top \end{pmatrix} \mathbb{R}^n - \mathbb{R}^{\bar{m}}_- \times \{0\}.$$

Furthermore, let us note that the latter condition is actually equivalent to the well-known *Mangasarian-Fromovitz constraint qualification* (MFCQ) for nonlinear programs:

$$\nabla h(\bar{x}) \text{ has full column rank}, \quad \exists\ d \in \mathbb{R}^n \text{ such that } \nabla g_{\mathcal{A}(\bar{x})}(\bar{x})^\top d < 0.$$

On the other hand, due to $\operatorname{lin} \varphi_d = \{0\} \times \{0\}$, the nondegeneracy condition immediately reduces to the condition

$$\begin{pmatrix} \nabla g_{\mathcal{A}(\bar{x})}(\bar{x})^\top \\ \nabla h(\bar{x})^\top \end{pmatrix} \mathbb{R}^n = \begin{pmatrix} \mathbb{R}^{\bar{m}} \\ \mathbb{R}^p \end{pmatrix}.$$

Thus, the point $\bar{x}$ is nondegenerate if and only if the matrix $(\nabla g_{\mathcal{A}(\bar{x})}(\bar{x}), \nabla h(\bar{x}))$ has full column rank. Consequently, if the MFCQ is satisfied at $\bar{x}$, then φ is C^2-decomposable at $\bar{x}$ with decomposition pair (φ_d, F). Additionally, if the *Linear Independency constraint qualification* (LICQ) holds at $\bar{x}$, then φ is C^2-fully decomposable.

Next, let $\bar{x} \in \operatorname{dom} \varphi$ be a stationary point of problem (5.3.23) (in the sense of Definition 5.1.1) and suppose that the MFCQ holds at $\bar{x}$. Then, the first order optimality conditions take the following form

$$\nabla f(\bar{x}) + \nabla g_{\mathcal{A}(\bar{x})}(\bar{x})\lambda + \nabla h(\bar{x})\mu = 0, \ \ (\lambda, \mu) \in \partial\varphi_d(0,0) = N_{\mathbb{R}^{\bar{m}}_-}(0) \times N_{\{0\}}(0) = \mathbb{R}^{\bar{m}}_+ \times \mathbb{R}^p$$

and the critical cone is given by

$$\begin{aligned} \mathcal{C}(\bar{x}) &= \{d \in \mathbb{R}^n : \nabla f(\bar{x})^\top d + \varphi_d(DF(\bar{x})d) = 0\} \\ &= \{d \in \mathbb{R}^n : \nabla f(\bar{x})^\top d = 0, \ \nabla g_i(\bar{x})^\top d \le 0 \ \forall \ i \in \mathcal{A}(\bar{x}), \ \nabla h(\bar{x})^\top d = 0\}. \end{aligned}$$

By Theorem 5.3.6, it follows that the condition

$$\begin{aligned} &\max_{(\lambda,\mu)\in\mathcal{M}(\bar{x})} \ d^\top \nabla^2 f(\bar{x})d + \sum_{i\in\mathcal{A}(\bar{x})} \lambda_i d^\top \nabla^2 g_i(\bar{x})d + \sum_{i=1}^{p} \mu_i d^\top \nabla^2 h_i(\bar{x})d = \\ &\max_{(\lambda,\mu)\in\mathcal{M}(\bar{x})} \ d^\top \nabla^2_{xx} L_r(\bar{x}, \lambda, \mu)d \ > \ 0, \quad \forall \ d \in \mathcal{C}(\bar{x}) \setminus \{0\} \end{aligned} \tag{5.3.24}$$

ensures local optimality of the stationary point $\bar{x}$. Moreover, if the LICQ is satisfied at $\bar{x}$, i.e., if φ is fully decomposable at $\bar{x}$, then the set of Lagrange multipliers $\mathcal{M}(\bar{x})$ reduces to a singleton and $\bar{x}$ is an isolated stationary point of (5.3.23). Here, $L_r : \mathbb{R}^n \times \mathbb{R}^{\bar{m}} \times \mathbb{R}^p \to \mathbb{R}$, $L_r(x, \lambda, \mu) := f(x) + \langle \lambda, g_{\mathcal{A}(\bar{x})}(x)\rangle + \langle \mu, h(x)\rangle$ denotes the reduced Lagrangian associated with problem (5.3.23). As usual, the decomposability of φ also implies that any local solution of the nonlinear program (5.3.23) has to satisfy the corresponding second order necessary conditions

$$\max_{(\lambda,\mu)\in\mathcal{M}(\bar{x})} \ d^\top \nabla^2_{xx} L_r(\bar{x}, \lambda, \mu)d \ \ge \ 0, \quad \forall \ d \in \mathcal{C}(\bar{x}). \tag{5.3.25}$$

Of course, this pair of second order conditions is well-known in nonlinear optimization. Here, in our case, the conditions (5.3.24) and (5.3.25) emerge as a special example of the general second order theory for decomposable functions and problems.

In summary, the latter examples demonstrate the diversity and wide applicability of decomposable functions in the context of nonsmooth optimization problems. Next, based on various results and examples in Shapiro [217, Section 2], we will study several, general classes of decomposable functions.

The following, first example was presented in [217, Example 2.1].

Example 5.3.16 (Max-type functions). Let $\varphi_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, ..., m$, be a family of twice continuously differentiable functions and let us consider

$$\varphi : \mathbb{R}^n \to \mathbb{R}, \quad \varphi(x) := \max_{1\le i\le m} \ \varphi_i(x).$$

Let $\bar{x} \in \mathbb{R}^n$ be arbitrary and let us define $\mathcal{A}(\bar{x}) := \{i : \varphi_i(\bar{x}) = \varphi(\bar{x})\}$. Then, a continuity argument implies $\mathcal{A}(x) \subset \mathcal{A}(\bar{x})$ for all x in a sufficiently small neighborhood U of $\bar{x}$ and, consequently, it holds

$$\varphi(x) = \varphi(\bar{x}) + \max_{i \in \mathcal{A}(\bar{x})} \{\varphi_i(x) - \varphi(\bar{x})\}, \quad \forall\, x \in U. \tag{5.3.26}$$

Let us set $\bar{m} := |\mathcal{A}(\bar{x})|$ and suppose that $i_1, ..., i_{\bar{m}}$ are the different elements of the index set $\mathcal{A}(\bar{x})$. Then, the functions $\varphi_d : \mathbb{R}^{\bar{m}} \to \mathbb{R}$ and $F : \mathbb{R}^n \to \mathbb{R}^{\bar{m}}$,

$$\varphi_d(y) := \max\{y_1, ..., y_{\bar{m}}\}, \quad F(x) = \varphi_{\mathcal{A}(\bar{x})}(x) - \varphi(\bar{x}) \cdot \mathbb{1} = \begin{pmatrix} \varphi_{i_1}(x) - \varphi(\bar{x}) \\ \vdots \\ \varphi_{i_{\bar{m}}}(x) - \varphi(\bar{x}) \end{pmatrix} \in \mathbb{R}^{\bar{m}}$$

have the following properties:

- F is of class $C^2(\mathbb{R}^n)$ and it holds $F(\bar{x}) = 0$.
- φ_d is convex, real valued, Lipschitz continuous, and positively homogeneous.
- Due to dom $\varphi_d = \mathbb{R}^{\bar{m}}$, Robinson's constraint qualification obviously holds at $\bar{x}$. Moreover, the derivative of F at $\bar{x}$ is given by $DF(\bar{x}) = \nabla \varphi_{\mathcal{A}(\bar{x})}(\bar{x})^\top$.

Thus, together with (5.3.26), this shows that φ is C^2-decomposable at $\bar{x}$ with decomposition pair (φ_d, F). If, in addition, the nondegeneracy condition

$$\mathbb{R}^{\bar{m}} = DF(\bar{x})\mathbb{R}^n - \text{lin } \varphi_d = \nabla \varphi_{\mathcal{A}(\bar{x})}(\bar{x})^\top \mathbb{R}^n - \mathbb{R} \cdot \mathbb{1}$$

is satisfied at $\bar{x}$, then φ is C^2-fully decomposable.

The next example is inspired by [217, Example 2.2].

Example 5.3.17 (Polyhedral functions). A function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is said to be *polyhedral* if and only if its epigraph epi φ is a *polyhedral set*, i.e., if epi φ can be represented in the following form

$$\exists\, C \in \mathbb{R}^{\ell \times (n+1)},\, c \in \mathbb{R}^\ell : \quad \text{epi } \varphi = \left\{(x, t) \in \mathbb{R}^{n+1} : C(x^\top, t)^\top - c \le 0\right\}.$$

Alternatively, by [208, Theorem 2.49], φ is a polyhedral function if only if it can be expressed as

$$\varphi(x) = \varphi_p(x) + \iota_K(x) := \max_{1 \le i \le m} \{a_i^\top x - \alpha_i\} + \iota_K(x),$$

where $a_i \in \mathbb{R}^n$, $\alpha_i \in \mathbb{R}$, $i = 1, ..., \ell$, and $K \subseteq \mathbb{R}^n$ is a polyhedral set. In particular, a polyhedral function is always convex and lower semicontinuous. Now, let $\bar{x} \in \text{dom } \varphi = K$ be arbitrary and let us set $\mathcal{A}(\bar{x}) := \{i : a_i^\top \bar{x} - \alpha_i = \varphi_p(\bar{x})\}$. Then, by using the calculus for max-type functions, see, e.g., [27, Example 2.68], it follows

$$\varphi_p'(\bar{x}; h) = \max_{i \in \mathcal{A}(\bar{x})} \langle a_i, h \rangle.$$

Now, as in the last example, the continuity of φ_p implies $\mathcal{A}(x) \subset \mathcal{A}(\bar{x})$ for all x in a certain neighborhood U of $\bar{x}$ and by applying (2.5.1), we obtain

$$\begin{aligned}\varphi_p(x) \geq \varphi_p'(\bar{x}; x - \bar{x}) + \varphi_p(\bar{x}) &= \max_{i\in\mathcal{A}(\bar{x})} \langle a_i, x - \bar{x}\rangle + \varphi_p(\bar{x}) \\ &\geq \max_{i\in\mathcal{A}(\bar{x})} \{a_i^\top x - \alpha_i\} = \max_{i\in\mathcal{A}(x)} \{a_i^\top x - \alpha_i\} = \varphi_p(x).\end{aligned}$$

On the other hand, since K is a polyhedral set, it follows $\mathcal{R}_K(\bar{x}) = \mathrm{cl}\ \mathcal{R}_K(\bar{x})$ and, consequently, Example 2.5.9 yields

$$\iota_K'(\bar{x}; h) = \iota_{\mathcal{R}_K(\bar{x})}(h) = \iota_{T_K(\bar{x})}(h), \quad \forall\ h \in \mathbb{R}^n.$$

Moreover, by using the structure of the (polyhedral) sets K and $T_K(\bar{x})$, we can easily establish

$$\iota_{T_K(\bar{x})}(x - \bar{x}) = \iota_{K-\bar{x}}(x - \bar{x}) = \iota_K(x)$$

for all x in a sufficiently small neighborhood of $\bar{x}$ (we refer to [208, Theorem 6.46 and Example 6.47] for more details). Thus, after choosing a smaller neighborhood $V \subset U$ of $\bar{x}$, if necessary, we have

$$\varphi(x) = \varphi(\bar{x}) + \varphi'(\bar{x}; x - \bar{x}), \quad \forall\ x \in V.$$

We proceed as in Example 5.3.12 and define the functions $\varphi_d : \mathbb{R}^n \to (-\infty, +\infty]$, $\varphi_d(y) := \varphi'(\bar{x}; y)$ and $F : \mathbb{R}^n \to \mathbb{R}^n$, $F(x) := x - \bar{x}$. Then, the pair (φ_d, F) has the following properties:

- F is of class $C^\infty(\mathbb{R}^n)$ and it holds $F(\bar{x}) = 0$.
- φ_d is a convex, proper, lower semicontinuous, and positively homogeneous function.
- Obviously, the nondegeneracy condition is satisfied at $\bar{x}$.

Hence, the polyhedral function φ is C^∞-fully decomposable at any point $\bar{x} \in \mathrm{dom}\ \varphi$ with decomposition pair (φ_d, F).

Apparently, since the ℓ_1-norm is a polyhedral function, the results in Example 5.3.12 are a direct consequence of the latter example. Moreover, since the composite function

$$\varphi(x) = \mu\|x\|_1 + \iota_{[a,b]}(x), \quad a, b \in \mathbb{R}^n$$

is also polyhedral, Example 5.3.17 immediately implies that ℓ_1-optimization problems with additional box constraints are C^∞-fully decomposable at any feasible point.

In the following, we analyze the decomposability of singular value-based functions, such as, e.g., the nuclear norm or the more general Ky Fan k-norm of a matrix. The example is quite involved and requires various tools from matrix and eigenvalue theory. Our construction essentially follows [220], [27, Example 3.98 and 3.140], [217, Example 2.3], and [63, Proposition 4.3].

Example 5.3.18 (Singular value optimization). Let $\bar{X} \in \mathbb{R}^{m\times n}$, $m \leq n$, be an arbitrary but fixed matrix and let $\sigma_1(\bar{X}) \geq ... \geq \sigma_m(\bar{X})$ denote the singular values of $\bar{X}$ in decreasing order. Furthermore, by $r_1, ..., r_{q+1}$, we denote the multiplicities and by $\mu_1 > ... > \mu_q > 0 = \mu_{q+1}$ the distinct values of the singular values $\sigma_1(\bar{X}), ..., \sigma_m(\bar{X})$, i.e., it holds

$$\mu_j = \sigma_{s_j+1}(\bar{X}) = ... = \sigma_{s_j+r_j}(\bar{X}), \quad s_j := \sum_{i=1}^{j-1} r_i, \quad j = 1, ..., q+1.$$

Moreover, let

$$\bar{X} = \bar{U}[\bar{\Sigma}\ 0](\bar{V}_1\ \bar{V}_2)^\top, \quad \bar{\Sigma} = \operatorname{diag}(\sigma_1(\bar{X}), ..., \sigma_m(\bar{X}))$$

be a corresponding singular value decomposition of $\bar{X}$. Here, the columns of the matrices $\bar{U} \in \mathbb{R}^{m\times m}$, $\bar{V} = (\bar{V}_1\ \bar{V}_2) \in \mathbb{R}^{n\times n}$ are formed by the pairwise orthonormal singular vectors $\bar{u}_1, ..., \bar{u}_m \in \mathbb{R}^m$ and $\bar{v}_1, ..., \bar{v}_n \in \mathbb{R}^n$, respectively. Moreover, for some arbitrary index sets $\mathcal{I}_u \subset \{1, ..., m\}$, $\mathcal{I}_v \subset \{1, ..., n\}$, we will use the abbreviations

$$\bar{U}_{\mathcal{I}_u} := \bar{U}_{[\cdot\mathcal{I}_u]} = (\bar{u}_i)_{i\in\mathcal{I}_u}, \quad \bar{V}_{\mathcal{I}_v} := \bar{V}_{[\cdot\mathcal{I}_v]} = (\bar{v}_i)_{i\in\mathcal{I}_v}.$$

Accordingly, for another matrix $X \in \mathbb{R}^{m\times n}$ let

$$X = U(X)[\Sigma(X)\ 0](V_1(X)\ V_2(X))^\top, \quad \Sigma(X) = \operatorname{diag}(\sigma_1(X), ..., \sigma_m(X)),$$

$U(X) \in \mathbb{R}^{m\times m}$, $V(X) = (V_1(X)\ V_2(X)) \in \mathbb{R}^{n\times n}$ be the singular value decomposition of X and let $U_{\mathcal{I}_u} := U(X)_{[\cdot\mathcal{I}_u]}$ and $V_{\mathcal{I}_v} := V(X)_{[\cdot\mathcal{I}_v]}$ denote the respective submatrices of $U(X)$ and $V(X)$. Now, for $j = 1, ..., q+1$, let us consider the index sets

$$\alpha_j := \{s_j + 1, ..., s_j + r_j\}, \quad \beta := \alpha_{q+1} \cup \{m+1, ..., n\}$$

and let us define the linear operator $\mathcal{B} : \mathbb{R}^{m\times n} \to \mathbb{S}^{m+n}$,

$$\mathcal{B}(X) := \begin{pmatrix} 0 & X \\ X^\top & 0 \end{pmatrix}.$$

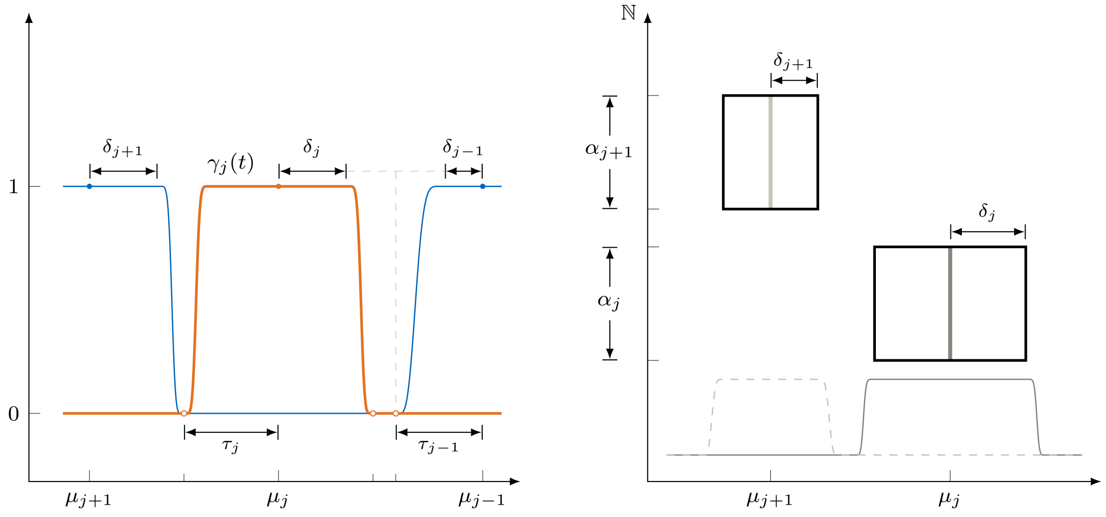
It is well-known that the symmetric matrix $\mathcal{B}(X)$ admits the following eigenvalue decomposition

$$\mathcal{B}(X) = P(X) \begin{pmatrix} \Sigma(X) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -\Sigma(X) \end{pmatrix} P(X)^\top, \quad P(X) := \frac{1}{\sqrt{2}} \begin{pmatrix} U & 0 & U \\ V_1 & \sqrt{2}V_2 & -V_1 \end{pmatrix},$$

where we dropped the argument X in the definition of $P(X)$ for a better readability. Let us also note that the matrix $P(X)$ is obviously orthogonal. Next, since the singular value functions $\sigma_i : \mathbb{R}^{m\times n} \to \mathbb{R}_+$ are globally Lipschitz continuous, see, e.g., [94, Section 8.6], there exists $\delta_j > 0$, such that

$$|\sigma_i(X) - \mu_j| < \delta_j, \quad \forall\ i \in \alpha_j \quad \text{and} \quad \delta_j + \delta_{j+1} < \mu_j - \mu_{j+1}$$

for all $j = 1, ..., q+1$ and all X in a certain neighborhood of $\bar{X}$. Thus, the singular values $\sigma_i(X)$, $i \in \alpha_j$, stay in bounded and disjoint boxes $[\mu_j - \delta_j, \mu_j + \delta_j]$. We will now define a



(a) (b)

Figure 5.2.: Illustration of the functions γ_j and of the distribution of the singular values.

function that separates these different singular value boxes. In particular, let us set

$$c(t) := \begin{cases} e^{-t^{-1}} & \text{if } t > 0, \\ 0 & \text{if } t \le 0, \end{cases} \quad c_j(t) := \frac{c(\tau_j^2 - t)}{c(\tau_j^2 - t) + c(t - \delta_j^2)}, \quad \gamma_j(t) = c_j((t - \mu_j)^2),$$

and

$$\tau_j := \frac{1}{2} \begin{cases} \mu_1 - \mu_2 + \delta_1 - \delta_2 & \text{if } j = 1, \\ \min\{\mu_{j-1} - \mu_j - \delta_{j-1} + \delta_j, \mu_j - \mu_{j+1} + \delta_j - \delta_{j+1}\} & \text{if } 2 \le j \le q, \\ \mu_q + \delta_{q+1} - \delta_q & \text{if } j = q+1, \end{cases}$$

then it holds $\gamma_j(t) = 1$ for all $t \in [\mu_j - \delta_j, \mu_j + \delta_j]$ and $\gamma_j(t) = 0$ for all $|t - \mu_j| > \tau_j$ and all $j = 1, ..., q+1$. Moreover, the functions γ_j are obviously of class C^∞ for all $j = 1, ..., q+1$. (The mappings γ_j and the underlying construction principle are also visualized in Figure 5.2). Let $p_i(X) \in \mathbb{R}^{m+n}$ denote the i-th column of the matrix $P(X)$ and let us define

$$\mathcal{P}_j(X) := \tilde{\mathcal{P}}_j(\mathcal{B}(X)) := \sum_{i=1}^{m+n} p_i(X)\gamma_j(\lambda_i(\mathcal{B}(X)))p_i(X)^\top = \begin{cases} P_{\alpha_j} P_{\alpha_j}^\top & \text{if } j \le q, \\ P_{\beta_{\mathcal{B}}} P_{\beta_{\mathcal{B}}}^\top & \text{if } j = q+1, \end{cases}$$

where

$$P_{\alpha_j} := P(X)_{[\cdot \alpha_j]}, \quad P_{\beta_{\mathcal{B}}} := P(X)_{[\cdot \beta_{\mathcal{B}}]}, \quad \beta_{\mathcal{B}} := \beta \cup \{r_{q+1} + n + 1, ..., n+m\}.$$

Since the columns of the matrices P_{α_j} and $P_{\beta_{\mathcal{B}}}$ span the eigenspace associated with the collection of eigenvalues

$$\{\lambda_i(\mathcal{B}(X)) : i \in \alpha_j\}, \quad j = 1, ..., q+1,$$

the mappings $\mathcal{P}_j(X)$ are the orthogonal projections onto these respective eigenspaces. Moreover, the functions $\tilde{\mathcal{P}}_j$, $j = 1, ..., q+1$, can be interpreted as *Löwner operators* of the matrix $\mathcal{B}(X)$. Thus, due to [18, Exercise V.3.9] or [64, Proposition 4], we can infer that $\tilde{\mathcal{P}}_j$ is (at least) twice continuously differentiable in a neighborhood of $\mathcal{B}(\bar{X})$ for every $j = 1, .., q+1$. Since $\mathcal{B}$ is a linear operator, this shows that, for all $j = 1, ..., q+1$, the function $\mathcal{P}_j(X)$ is twice continuously differentiable in a neighborhood $\mathcal{N}$ of $\bar{X}$. In the following, we will distinguish two different cases.

Case 1: $1 \leq j \leq q$. Let us define

$$Q_j(X) := \mathcal{P}_j(X)\bar{P}_{\alpha_j} \in \mathbb{R}^{m+n\times r_j}, \quad \mathcal{Q}_j(X) := Q_j(X)(Q_j(X)^\top Q_j(X))^{-\frac{1}{2}},$$

where $\bar{P}_{\alpha_j} = P(\bar{X})_{[\cdot\alpha_j]}$. Since $\mathcal{P}_j(X)$ is the orthogonal projection onto the eigenspace associated with the set of eigenvalues $\{\lambda_i(\mathcal{B}(X)) : i \in \alpha_j\}$, the columns of $\mathcal{Q}_j(X)$ are elements of the subspace sp $P(X)_{[\cdot\alpha_j]}$. Moreover, due to $Q_j(\bar{X}) = \bar{P}_{\alpha_j}$, it follows

$$Q_j(\bar{X})^\top Q_j(\bar{X}) = \bar{P}_{\alpha_j}^\top \bar{P}_{\alpha_j} = \frac{1}{2}(\bar{U}_{\alpha_j}^\top \bar{U}_{\alpha_j} + \bar{V}_{\alpha_j}^\top \bar{V}_{\alpha_j}) = I \in \mathbb{R}^{r_j\times r_j}.$$

Consequently, for all X in a neighborhood of $\bar{X}$, the inverse matrix root $(Q_j(X)^\top Q_j(X))^{-\frac{1}{2}}$ is well-defined and the matrix $\mathcal{Q}_j(X)$ has full column rank. Additionally, it holds

$$\mathcal{Q}_j(X)^\top \mathcal{Q}_j(X) = (Q_j(X)^\top Q_j(X))^{-\frac{1}{2}} Q_j(X)^\top Q_j(X)(Q_j(X)^\top Q_j(X))^{-\frac{1}{2}} = I.$$

Thus, the columns of $\mathcal{Q}_j(X)$ are pairwise orthonormal and we obtain

$$\text{sp } \mathcal{Q}_j(X) = \text{sp } P(X)_{[\cdot\alpha_j]}. \tag{5.3.27}$$

Finally, since the inverse matrix root can be written as a specific Löwner operator,

$$\mathbb{S}^\ell \ni Y = P(Y)\Lambda(Y)P(Y)^\top, \quad Y^{-\frac{1}{2}} = \sum_{i=1}^{\ell} p_i(Y)\varrho(\lambda_i(Y))p_i(Y), \quad \varrho(t) := \frac{1}{\sqrt{t}},$$

and since ϱ is C^∞ on $\mathbb{R}\setminus\{0\}$, the function $\mathcal{Q}_j(X)$ is also twice continuously differentiable in a certain neighborhood of $\bar{X}$. Now, let us define

$$\Xi_j : \mathbb{R}^{m\times n} \to \mathbb{S}^{r_j}, \quad X \mapsto \Xi_j(X) := \mathcal{Q}_j(X)^\top \mathcal{B}(X)\mathcal{Q}_j(X).$$

Then, there exists a neighborhood $\mathcal{N}_j \subset \mathcal{N}$ of $\bar{X}$, such that the mapping Ξ_j has the following properties:

- $\Xi_j(X)$ is twice continuously differentiable on $\mathcal{N}_j$.
- It holds $\Xi_j(\bar{X}) = \mathcal{Q}_j(\bar{X})^\top \mathcal{B}(\bar{X})\mathcal{Q}_j(\bar{X}) = \frac{1}{2}(\bar{V}_{\alpha_j}^\top \bar{X}^\top \bar{U}_{\alpha_j} + \bar{U}_{\alpha_j}^\top \bar{X}\bar{V}_{\alpha_j}) = \mu_j I$.
- For all $X \in \mathcal{N}_j$, the eigenvalues of the symmetric matrix $\Xi_j(X)$ coincide with the set of singular values

$$\{\lambda_i(\Xi_j(X)) : i = 1, ..., r_j\} = \{\sigma_i(X) : i \in \alpha_j\}.$$

- The derivative mapping $D\Xi_j(\bar X) : \mathbb{R}^{m\times n} \to \mathbb{S}^{r_j}$ is onto and satisfies
$$D\Xi_j(\bar X)[H] = \bar P_{\alpha_j}^\top \mathcal{B}(H)\bar P_{\alpha_j} = \mathrm{sym}(\bar U_{\alpha_j}^\top H \bar V_{\alpha_j}).$$

Since the third statement follows from (5.3.27) and from the orthogonality of $\mathcal{Q}_j(X)$, we only need to prove the last part. First, let us consider the derivative of the matrix root mapping

$$\mathcal{S} : \mathbb{S}^\ell \to \mathbb{S}^\ell, \quad Y \mapsto \mathcal{S}(Y) := Y^{\frac{1}{2}}.$$

For every $Y \in \mathbb{S}^\ell_+$, the function $\mathcal{S}(Y)$ is uniquely characterized by the equation

$$\mathcal{S}(Y)\cdot\mathcal{S}(Y) = Y.$$

Consequently, the derivative of $\mathcal{S}$ has to satisfy

$$D\mathcal{S}(Y)[H]\cdot\mathcal{S}(Y) + \mathcal{S}(Y)\cdot D\mathcal{S}(Y)[H] = H, \quad \forall\ H \in \mathbb{S}^\ell.$$

In particular, in the case $Y = I$, we readily obtain $D\mathcal{S}(I)[H] = \frac{1}{2}H$. We then have

$$\begin{aligned} D\mathcal{Q}_j(\bar X)[H] &= DQ_j(\bar X)[H] - Q_j(\bar X)\cdot D\mathcal{S}(I)[(DQ_j(\bar X)[H])^\top Q_j(\bar X) + Q_j(\bar X)^\top DQ_j(\bar X)[H]] \\ &= M\bar P_{\alpha_j} - \frac{1}{2}\bar P_{\alpha_j}(\bar P_{\alpha_j}^\top M^\top \bar P_{\alpha_j} + \bar P_{\alpha_j}^\top M \bar P_{\alpha_j}) = (I - \bar P_{\alpha_j}\bar P_{\alpha_j}^\top) M \bar P_{\alpha_j}, \end{aligned}$$

where $M := D\mathcal{P}_j(\bar X)[H] \in \mathbb{S}^{m+n}$, and it follows

$$D\Xi_j(\bar X)[H] = (D\mathcal{Q}_j(\bar X)[H])^\top \mathcal{B}(\bar X)\bar P_{\alpha_j} + \bar P_{\alpha_j}^\top(\mathcal{B}(H)\bar P_{\alpha_j} + \mathcal{B}(\bar X) D\mathcal{Q}_j(\bar X)[H]).$$

By using $\bar U_{\alpha_j}^\top \bar X = \mu_j \bar V_{\alpha_j}$ and $\bar V_{\alpha_j}^\top \bar X^\top = \mu_j \bar U_{\alpha_j}^\top$, we have $\bar P_{\alpha_j}^\top \mathcal{B}(\bar X) = \mu_i \bar P_{\alpha_j}^\top$. Thus, it immediately follows

$$(D\mathcal{Q}_j(\bar X)[H])^\top \mathcal{B}(\bar X)\bar P_{\alpha_j} = \bar P_{\alpha_j}^\top \mathcal{B}(\bar X) D\mathcal{Q}_j(\bar X)[H] = 0$$

and we establish $D\Xi_j(\bar X)[H] = \bar P_{\alpha_j}^\top \mathcal{B}(H)\bar P_{\alpha_j}$. Moreover, this mapping is clearly onto.

Case 2: $j = q+1$. In this case, the projection $\mathcal{P}_{q+1}(X)$ takes the following form

$$\mathcal{P}_{q+1}(X) = P_{\beta_\mathcal{B}} P_{\beta_\mathcal{B}}^\top = \begin{pmatrix} \mathcal{U}_{q+1}(X) & 0 \\ 0 & \mathcal{V}_{q+1}(X) \end{pmatrix},$$

where $\mathcal{U}_{q+1}(X) := U_{\alpha_{q+1}} U_{\alpha_{q+1}}^\top$ and $\mathcal{V}_{q+1}(X) := V_\beta V_\beta^\top$. We now define

$$\begin{aligned} L_{q+1}(X) &:= \mathcal{U}_{q+1}(X)\bar U_{\alpha_{q+1}}, & \mathcal{L}_{q+1}(X) &:= L_{q+1}(X)(L_{q+1}(X)^\top L_{q+1}(X))^{-\frac{1}{2}}, \\ R_{q+1}(X) &:= \mathcal{V}_{q+1}(X)\bar V_\beta, & \mathcal{R}_{q+1}(X) &:= R_{q+1}(X)(R_{q+1}(X)^\top R_{q+1}(X))^{-\frac{1}{2}}. \end{aligned}$$

Here, the mappings $\mathcal{U}_{q+1}(X)$ and $\mathcal{V}_{q+1}(X)$ are the orthogonal projections onto the left and right eigenspaces associated with the set of singular values $\{\sigma_i(X) : i \in \alpha_{q+1}\}$. Hence, the columns of $\mathcal{L}_{q+1}(X)$ and $\mathcal{R}_{q+1}(X)$ are elements of the subspaces sp $U(X)_{[\cdot\alpha_{q+1}]}$ and sp $V(X)_{[\cdot\beta]}$, respectively. Moreover, due to

$$L_{q+1}(\bar{X}) = \bar{U}_{\alpha_{q+1}}, \quad R_{q+1}(\bar{X}) = \bar{V}_\beta,$$

the matrices $\mathcal{L}_{q+1}(X)$ and $\mathcal{R}_{q+1}(X)$ have full column rank in a neighborhood of $\bar{X}$. As in the first case, it can be easily shown that the functions $\mathcal{L}_{q+1}(X)$ and $\mathcal{R}_{q+1}(X)$ are twice continuously differentiable near $\bar{X}$ and the columns of $\mathcal{L}_{q+1}(X)$ and $\mathcal{R}_{q+1}(X)$ are pairwise orthonormal. Consequently, it readily follows

$$\text{sp } \mathcal{L}_{q+1}(X) = \text{sp } U(X)_{[\cdot\alpha_{q+1}]}, \quad \text{sp } \mathcal{R}_{q+1}(X) = \text{sp } V(X)_{[\cdot\beta]}. \tag{5.3.28}$$

We now define

$$\Xi_{q+1} : \mathbb{R}^{m\times n} \to \mathbb{R}^{r_{q+1}\times|\beta|}, \quad X \mapsto \Xi_{q+1}(X) := \mathcal{L}_{q+1}(X)^\top X \mathcal{R}_{q+1}(X).$$

Then there exists a neighborhood $\mathcal{N}_{q+1} \subset \mathcal{N}$, such that Ξ_{q+1} has the following properties:

- $\Xi_{q+1}(X)$ is twice continuously differentiable on $\mathcal{N}_{q+1}$.

- It holds $\Xi_{q+1}(\bar{X}) = 0$.

- For all $X \in \mathcal{N}_{q+1}$ the singular values of the matrix $\Xi_{q+1}(X)$ coincide with the set of singular values

$$\{\sigma_i(\Xi_{q+1}(X)) : i = 1, ..., r_{q+1}\} = \{\sigma_i(X) : i \in \alpha_{q+1}\}.$$

- The derivative $D\Xi_{q+1}(\bar{X}) : \mathbb{R}^{m\times n} \to \mathbb{R}^{r_{q+1}\times|\beta|}$ is onto and it holds

$$D\Xi_{q+1}(\bar{X})[H] = \bar{U}_{\alpha_{q+1}}^\top H \bar{V}_\beta.$$

Since singular values are invariant under left and right orthogonal transformations, the third part follows again from (5.3.28). The derivative $D\Xi_{q+1}(\bar{X})[H]$ can be computed as in the last case. (Here, we have to use $\bar{U}_{\alpha_{q+1}}^\top \bar{X} = 0$ and $\bar{V}_\beta^\top \bar{X}^\top = 0$).

Let us note that the separate discussion of the zero singular values cannot be avoided in general. In particular, since the mapping $H \mapsto \bar{P}_{\beta_\mathcal{B}}^\top \mathcal{B}(H)\bar{P}_{\beta_\mathcal{B}}$ is typically not onto, we cannot reuse the basic construction principle from the first case to define $\Xi_{q+1}(X)$.

Finally, let us consider the so-called Ky Fan k-norm

$$\|\cdot\|_{(k)} : \mathbb{R}^{m\times n} \to \mathbb{R}_+, \quad \|X\|_{(k)} := \sum_{i=1}^k \sigma_i(X), \quad k \in \{1, ..., m\},$$

which denotes the sum of the k-largest singular values. Using our latter constructions, we will show that $\|\cdot\|_{(k)}$ is C^2-fully decomposable at every $\bar{X} \in \mathbb{R}^{m\times n}$. Again, depending on the singular value $\sigma_k(\bar{X})$, we have to discuss two different cases.

Case 1: $\sigma_k(\bar{X}) > 0$. Then, there exists $1 \le q_0 \le q$ such that $\sigma_k(\bar{X}) \in \alpha_{q_0}$ and we define

$$F : \mathbb{R}^{m\times n} \to \mathbb{R} \times \mathbb{S}^{r_{q_0}}, \quad F(X) := \begin{pmatrix} \sum_{j=1}^{q_0-1} \operatorname{tr}(\Xi_j(X) - \Xi_j(\bar{X})) \\ \Xi_{q_0}(X) - \mu_{q_0} I \end{pmatrix},$$

$$\varphi_d : \mathbb{R} \times \mathbb{S}^{r_{q_0}} \to \mathbb{R}, \quad \varphi_d(t, Y) := t + s(Y)_{(k-s_{q_0})}, \quad s(Y)_{(k-s_{q_0})} := \sum_{i=1}^{k-s_{q_0}} \lambda_i(Y).$$

Here, the function $s(\cdot)_{(\ell)}$ denotes the sum of the ℓ-largest eigenvalues of a symmetric matrix. Furthermore, the Ky Fan k-norm can be represented as follows:

$$\begin{aligned} \|X\|_{(k)} = \sum_{j=1}^{q_0-1} \sum_{i\in\alpha_j} \sigma_i(X) + \sum_{i=s_{q_0}+1}^{k} \sigma_i(X) &= \sum_{j=1}^{q_0-1} \sum_{i=1}^{r_j} \lambda_i(\Xi_j(X)) + \sum_{i=1}^{k-s_{q_0}} \lambda_i(\Xi_{q_0}(X)) \\ &= \sum_{j=1}^{q_0-1} \operatorname{tr}(\Xi_j(X)) + s(\Xi_{q_0}(X))_{(k-s_{q_0})}. \end{aligned}$$

Now, setting $\hat{\mathcal{N}} := \bigcap_j \mathcal{N}_j$, the functions φ_d and F have the following properties:

- Clearly, our preceding discussion implies that F is twice continuously differentiable on the open set $\hat{\mathcal{N}}$. Moreover, it holds $F(\bar{X}) = 0$.
- Since $s(\cdot)_{(k-s_{q_0})}$ is convex and positively homogeneous, φ_d is a convex, real valued, and positively homogeneous mapping.
- Due to $\lambda_i(Y + \kappa I) = \lambda_i(Y) + \kappa$ for all κ and i, our latter calculation implies

$$\|X\|_{(k)} = \varphi_d(F(X)) + \|\bar{X}\|_{(k)}, \quad \forall\ X \in \hat{\mathcal{N}}.$$

- The derivative mapping $DF(\bar{X}) : \mathbb{R}^{m\times n} \to \mathbb{R} \times \mathbb{S}^{r_{q_0}}$ is given by

$$DF(\bar{X})[H] = \begin{pmatrix} \sum_{j=1}^{q_0-1} \operatorname{tr}(D\Xi_j(\bar{X})[H]) \\ D\Xi_{q_0}(\bar{X})[H] \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{q_0-1} \operatorname{tr}(\bar{P}_{\alpha_j}^\top \mathcal{B}(H) \bar{P}_{\alpha_j}) \\ \operatorname{sym}(\bar{U}_{\alpha_{q_0}}^\top H \bar{V}_{\alpha_{q_0}}) \end{pmatrix}$$

  and it can be easily shown that the function $DF(\bar{X})$ is onto. Consequently, the nondegeneracy condition is satisfied at $\bar{X}$.

Case 2: $\sigma_k(\bar{X}) = 0$. In this situation, it holds $\sigma_k(\bar{X}) \in \alpha_{q+1}$ and we define

$$F : \mathbb{R}^{m\times n} \to \mathbb{R} \times \mathbb{R}^{r_{q+1}\times|\beta|}, \quad F(X) := \begin{pmatrix} \sum_{j=1}^{q} \operatorname{tr}(\Xi_j(X) - \Xi_j(\bar{X})) \\ \Xi_{q+1}(X) \end{pmatrix},$$

$$\varphi_d : \mathbb{R} \times \mathbb{R}^{r_{q+1}\times|\beta|} \to \mathbb{R}, \quad \varphi_d(t, Y) := t + \|Y\|_{(k-s_{q+1})}.$$

Again, setting $\hat{\mathcal{N}} := \bigcap_j \mathcal{N}_j$, the mappings φ_d and F have the following properties:

- F is twice continuously differentiable on the neighborhood $\hat{\mathcal{N}}$ and it holds $F(\bar{X}) = 0$.

- Obviously, φ_d is a convex, real valued, and positively homogeneous function.
- As in the last case and due to $\mu_{q+1} = 0$, it follows
$$\|X\|_{(k)} = \varphi_d(F(X)) + \|\bar{X}\|_{(k)}, \quad \forall\, X \in \hat{\mathcal{N}}.$$
- The derivative mapping $DF(\bar{X}) : \mathbb{R}^{m\times n} \to \mathbb{R} \times \mathbb{R}^{r_{q+1}\times|\beta|}$ is given by
$$DF(\bar{X})[H] = \begin{pmatrix} \sum_{j=1}^{q} \mathrm{tr}(D\Xi_j(\bar{X})[H]) \\ D\Xi_{q+1}(\bar{X})[H] \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{q} \mathrm{tr}(\bar{P}_{\alpha_j}^\top \mathcal{B}(H)\bar{P}_{\alpha_j}) \\ \bar{U}_{\alpha_{q+1}}^\top H \bar{V}_\beta \end{pmatrix}$$
and is also onto.

Combining the latter cases, we establish that the Ky Fan k-norm is C^2-fully decomposable at each $\bar{X} \in \mathbb{R}^{m\times n}$. For instance, this also proves that the spectral and the nuclear norm are fully decomposable matrix functions.

As mentioned at several points in this section, the concept of decomposable functions is closely connected to the notion of cone reducible sets in constrained optimization. In the following, we will clarify and explain this connections in some more detail. We start with a definition of reducible sets.

Definition 5.3.19 (cf. [27, Definition 3.135]). *Let $K \subset \mathbb{R}^n$ and $C \subset \mathbb{R}^m$ be two convex, closed sets. The set K is said to be* C^ℓ-reducible *to the set C, at a point $\bar{x} \in K$, if there exists a neighborhood U of $\bar{x}$ and an ℓ-times continuously differentiable function $F : U \to \mathbb{R}^m$ such that:*

- *The derivative mapping $DF(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is onto.*
- *It holds $K \cap U = \{x \in U : F(x) \in C\}$.*

We say that the reduction is pointed *if the tangent cone $T_C(F(\bar{x}))$ is a pointed cone. If, in addition, the set $C - F(\bar{x})$ is a pointed, convex, and closed cone, we say that K is* C^ℓ-cone reducible. *We can assume without loss of generality that $F(\bar{x}) = 0$.*

Let us note that a cone $C \subset \mathbb{R}^n$ is said to be pointed if and only if its corresponding lineality space lin C is $\{0\}$. The following example is taken from [217] and connects cone reducible sets and decomposable functions.

Example 5.3.20 (Cone reducible sets and decomposability). Consider the indicator function $\varphi(x) = \iota_K(x)$ and a point $\bar{x} \in K$, and suppose that the set K is C^ℓ-cone reducible to the set $C \subset \mathbb{R}^m$ at $\bar{x}$. Then, due to Definition 5.3.19, there exist a neighborhood U of $\bar{x}$ and an ℓ-times continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^m$ such that
$$F(\bar{x}) = 0, \quad \varphi(x) = \varphi(\bar{x}) + \iota_C(F(x)), \quad \forall\, x \in U.$$

Since C is a convex, closed cone, the function $\varphi_d(y) := \iota_C(y)$ is convex, proper, lower semicontinuous, and positively homogeneous. Moreover, since the mapping $DF(\bar{x})$ is onto, the nondegeneracy condition is automatically satisfied. Consequently, φ is C^ℓ-fully decomposable at $\bar{x}$ with decomposition pair (φ_d, F). On the other hand, if the indicator function $\varphi(x) = \iota_K(x)$ is C^ℓ-fully decomposable at $\bar{x} \in K$ and if the derivative mapping of the corresponding decomposition function F is onto, then it can be easily shown that the set K is C^ℓ(-cone) reducible at $\bar{x}$.

Thus, the concept of decomposability can also be applied to optimization problems with cone-reducible constraints. The class of cone-reducible sets contains many different and interesting examples. For instance, any polyhedral set and the cone of positive semidefinite matrices $\mathbb{S}^n_+ \subset \mathbb{R}^{n\times n}$ are C^∞-cone reducible; see [27, Example 3.139 and 3.140]. Further examples comprise the second order cone, [25],

$$\mathcal{K} = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \|x\|_2 \leq t\} = \text{epi } \|\cdot\|_2,$$

or the epigraph of the Ky-Fan k-norm, which also includes the epigraph of the nuclear norm as a special case, we refer to [63, Chapter 4] for more details. Let us also point out that decomposable functions and cone-reducible sets share many of their characteristic, second order properties. In particular, similar to decomposable functions, the curvature term of a cone-reducible set reduces to a quadratic function on the critical cone. More information on reducible sets and on their specific properties can be found in [27, Section 3.4.4].

In the following, we establish a connection between full decomposability of a function φ and cone-reducibility of the corresponding epigraph epi φ.

Corollary 5.3.21. *Suppose that the function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is C^ℓ-fully decomposable at some point $\bar{x} \in$ dom φ with decomposition pair (φ_d, F). If the mapping $DF(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is onto, then the epigraph* epi φ *is C^ℓ-reducible to the set* epi φ_d*, at the point $(\bar{x}, \varphi(\bar{x}))$. Additionally, if the function φ_d is pointed, i.e., if* lin $\varphi_d = \{0\}$*, then the epigraph* epi φ *is C^ℓ-cone reducible at $(\bar{x}, \varphi(\bar{x}))$.*

Proof. Let φ be C^ℓ-fully decomposable at $\bar{x} \in$ dom φ with decomposition pair (φ_d, F). Then, by Definition 5.3.1, there exists a neighborhood U of $\bar{x}$ such that

$$\varphi(x) = \varphi(\bar{x}) + \varphi_d(F(x)), \quad \forall\ x \in U.$$

and for any $(x,t) \in \text{epi } \varphi \cap U \times \mathbb{R}$, it follows

$$\varphi(x) \leq t \quad \iff \quad \varphi_d(F(x)) \leq t - \varphi(\bar{x}) \quad \iff \quad (F(x), t - \varphi(\bar{x})) \in \text{epi } \varphi_d.$$

Clearly, this implies

$$\text{epi } \varphi \cap U \times \mathbb{R} = \{(x,t) \in U \times \mathbb{R} : (F(x), t - \varphi(\bar{x})) \in \text{epi } \varphi_d\}.$$

Furthermore, since the mapping $DF(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^m$ is supposed to be onto, this also shows that the epigraph epi φ is C^ℓ-reducible to the set epi φ_d at $(\bar{x}, \varphi(\bar{x}))$. Now, since the function φ_d is convex, proper, lower semicontinuous, and positively homogeneous, the set epi φ_d is a convex, nonempty, and closed cone. Moreover, due to Lemma 2.5.3 and (5.3.5), we have

$$T_{\text{epi } \varphi_d}(F(\bar{x}), \varphi(\bar{x}) - \varphi(\bar{x})) = T_{\text{epi } \varphi_d}(0, 0) = \text{epi } \varphi_d^\downarrow(0; \cdot) = \text{epi } \varphi_d.$$

Consequently, under the additional assumption lin $\varphi_d = \{0\}$, the cone epi φ_d is pointed and epi φ is C^ℓ-cone reducible at $(\bar{x}, \varphi(\bar{x}))$. □

We conclude this subsection with a simple sum and chain rule for decomposable functions.

Lemma 5.3.22. *Let the functions $\varphi^1, \varphi^2 : \mathbb{R}^n \to (-\infty, +\infty]$ be C^ℓ-decomposable at some point $\bar{x} \in \mathbb{R}^n$ with corresponding decomposition pairs (φ^1_d, F^1) and (φ^2_d, F^2), and let φ^1 be real valued. Then, the function $\varphi \equiv \alpha\varphi^1 + \beta\varphi^2$ is C^ℓ-decomposable at $\bar{x}$ for every choice $\alpha, \beta \geq 0$.*

Proof. It can be easily shown that the functions $\varphi_d : \mathbb{R}^{m_1+m_2} \to (-\infty, +\infty]$, $F : \mathbb{R}^n \to \mathbb{R}^{m_1+m_2}$,

$$\varphi_d(y, z) := \alpha\varphi^1_d(y) + \beta\varphi^2_d(z), \quad F(x) := \begin{pmatrix} F^1(x) \\ F^2(x) \end{pmatrix}$$

form a decomposition pair (φ_d, F) of the mapping φ. Moreover, since the function φ^1 is real valued, Robinson's constraint qualification

$$DF(\bar{x})\mathbb{R}^n - \text{dom } \varphi_d = \begin{pmatrix} DF^1(\bar{x}) \\ DF^2(\bar{x}) \end{pmatrix} \mathbb{R}^n - \begin{pmatrix} \mathbb{R}^{m_1} \\ \text{dom } \varphi^2_d \end{pmatrix} = \mathbb{R}^{m_1+m_2}$$

immediately follows from the condition $DF^2(\bar{x})\mathbb{R}^n - \text{dom } \varphi^2_d = \mathbb{R}^{m_2}$. This establishes the desired C^ℓ-decomposability of φ. □

Lemma 5.3.23 (Chain rule). *Let $\varphi : \mathbb{R}^m \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping and let $G : \mathbb{R}^n \to \mathbb{R}^m$ be ℓ-times continuously differentiable in a neighborhood of some point $\bar{x} \in$ dom φ. Suppose that φ is C^ℓ-decomposable at $G(\bar{x})$ and that Robinson's constraint qualification*

$$0 \in \text{int}\{G(\bar{x}) + DG(\bar{x})\mathbb{R}^n - \text{dom } \varphi\} \tag{5.3.29}$$

holds at $\bar{x}$. Then, the composite function $\varphi \circ G$ is C^ℓ-decomposable at $\bar{x}$. Moreover, if φ is C^ℓ-fully decomposable at $G(\bar{x})$ and if the nondegeneracy condition

$$DG(\bar{x})\mathbb{R}^n + \text{lin } N_{\partial\varphi(G(\bar{x}))}(\lambda) = \mathbb{R}^m, \quad \lambda \in \partial\varphi(G(\bar{x})),$$

is satisfied at $\bar{x}$, then $\varphi \circ G$ is C^ℓ-fully decomposable at $\bar{x}$.

Proof. Since φ is C^ℓ-decomposable at $G(\bar{x})$, there exist functions $F : \mathbb{R}^m \to \mathbb{R}^p$, $\varphi_d : \mathbb{R}^p \to (-\infty, +\infty]$ such that

$$\varphi(y) = \varphi(G(\bar{x})) + \varphi_d(F(y)) \tag{5.3.30}$$

for all y in a neighborhood $V \subset \mathbb{R}^m$ of $G(\bar{x})$. Moreover, since G is continuous near $\bar{x}$, there exists an open, nonempty set $U \subset \mathbb{R}^n$ such that $\bar{x} \in U$ and $G(U) \subset V$. Then, (5.3.30) implies

$$(\varphi \circ G)(x) = (\varphi \circ G)(\bar{x}) + \varphi_d(F(G(x))), \quad \forall\, x \in U.$$

Now, our goal is to show that $\varphi \circ G$ is decomposable at $\bar{x}$ with respect to the decomposition pair $(\varphi_d, F \circ G)$. Clearly, the composite function $F \circ G : U \to \mathbb{R}^p$ is ℓ-times continuously differentiable and we have $(F \circ G)(\bar{x}) = 0$. Since the function φ_d is part of the decomposition pair (φ_d, F), it is necessarily convex, proper, lower semicontinuous, and positively homogeneous. We have already seen that Robinson's constraint qualification (5.3.29) can be equivalently reformulated as

$$0 \in \operatorname{int} \left\{ \begin{pmatrix} G(\bar{x}) \\ \varphi(G(\bar{x})) \end{pmatrix} + \begin{pmatrix} DG(\bar{x})\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - \text{epi } \varphi \right\}.$$

Furthermore, by [27, Proposition 2.97 and Corollary 2.98] and Lemma 2.5.3, this condition is also equivalent to

$$(5.3.31) \quad \begin{pmatrix} \mathbb{R}^m \\ \mathbb{R} \end{pmatrix} = \begin{pmatrix} DG(\bar{x})\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - T_{\text{epi } \varphi}(G(\bar{x}), \varphi(G(\bar{x}))) = \begin{pmatrix} DG(\bar{x})\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - \text{epi } \varphi^{\downarrow}(G(\bar{x}); \cdot).$$

Since φ is decomposable at $G(\bar{x})$, Robinson's constraint qualification (5.3.2), (with respect to φ_d and F), is satisfied at $G(\bar{x})$ and, by applying Lemma 2.5.6, we have

$$\begin{aligned} \varphi^{\downarrow}(G(\bar{x}); h) &= (\varphi_d \circ F)^{\downarrow}(G(\bar{x}); h) \\ &= \varphi_d^{\downarrow}(F(G(\bar{x})); DF(G(\bar{x}))h) = \varphi_d^{\downarrow}(0; DF(G(\bar{x}))h) = \varphi_d(DF(G(\bar{x}))h). \end{aligned}$$

This immediately establishes

$$\Phi \cdot \text{epi } \varphi^{\downarrow}(G(\bar{x}); \cdot) \subset \text{epi } \varphi_d, \quad \Phi := \begin{pmatrix} DF(G(\bar{x})) & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{p+1 \times m+1}$$

and, by multiplying (5.3.31) with Φ and subtracting epi φ_d, we obtain the following inclusion

$$\begin{pmatrix} D(F \circ G)(\bar{x})\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - (\text{epi } \varphi_d + \text{epi } \varphi_d) \supset \begin{pmatrix} DF(G(\bar{x}))\mathbb{R}^m \\ \mathbb{R} \end{pmatrix} - \text{epi } \varphi_d = \begin{pmatrix} \mathbb{R}^p \\ \mathbb{R} \end{pmatrix}.$$

Moreover, since φ_d is convex and positively homogeneous, the epigraph epi φ_d is a convex cone and it follows epi $\varphi_d +$ epi $\varphi_d \subset$ epi φ_d, (see, e.g., [11, Proposition 6.4]). Together with the latter inclusion, this yields

$$D(F \circ G)(\bar{x})\mathbb{R}^n - \text{dom } \varphi_d = \mathbb{R}^p.$$

Hence, Robinson's constraint qualification is satisfied at $\bar{x}$, (with respect to the decomposition pair $(\varphi_d, F \circ G)$), and consequently, φ is C^ℓ-decomposable at $\bar{x}$. Now, let us suppose that φ is C^ℓ-fully decomposable at $G(\bar{x})$. Then, due to

$$\text{lin } N_{\partial\varphi(G(\bar{x}))}(\lambda) = \text{lin } \varphi^{\downarrow}(G(\bar{x}); \cdot),$$

we can reuse and adapt the last steps of the proof to verify that the nondegeneracy condition,

$$D(F \circ G)(\bar{x})\mathbb{R}^n - \text{lin } \varphi_d = \mathbb{R}^p,$$

is fulfilled at $\bar{x}$. This finally shows that φ is C^ℓ-fully decomposable at $\bar{x}$ with decomposition pair $(\varphi_d, F \circ G)$. □

Remark 5.3.24. In [218, Proposition 3.2], Shapiro showed that sets of the form

$$S = \{x \in \mathbb{R}^n : G(x) \in K\}$$

are cone reducible at some point $\bar{x}$ if the set $K \subset \mathbb{R}^m$ is cone reducible at $G(\bar{x})$ and if the nondegeneracy condition holds at $\bar{x}$. Hence, Lemma 5.3.23 transfers and extends this result to (fully) decomposable functions. Let us further note that similar computational results have also been established by Rockafellar et al., [193, 194, 208], for the composition and sum of (fully)-amenable functions. Moreover, we want to emphasize that chain rules for amenable functions are also available for general, nonconvex functions φ. For more details we refer to [193, 194] and [208, Section 10.F].

5.3.4. The curvature of fully decomposable functions

In this subsection, we will discuss an essential property of fully decomposable functions in some more detail. In particular, let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and suppose that the mapping $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is convex, proper, lower semicontinuous, and C^2-fully decomposable at some stationary point $\bar{x} \in \operatorname{dom} \varphi$ of our initial problem

$$\min_x \; \psi(x) = f(x) + \varphi(x). \tag{5.3.32}$$

Then, under the strict complementarity condition, we will establish a connection between the curvature term $\xi^*_{\varphi,h}$, which is associated with the problem (5.3.32), and the Fréchet derivative of the proximity operator $\operatorname{prox}^\Lambda_\varphi$ at $\bar{x} - \Lambda^{-1}\nabla f(\bar{x})$. This is one of the most crucial steps in order to combine the second order conditions (5.3.9) and (5.3.11), which are based on the knowledge of a specific decomposition pair (φ_d, F) of φ, and to prove nonsingularity conditions for the generalized derivatives of the nonsmooth function

$$F^\Lambda(\bar{x}) = \bar{x} - \operatorname{prox}^\Lambda_\varphi(\bar{x} - \Lambda^{-1}\nabla f(\bar{x})), \quad \Lambda \in \mathbb{S}^n_{++}.$$

Now, let (φ_d, F) be a decomposition pair of the function φ. The corresponding composite optimization problem is given by

$$\min_x \; \psi_c(x) = f(x) + \varphi_d(F(x)) + \bar{c}, \quad \bar{c} = \varphi(\bar{x}). \tag{5.3.33}$$

Moreover, by Definition 5.1.5, the critical cone of the problems (5.3.32) and (5.3.33), has the following equivalent representations

$$\begin{aligned} \mathcal{C}(\bar{x}) = N_{\partial\varphi(\bar{x})}(-\nabla f(\bar{x})) &= \{h \in \mathbb{R}^n : \psi^\downarrow(\bar{x}; h) = 0\} \\ &= \{h \in \mathbb{R}^n : \nabla f(\bar{x})^\top h + \varphi_d(DF(\bar{x})h) = 0\} \\ &= \{h \in \mathbb{R}^n : DF(\bar{x})h \in N_{\partial\varphi_d(0)}(\bar{\lambda})\}, \end{aligned}$$

where $\bar{\lambda} \in \mathcal{M}(\bar{x})$ is an associated (unique) Lagrange multiplier. Let us recall that the strict complementarity condition is said to be satisfied at $\bar{x}$ if and only if

$$-\nabla f(\bar{x}) \in \operatorname{ri} \partial\varphi(\bar{x}). \tag{5.3.34}$$

The next lemma shows that the nondegeneracy condition guarantees equivalence of the latter condition and of the respective strict complementarity condition for the composite problem (5.3.33) (in the sense of Definition 5.1.8). This result is motivated by a discussion in [217, Section 4]. However, our proof uses different arguments.

Lemma 5.3.25. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. Suppose that $\bar{x} \in \operatorname{dom} \varphi$ is a stationary point of problem $(\mathcal{P})$ and φ is C^1-fully decomposable at $\bar{x}$ with decomposition pair (φ_d, F). Then, the following conditions are equivalent:*

$$\begin{aligned} -\nabla f(\bar{x}) \in \operatorname{ri} \partial\varphi(\bar{x}) \iff &\ \mathcal{C}(\bar{x}) \text{ is a subspace} \\ \iff &\ N_{\partial\varphi_d(0)}(\bar{\lambda}) \text{ is a subspace} \iff \bar{\lambda} \in \operatorname{ri} \partial\varphi_d(0), \end{aligned}$$

where $\bar{\lambda} \in \mathcal{M}(\bar{x})$ is the corresponding, unique Lagrange multiplier of the problem (5.3.33)*.*

Proof. By Lemma 5.1.10, we only need to prove the second equivalence. Moreover, from the above discussion, it immediately follows $DF(\bar{x})\mathcal{C}(\bar{x}) \subset N_{\partial\varphi_d(0)}(\bar{\lambda})$ and, due to the correspondence

$$h \in \mathcal{C}(\bar{x}) \quad \iff \quad DF(\bar{x})h \in N_{\partial\varphi_d(0)}(\bar{\lambda}),$$

it can be easily shown that the set $\mathcal{C}(\bar{x})$ is a subspace if $N_{\partial\varphi_d(0)}(\bar{\lambda})$ is a subspace. Now, on the other hand, let $z_1, z_2 \in N_{\partial\varphi_d(0)}(\bar{\lambda})$ be arbitrary and let us suppose that the critical cone $\mathcal{C}(\bar{x})$ is a subspace. Then, the nondegeneracy condition implies

$$\exists\ h_i \in \mathbb{R}^n,\ y_i \in \operatorname{lin} N_{\partial\varphi_d(0)}(\bar{\lambda}) \quad \text{such that} \quad z_i = DF(\bar{x})h_i + y_i, \quad i = 1, 2.$$

Furthermore, for $i = 1, 2$, we have

$$0 \geq \langle z_i, \lambda - \bar{\lambda} \rangle = \langle DF(\bar{x})h_i + y_i, \lambda - \bar{\lambda} \rangle = \langle DF(\bar{x})h_i, \lambda - \bar{\lambda} \rangle, \quad \forall\ \lambda \in \partial\varphi_d(0).$$

This establishes $DF(\bar{x})h_i \in N_{\partial\varphi_d(0)}(\bar{\lambda})$ and consequently, it follows $h_i \in \mathcal{C}(\bar{x})$ for all $i = 1, 2$. Next, since the sets $\mathcal{C}(\bar{x})$ and $\operatorname{lin} N_{\partial\varphi_d(0)}(\bar{\lambda})$ are subspaces and the normal cone $N_{\partial\varphi_d(0)}(\bar{\lambda})$ is a convex cone, we obtain

$$\begin{aligned} z_1 + z_2 = DF(\bar{x})(h_1 + h_2) + (y_1 + y_2) &\in DF(\bar{x})\mathcal{C}(\bar{x}) + \operatorname{lin} N_{\partial\varphi_d(0)}(\bar{\lambda}) \\ &\subset N_{\partial\varphi_d(0)}(\bar{\lambda}) + N_{\partial\varphi_d(0)}(\bar{\lambda}) \subset N_{\partial\varphi_d(0)}(\bar{\lambda}) \end{aligned}$$

and

$$\alpha z_1 = DF(\bar{x})(\alpha h_1) + (\alpha y_1) \in DF(\bar{x})\mathcal{C}(\bar{x}) + \operatorname{lin} N_{\partial\varphi_d(0)}(\bar{\lambda}) \subset N_{\partial\varphi_d(0)}(\bar{\lambda}),$$

for all $\alpha \in \mathbb{R}$. This shows that the normal cone $N_{\partial\varphi_d(0)}(\bar{\lambda})$ is a subspace and concludes the proof. □

In the following, we state one of the main results of this work.

Theorem 5.3.26 (Curvature via proximity operators). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be convex, proper, and lower semicontinuous. Moreover, let $\bar{x} \in \operatorname{dom} \varphi$ be a stationary point of problem $(\mathcal{P})$ and suppose that φ is C^2-fully decomposable at $\bar{x}$. If the strict complementarity condition holds* *at $\bar{x}$, then the proximity operator* $\mathrm{prox}_\varphi^\Lambda$ *is Fréchet differentiable at* $\bar{u} := \bar{x} - \Lambda^{-1}\nabla f(\bar{x})$ *for every* $\Lambda \in \mathbb{S}^n_{++}$ *and it follows*

$$-\xi^*_{\varphi,h}(-\nabla f(\bar{x})) = \langle h, (\Lambda^{-\frac{1}{2}}\mathcal{Q}^\Lambda_\varphi(\bar{u})^+\Lambda^{\frac{1}{2}} - I)h\rangle_\Lambda, \quad \forall\ h \in \mathcal{C}(\bar{x}),$$

where $\mathcal{Q}^\Lambda_\varphi(\bar{u})^+$ *denotes the pseudoinverse of the matrix* $\mathcal{Q}^\Lambda_\varphi(\bar{u}) := \Lambda^{\frac{1}{2}}D\mathrm{prox}^\Lambda_\varphi(\bar{u})\Lambda^{-\frac{1}{2}}$.

The proof of this theorem relies on a well-considered combination of various results and concepts and beautifully connects different fields in optimization. In particular, the proof consists of the following steps and concepts:

Second order subderivatives.

- At first, the curvature term $\xi^*_{\varphi,h}(-\nabla f(\bar{x}))$, which is based on parabolic second order epiderivatives, will be connected to another, already mentioned, epigraphical framework – the so-called second order subderivatives.

Second order sensitivity analysis of the Moreau envelope.

- Since decomposable functions are outer second order regular and twice epidifferentiable at $\bar{x}$, it can be shown that the proximity operator of φ is directionally differentiable at $\bar{x}$. This result is presented in [27, Example 4.106] and requires a deep background and discussion of differentiability and sensitivity properties of minimum value functions.

- In the case of full decomposability, the directional derivative of the proximity operator can be characterized as the unique minimizer of a specific, convex and quadratic optimization problem. The strict complementarity condition then will imply Fréchet differentiability. Let us note that similar properties were obtained by Shapiro, [212], for metric projections onto convex, cone-reducible sets. Thus, differentiability of the proximity operator can be seen as a "translation" and extension of Shapiro's work to the proximal setting.

$\mathcal{VU}$-theory, sub-Lagrangians and the $\mathcal{U}$-Hessian.

- By using the theory of the quadratic sub-Lagrangian, [103], the strict complementarity condition, and a slightly adapted result of Mifflin and Sagastizábal, [154], it is possible to show that the convex, decomposable function φ admits a so-called $\mathcal{U}$-Hessian at $\bar{x}$ which will be precisely given by the generalized quadratic $\xi^*_{\varphi,h}(-\nabla f(\bar{x}))$.

- Finally, the theory and results developed in [132] and [130, Section 5], connect the (existing) $\mathcal{U}$-Hessian of φ with the $\mathcal{U}$-Hessian of the corresponding Moreau envelope of φ and the derivative of the proximity operator.

In the next paragraphs, we will introduce all relevant and necessary concepts and tools for the proof of Theorem 5.3.26 step by step. We start with a brief discussion of second order subderivatives.

Second order subderivatives

Let $\varrho : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper functional and let $x \in \operatorname{dom} \varrho$ be arbitrary. Then, the so-called *second order subderivative* of ϱ at x relative to $y \in \mathbb{R}^n$ is defined as the following epi-limit

$$\mathrm{d}^2\varrho(x|y) := \operatorname*{e\text{-}lim}_{t \downarrow 0} \, \Delta_t^2 \, \varrho(x|y), \quad \Delta_t^2 \, \varrho(x|y)(h) := \frac{\varrho(x+th) - \varrho(x) - t\langle y, h\rangle}{\frac{1}{2}t^2}.$$

If the latter epi-limit exists and if it holds $\mathrm{d}^2\varrho(x|y)(0) > -\infty$, then ϱ is said to be *twice epi-subdifferentiable* at x relative to y. Alternatively, the second order subderivative can also be defined by means of the lower and upper epi-limits:

$$\mathrm{d}^2_-\varrho(x|y)(h) := \liminf_{t\downarrow 0,\, \tilde{h}\to h} \, \Delta_t^2 \, \varrho(x|y)(\tilde{h}), \quad \mathrm{d}^2_+\varrho(x|y)(h) = \sup_{(t_k)_k \in \mathcal{N}_0} \, \liminf_{k\to\infty,\, \tilde{h}\to h} \, \Delta_{t_k}^2 \, \varrho(x|y)(\tilde{h}).$$

Here, as usual, the limits $\mathrm{d}^2_-\varrho(x|y)(h)$ and $\mathrm{d}^2_+\varrho(x|y)(h)$ denote the *lower* and *upper second order subderivative* of ϱ at x relative to y, in the direction $h \in \mathbb{R}^n$, and ϱ is twice epi-subdifferentiable relative to y if and only if these two limits coincide for all $h \in \mathbb{R}^n$ and it holds $\mathrm{d}^2_-\varrho(x|y)(0) > -\infty$. Of course, in the latter case, the common limit is just the second order subderivative. Let us state some basic properties of second order subderivatives:

- Suppose that ϱ is convex and subdifferentiable at $x \in \operatorname{dom} \varrho$, then it holds

$$\mathrm{d}^2_-\varrho(x|\lambda)(h) \geq \liminf_{t\downarrow 0,\, \tilde{h}\to h} \frac{\langle \lambda, t\tilde{h}\rangle - t\langle \lambda, \tilde{h}\rangle}{\frac{1}{2}t^2} = 0, \quad \forall\, \lambda \in \partial\varrho(x), \quad \forall\, h \in \mathbb{R}^n.$$

- Let ϱ be twice epi-subdifferentiable at $x \in \operatorname{dom} \varrho$ relative to y and let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable in a neighborhood of x, then, for $w := y + \nabla f(x)$ and for all $h \in \mathbb{R}^n$, it holds

$$\mathrm{d}^2(f + \varrho)(x|w)(h) = h^\top \nabla^2 f(x) h + \mathrm{d}^2\varrho(x|y)(h).$$

- Let $\varrho : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, lower semicontinuous, and positively homogeneous function. Then, for all $\lambda \in \partial\varrho(0)$, it holds

$$\mathrm{d}^2\varrho(0|\lambda)(h) = \begin{cases} 0 & \text{if } \varrho(h) = \langle \lambda, h\rangle, \\ +\infty & \text{if } \varrho(h) > \langle \lambda, h\rangle. \end{cases} \tag{5.3.35}$$

Proof. The second claim can be easily shown by using a second order Taylor expansion of f. Now, let us briefly prove the third part. Let $\lambda \in \partial\varrho(0)$ be arbitrary. Then, by Lemma 2.5.13 and (5.3.5), it follows

$$\varrho(h) = \varrho^{\downarrow}(0; h) = \sigma_{\partial\varrho(0)}(h) \geq \langle \lambda, h\rangle, \quad \forall\, h \in \mathbb{R}^n. \tag{5.3.36}$$

Thus, the case "$\varrho(h) < \langle \lambda, h\rangle$" cannot occur. Next, let us consider a point $h \in \mathbb{R}^n$ with $\varrho(h) > \langle \lambda, h\rangle$ and let $(t_k)_k$, $t_k \downarrow 0$, and $(h^k)_k$, $h^k \to h$, be arbitrary. Then, the epi-convergence of the (first order) difference quotients $\Delta_t\, \varrho(0)$ implies

$$\liminf_{k\to\infty}\ \frac{\varrho(0+t_kh^k)-\varrho(0)}{t_k}-\langle\lambda,h^k\rangle=\liminf_{k\to\infty}\ \Delta_{t_k}\,\varrho(0)(h^k)-\langle\lambda,h^k\rangle\geq\varrho^{\downarrow}(0;h)-\langle\lambda,h\rangle>0.$$

Hence, for every pair of sequences $\mathcal{T}:=(t_k)_k$, $\mathcal{H}:=(h^k)_k$, there exists $K(\mathcal{T},\mathcal{H})\in\mathbb{N}$ such that

$$\Delta_{t_k}\,\varrho(0)(h^k)-\langle\lambda,h^k\rangle>0,\quad\forall\ k\geq K(\mathcal{T},\mathcal{H}).$$

This immediately establishes

$$\mathrm{d}^2_{-}\,\varrho(0|\lambda)(h)=\liminf_{t\downarrow0,\,\tilde h\to h}\ \Delta^2_t\,\varrho(0|\lambda)(\tilde h)=\liminf_{t\downarrow0,\,\tilde h\to h}\ \frac{2}{t}(\Delta_t\,\varrho(0)(\tilde h)-\langle\lambda,\tilde h\rangle)=+\infty.$$

Now, on the other hand, let us suppose that $h\in\mathbb{R}^n$ satisfies $\varrho(h)=\langle\lambda,h\rangle$. Again, let $(t_k)_k$, $t_k\downarrow0$, and $(h^k)_k$, $h^k\to h$, be two arbitrary sequences. Then, inequality (5.3.36) implies

$$\liminf_{k\to\infty}\ \Delta^2_{t_k}\,\varrho(0|\lambda)(h^k)=\liminf_{k\to\infty}\ \frac{\varrho(h^k)-\langle\lambda,h^k\rangle}{\frac{1}{2}t_k}\geq0.$$

Moreover, due to $h\in\operatorname{dom}\,\varrho$, the function ϱ is directionally epidifferentiable at h. In particular, there exists $(w^k)_k$, $w^k\to0$, such that

$$\limsup_{k\to\infty}\ \Delta_{t_k}\,\varrho(h)(w^k)\leq\varrho^{\downarrow}(h;0)\leq0.$$

Next, let us consider and define the specific sequence $(h^k)_k$ with $h^k:=h+\frac{1}{2}t_kw^k$, $k\in\mathbb{N}$. It holds

$$\limsup_{k\to\infty}\ \Delta^2_{t_k}\,\varrho(0|\lambda)(h^k)=\limsup_{k\to\infty}\frac{\varrho(h+\frac{1}{2}t_kw^k)-\varrho(h)}{\frac{1}{2}t_k}-\langle\lambda,w^k\rangle\leq\varrho^{\downarrow}(h;0)-\langle\lambda,0\rangle\leq0.$$

Consequently, the epi-convergence of the difference quotients $\Delta^2_t\,\varrho(0|\lambda)$ and formula (5.3.35) follow from Lemma 2.4.3. □

Second order subderivatives were extensively studied by Rockafellar [206, 207] and Poliquin and Rockafellar [193, 194] in the context of amenable functions. A thorough discussion of second order subderivatives can also be found in [208, Chapter 13]. Next, we present a connection between the second order parabolic epiderivative and the subderivative of a fully decomposable function. Let us note that similar results were obtained in [206, Theorems 4.5 and 4.7] or [208, Theorems 13.67] for fully amenable functions.

Lemma 5.3.27. *Let $f:\mathbb{R}^n\to\mathbb{R}$ be a twice continuously differentiable function and let $\varphi:\mathbb{R}^n\to(-\infty,+\infty]$ be convex, proper, and lower semicontinuous. Furthermore, let $\bar x\in\operatorname{dom}\,\varphi$ be a stationary point of problem $(\mathcal{P})$ and suppose that the mapping φ is C^2-fully decomposable at $\bar x$ with decomposition pair (φ_d,F). Then, φ is twice epi-subdifferentiable at $\bar x$ relative to*

$\bar{g} := -\nabla f(\bar{x})$ *and the second order subderivative of* φ *at* $\bar{x}$ *is given by*

$$\mathrm{d}^2\varphi(\bar{x}|\bar{g})(h) = \begin{cases} -\xi^*_{\varphi,h}(\bar{g}) = \langle \bar{\lambda}, D^2F(\bar{x})[h,h]\rangle & \text{if } h \in \mathcal{C}(\bar{x}), \\ +\infty & \text{otherwise,} \end{cases}$$

where $\bar{\lambda} \in \mathcal{M}(\bar{x})$ *denotes the associated, unique Lagrange multiplier of problem* (5.3.33).

Proof. At first, in the case $h \notin \mathcal{C}(\bar{x})$, the same arguments as in the proof of formula (5.3.35) can be used to establish

$$\mathrm{d}^2_-\varphi(\bar{x}|\bar{g})(h) = +\infty,$$

see also [208, Proposition 13.5] for a similar, general result. Next, let $(t_k)_k$, $t_k \downarrow 0$, be arbitrary and suppose that $h \in \mathcal{C}(\bar{x})$ is an element of the critical cone. Moreover, let us consider another arbitrary sequence $(h^k)_k$ with $h^k \to h$. Then, for any $k \in \mathbb{N}$ sufficiently large, the decomposability of φ, a second order Taylor expansion of F at $\bar{x}$, and $\bar{\lambda} \in \mathcal{M}(\bar{x}) \subset \partial\varphi_d(0)$ yield

$$\begin{aligned} \Delta^2_{t_k}\,\varphi(\bar{x}|\bar{g})(h^k) &= \frac{\varphi_d(F(\bar{x}+t_kh^k)) - \varphi_d(F(\bar{x})) - t_k\langle \bar{g}, h^k\rangle}{\frac{1}{2}t_k^2} \\ &\geq \frac{\langle \bar{\lambda}, F(\bar{x}+t_kh^k)\rangle - t_k\langle \bar{\lambda}, DF(\bar{x})h^k\rangle}{\frac{1}{2}t_k^2} = \langle \bar{\lambda}, D^2F(\bar{x})[h^k,h^k]\rangle + o(1). \end{aligned}$$

Of course, by Lemma 5.3.9, this immediately implies

$$\liminf_{k\to\infty}\ \Delta^2_{t_k}\,\varphi(\bar{x}|\bar{g})(h^k) \geq \langle \bar{\lambda}, D^2F(\bar{x})[h,h]\rangle = -\xi^*_{\varphi,h}(\bar{g}).$$

On the other hand, the full decomposability of φ and Lemma 5.3.9 also ensure the existence of a point $\hat{w} \in \mathbb{R}^n$ such that

$$-\xi^*_{\varphi,h}(\bar{g}) = \inf_{w\in\mathbb{R}^n}\ \varphi^{\downarrow\downarrow}(\bar{x};h,w) - \langle \bar{g}, w\rangle = \varphi^{\downarrow\downarrow}(\bar{x};h,\hat{w}) - \langle \bar{g}, \hat{w}\rangle.$$

Furthermore, by Lemma 5.3.5, we know that the mapping φ is twice (parabolically) directionally epidifferentiable at $\bar{x}$, in the direction $h \in \mathcal{C}(\bar{x})$. Hence, the epi-convergence of the (parabolic) difference quotients $\Delta^2_t\,\varphi(\bar{x};h)$ and Lemma 2.4.3 imply that there exists a sequence $(w^k)_k$, $w^k \to \hat{w}$, such that

$$\limsup_{k\to\infty}\ \Delta^2_{t_k}\,\varphi(\bar{x};h)(w^k) = \limsup_{k\to\infty}\ \frac{\varphi(\bar{x}+t_kh+\frac{1}{2}t_k^2w^k) - \varphi(\bar{x}) - t_k\varphi^{\downarrow}(\bar{x};h)}{\frac{1}{2}t_k^2} \leq \varphi^{\downarrow\downarrow}(\bar{x};h,\hat{w}).$$

Now, as in the last proof, let us define $h^k := h + \frac{1}{2}t_kw^k$. Then, clearly, it holds $h^k \to h$ and we obtain

$$\begin{aligned} \limsup_{k\to\infty}\ \Delta^2_{t_k}\,\varphi(\bar{x}|\bar{g})(h^k) &= \limsup_{k\to\infty}\ \left\{\Delta^2_{t_k}\,\varphi(\bar{x};h)(w^k) - \langle \bar{g}, w^k\rangle\right\} \\ &\leq \varphi^{\downarrow\downarrow}(\bar{x};h,\hat{w}) - \langle \bar{g}, \hat{w}\rangle = -\xi^*_{\varphi,h}(\bar{g}), \end{aligned}$$

where we used $h \in \mathcal{C}(\bar{x})$. Thus, for every sequence $(t_k)_k$, $t_k \downarrow 0$, it follows

$$\begin{cases} \liminf\limits_{k\to\infty} \ \Delta^2_{t_k} \varphi(\bar{x}|\bar{g})(h^k) \ \geq -\xi^*_{\varphi,h}(\bar{g}) & \text{for every sequence } h^k \to h, \\ \limsup\limits_{k\to\infty} \ \Delta^2_{t_k} \varphi(\bar{x}|\bar{g})(h^k) \leq -\xi^*_{\varphi,h}(\bar{g}) & \text{for some sequence } h^k \to h. \end{cases}$$

By Lemma 2.4.3, this shows that φ is twice epi-subdifferentiable at $\bar{x}$ relative to $\bar{g}$ with $\mathrm{d}^2\varphi(\bar{x}|\bar{g})(h) = -\xi^*_{\varphi,h}(\bar{g})$ for all $h \in \mathcal{C}(\bar{x})$. □

Remark 5.3.28. Let us note that the first part of this proof is based on [206, Theorem 4.5], while the derivation of the "limsup-inequality" is motivated by the proof of [27, Proposition 3.103]. However, we also want to mention that our argumentation might not be optimal since it requires full decomposability of φ and existence of a maximizer $\hat{w}$ of the curvature term

$$-\xi^*_{\varphi,h}(\bar{g}) = -\sup_{w} \ \langle \bar{g}, w\rangle - \varphi^{\downarrow\downarrow}(\bar{x}; h, w) = \varphi^{\downarrow\downarrow}(\bar{x}; h, \hat{w}) - \langle \bar{g}, \hat{w}\rangle.$$

It is an interesting question whether a direct discussion of the second order subderivative as in [206] can lead to similar and more general results for C^2-decomposable functions.

Second order sensitivity analysis of the Moreau envelope

In Lemma 3.1.5, we have already seen that the Moreau envelope $\mathrm{env}^\Lambda_\varphi$, $\Lambda \in \mathbb{S}^n_{++}$, of a convex, proper, and lower semicontinuous function $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is convex and continuously differentiable. Moreover, in section 3.3, we have discussed several second order properties of the proximity operator $\mathrm{prox}^\Lambda_\varphi$ that can be primarily traced back to the convexity of the Moreau envelope and to the firm nonexpansiveness of $\mathrm{prox}^\Lambda_\varphi$. In this subsection, we will additionally assume that the mapping φ is outer second order regular and twice directionally epidifferentiable at a certain point of interest. This extra information will then allow us to refine our basic differentiability results and to establish general, first and second order directional differentiability of the proximity operator and of the Moreau envelope, respectively.

The theoretical statements in this paragraph are essentially based on [23, Section 7.3] and [27, Example 4.106]. However, let us emphasize that the results of Bonnans, Cominetti, and Shapiro, [23, 27], rely on a number of involved second order sensitivity results for the minimum value function of a general optimization problem. Here, in our specific situation, these results and the corresponding proofs can be expressed in a compact, self-contained, and simplified form, which will be presented in the following. For a more abstract formulation and more details on second order sensitivity analysis we refer to [23] and [27, Section 4.7].

In the following, we will always assume that $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous function. Furthermore, we consider the specific problem

$$\min_{x} \ \varphi(x) + \frac{1}{2}\|\bar{u} - x\|^2_\Lambda, \tag{5.3.37}$$

where $\bar{u} \in \mathbb{R}^n$ is a fixed point and $\Lambda \in \mathbb{S}^n_{++}$ is an arbitrary parameter matrix. The unique, optimal solution of this problem is given by the proximity operator $\bar{p} := \mathrm{prox}^\Lambda_\varphi(\bar{u})$ and the optimal value of problem (5.3.37) coincides with the Moreau envelope $\mathrm{env}_\varphi^\Lambda(\bar u)$. In particular, it holds

$$\mathrm{env}_\varphi^\Lambda(\bar u) = \varphi(\bar p) + \frac{1}{2}\|\bar u - \bar p\|_\Lambda^2. \tag{5.3.38}$$

By Remark 3.1.6, the function φ is subdifferentiable at $\bar p$ and the corresponding first order optimality conditions take the following form

$$\varphi^\downarrow(\bar p; h) + \langle \Lambda(\bar p - \bar u), h\rangle = \varphi^\downarrow(\bar p; h) - \langle \nabla \mathrm{env}_\varphi^\Lambda(\bar u), h\rangle \geq 0, \quad \forall\, h \in \mathbb{R}^n. \tag{5.3.39}$$

Thus, the critical cone associated with problem (5.3.37) can be defined as follows:

$$\mathcal{C}_{\bar u}^\Lambda(\bar p) := \{h \in \mathbb{R}^n : \varphi^\downarrow(\bar p; h) - \nabla \mathrm{env}_\varphi^\Lambda(\bar u)^\top h = 0\} = N_{\partial\varphi(\bar p)}(\nabla \mathrm{env}_\varphi^\Lambda(\bar u)).$$

In the next lemmas, we want to analyze stability and sensitivity properties of problem (5.3.37) and of the Moreau envelope $\mathrm{env}_\varphi^\Lambda$ along parabolic paths of the form

$$u(t) := \bar u + td + \frac{1}{2}t^2 r + o(t^2), \quad d, r \in \mathbb{R}^n.$$

By adapting the proof of [27, Theorem 4.100] and by using the concept of outer second order regularity, we obtain the following result.

Lemma 5.3.29. *Let $\Lambda \in \mathbb{S}_{++}^n$ be arbitrary and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be convex, proper and lower semicontinuous. Suppose that the function φ is outer second order regular at $\bar p = \mathrm{prox}_\varphi^\Lambda(\bar u)$ in all directions $h \in \mathcal{C}_{\bar u}^\Lambda(\bar p)$. Then, it holds*

$$\begin{aligned}\liminf_{t\downarrow 0}\ & \frac{\mathrm{env}_\varphi^\Lambda(u(t)) - \mathrm{env}_\varphi^\Lambda(\bar u) - \nabla \mathrm{env}_\varphi^\Lambda(\bar u)^\top d}{\frac{1}{2}t^2} \\ & \geq \langle \nabla \mathrm{env}_\varphi^\Lambda(\bar u), r\rangle + \min_{h \in \mathcal{C}_{\bar u}^\Lambda(\bar p)} \left\{\|d - h\|_\Lambda^2 - \xi_{\varphi,h}^*(\nabla \mathrm{env}_\varphi^\Lambda(\bar u))\right\},\end{aligned} \tag{5.3.40}$$

where $\xi_{\varphi,h}(\cdot) = \varphi_-^{\downarrow\downarrow}(\bar p; h, \cdot)$.

Proof. Let $(t_k)_k$, $t_k \downarrow 0$, be a sequence such that the limes inferior on the left side of (5.3.40) is attained as $(t_k)_k$ converges to zero. Furthermore, let us set $u^k = u(t_k)$, $p^k = \mathrm{prox}_\varphi^\Lambda(u^k)$, and $\bar p := \mathrm{prox}_\varphi^\Lambda(\bar u)$. By defining $h^k := t_k^{-1}(p^k - \bar p)$, it holds $p^k = \bar p + t_k h^k$ and

$$\|h^k\|_\Lambda = \frac{1}{t_k}\|p^k - \bar p\|_\Lambda \leq \frac{1}{t_k}\|u^k - \bar u\|_\Lambda \leq \|d + \tfrac{1}{2}t_k r + o(t_k)\|_\Lambda,$$

where we used the Lipschitz continuity of the proximity operator. Consequently, the sequence $(h^k)_k$ is bounded and there exists $h \in \mathbb{R}^n$ and a subsequence of $(h^k)_k$ that converges to h. In the following, without loss of generality, we will drop the additional index of the subsequence for a better readability. By further setting $w^k := 2t_k^{-1}(h^k - h)$, the proximal path p^k can be written in the form

$$p^k = \bar p + t_k h + \frac{1}{2}t_k^2 \cdot [2t_k^{-1}(h^k - h)] = \bar p + t_k h + \frac{1}{2}t_k^2 w^k.$$

Now, a simple calculation yields

$$\begin{aligned}
&\|p^k - u^k\|_\Lambda^2 \\
&\quad = \|\bar{p} - \bar{u} - t_k(d - h + \tfrac{1}{2}t_k(r - w^k) - o(t_k))\|_\Lambda^2 \\
&\quad = \|\bar{p} - \bar{u}\|_\Lambda^2 - 2t_k\langle \Lambda(\bar{p} - \bar{u}), d - h + \tfrac{1}{2}t_k(r - w^k)\rangle \\
&\qquad\quad + t_k^2\|d - h + \tfrac{1}{2}t_k(r - w^k) - o(t_k)\|_\Lambda^2 + o(t_k^2) \\
&\quad = \|\bar{p} - \bar{u}\|_\Lambda^2 + 2t_k\langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), d - h\rangle + t_k^2\left\{\langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), r - w^k\rangle + \|d - h\|_\Lambda^2\right\} + o(t_k^2).
\end{aligned}$$

Hence, by applying the definition of the Moreau envelope (5.3.38), we obtain

$$\begin{aligned}
&\mathrm{env}_\varphi^\Lambda(u^k) - \mathrm{env}_\varphi^\Lambda(\bar{u}) - t_k\nabla\mathrm{env}_\varphi^\Lambda(\bar{u})^\top d \\
&\quad = \varphi(p^k) - \varphi(\bar{p}) + \frac{1}{2}\|p^k - u^k\|_\Lambda^2 - \frac{1}{2}\|\bar{p} - \bar{u}\|_\Lambda^2 - t_k\langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), d\rangle \\
&\quad = \varphi(p^k) - \varphi(\bar{p}) - t_k\langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), h\rangle + \frac{1}{2}t_k^2\left\{\langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), r - w^k\rangle + \|d - h\|_\Lambda^2\right\} + o(t_k^2).
\end{aligned}$$

Dividing both sides by t_k and taking the limit, $k \to \infty$, this establishes

$$\begin{aligned}
0 &= \lim_{k\to\infty}\left\{\frac{\mathrm{env}_\varphi^\Lambda(u^k) - \mathrm{env}_\varphi^\Lambda(\bar{u})}{t_k} - \nabla\mathrm{env}_\varphi^\Lambda(\bar{u})^\top d\right\} \\
&= \liminf_{k\to\infty}\left\{\frac{\varphi(p^k) - \varphi(\bar{p})}{t_k} - \langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), h\rangle + o(t_k)\right\} \\
&\geq \liminf_{k\to\infty}\left\{\varphi^\downarrow(\bar{p}; h + \tfrac{1}{2}t_kw^k) + o(t_k)\right\} - \langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), h\rangle \geq \varphi^\downarrow(\bar{p}; h) - \langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), h\rangle,
\end{aligned}$$

where we used $t_kw^k \to 0$, the convexity of φ, and the properties of the epiderivative $\varphi^\downarrow(\bar{p}; \cdot)$. On the other hand, since $\bar{p}$ is a solution of the minimization problem (5.3.37), the optimality conditions (5.3.39) imply

$$\varphi^\downarrow(\bar{p}; h) - \langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), h\rangle = 0$$

and, consequently, it follows $h \in \mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$. Next, by combining the last results, we get

$$\varphi(p^k) - \varphi(\bar{p}) - t_k\varphi^\downarrow(\bar{p}; h) = \tfrac{1}{2}t_k^2\tau^k,$$

where the remainder τ^k is defined as $\tau^k := 2t_k^{-2}[\mathrm{env}_\varphi^\Lambda(u^k) - \mathrm{env}_\varphi^\Lambda(\bar{u}) - \nabla\mathrm{env}_\varphi^\Lambda(\bar{u})^\top d] - \langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), r - w^k\rangle - \|d - h\|_\Lambda^2 + o(1)$ and satisfies $t_k\tau^k \to 0$. Thus, since φ is outer second order regular at $\bar{p}$ in all directions $h \in \mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$, there exist sequences $(\tilde{w}^k)_k$ and $(\tilde{\tau}^k)_k$ such that $\tilde{w}^k - w^k \to 0$, $\tilde{\tau}^k - \tau^k \to 0$, and $\tilde{\tau}^k \geq \varphi_-^{\downarrow\downarrow}(\bar{p}; h, \tilde{w}^k)$. Finally, we have

$$\begin{aligned}
&\frac{\mathrm{env}_\varphi^\Lambda(u^k) - \mathrm{env}_\varphi^\Lambda(\bar{u}) - \nabla\mathrm{env}_\varphi^\Lambda(\bar{u})^\top d}{\tfrac{1}{2}t_k^2} \\
&\qquad\qquad \geq \varphi_-^{\downarrow\downarrow}(\bar{p}; h, \tilde{w}^k) + \langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), r - w^k\rangle + \|d - h\|_\Lambda^2 + o(1) \\
&\qquad\qquad \geq \langle \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), r\rangle + \|d - h\|_\Lambda^2 - \xi_{\varphi,h}^*(\nabla\mathrm{env}_\varphi^\Lambda(\bar{u})) + o(1).
\end{aligned}$$

Moreover, by taking the limit $k \to \infty$ over both sides of the latter inequality and using $h \in \mathcal{C}_{\bar{u}}^{\Lambda}(\bar{p})$, we clearly obtain (5.3.40). This concludes the proof of Lemma 5.3.29. □

In order to derive an upper bound for the parabolic difference quotient in (5.3.40), we will now discuss the second order behaviour of the Moreau envelope $\mathrm{env}_{\varphi}^{\Lambda}$ along fixed proximal paths

$$p(t) := \bar{p} + th + \frac{1}{2}t^2 w + o(t^2), \quad \bar{p} = \mathrm{prox}_{\varphi}^{\Lambda}(\bar{u}), \quad h \in \mathcal{C}_{\bar{u}}^{\Lambda}(\bar{p}), \tag{5.3.41}$$

where $w \in \mathbb{R}^n$ and the $o(t^2)$-term are chosen such that the path $p(t)$ satisfies $p(t) \in \mathrm{dom}\,\varphi$ for all $t \geq 0$ sufficiently small. Due to the definition of the inner second order tangent set and Lemma 5.2.4, it follows for all $t \geq 0$

$$\begin{pmatrix} \bar{p} \\ \varphi(\bar{p}) \end{pmatrix} + t \begin{pmatrix} h \\ \varphi^{\downarrow}(\bar{p}; h) \end{pmatrix} + \frac{1}{2}t^2 \begin{pmatrix} w \\ \varphi_{+}^{\downarrow\downarrow}(\bar{p}; h, w) \end{pmatrix} + o(t^2) \in \mathrm{epi}\,\varphi. \tag{5.3.42}$$

Hence, since the epiderivative $\varphi^{\downarrow}(\bar{p}; h)$ is finite, the path $p(t)$ is feasible if there exists $w \in \mathbb{R}^n$ with $\varphi_{+}^{\downarrow\downarrow}(\bar{p}; h, w) < \infty$. The following result is based on [27, Proposition 4.83].

Lemma 5.3.30. *Let $\Lambda \in \mathbb{S}_{++}^n$ be arbitrary and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper and lower semicontinuous function. Then, it holds*

$$\begin{aligned} \limsup_{t \downarrow 0} \; & \frac{\mathrm{env}_{\varphi}^{\Lambda}(u(t)) - \mathrm{env}_{\varphi}^{\Lambda}(\bar{u}) - \nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u})^\top d}{\frac{1}{2}t^2} \\ & \leq \langle \nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u}), r \rangle + \min_{h \in \mathcal{C}_{\bar{u}}^{\Lambda}(\bar{p})} \left\{ \|d - h\|_{\Lambda}^2 - \zeta_{\varphi,h}^{*}(\nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u})) \right\}, \end{aligned} \tag{5.3.43}$$

where $\zeta_{\varphi,h}(\cdot) = \varphi_{+}^{\downarrow\downarrow}(\bar{p}; h, \cdot)$.

Proof. Clearly, if the upper second order directional epiderivative $\varphi_{+}^{\downarrow\downarrow}(\bar{p}; h, \cdot)$ is infinite for all $h \in \mathcal{C}_{\bar{u}}^{\Lambda}(\bar{p})$, then, due to

$$-\zeta_{\varphi,h}^{*}(\nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u})) = \inf_{w} \; \varphi_{+}^{\downarrow\downarrow}(\bar{p}; h, w) - \langle \nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u}), w \rangle,$$

the right side of (5.3.43) equals $+\infty$ and the inequality in Lemma 5.3.30 is trivially satisfied. Otherwise, there exist $h \in \mathcal{C}_{\bar{u}}^{\Lambda}(\bar{p})$ and $w \in \mathbb{R}^n$ such that we can construct a feasible proximal path $p(t)$ of the form (5.3.41). Then, by using the calculations of the proof of Lemma 5.3.29 and (5.3.42), it follows

$$\begin{aligned} & \mathrm{env}_{\varphi}^{\Lambda}(u(t)) - \mathrm{env}_{\varphi}^{\Lambda}(\bar{u}) - t\nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u})^\top d \\ & \quad \leq \varphi(p(t)) - \varphi(\bar{p}) + \frac{1}{2}\|p(t) - u(t)\|_{\Lambda}^2 - \frac{1}{2}\|\bar{p} - \bar{u}\|_{\Lambda}^2 - t\langle \nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u}), d \rangle \\ & \quad = \varphi(p(t)) - \varphi(\bar{p}) - t\langle \nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u}), h \rangle + \frac{1}{2}t^2 \left\{ \langle \nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u}), r - w \rangle + \|d - h\|_{\Lambda}^2 \right\} + o(t^2) \\ & \quad \leq \frac{1}{2}t^2 \left\{ \varphi_{+}^{\downarrow\downarrow}(\bar{p}; h, w) + \langle \nabla \mathrm{env}_{\varphi}^{\Lambda}(\bar{u}), r - w \rangle + \|d - h\|_{\Lambda}^2 \right\} + o(t^2). \end{aligned}$$

By taking the infimum with respect to w and minimizing over $h \in \mathcal{C}^{\Lambda}_{\bar{u}}(\bar{p})$, we readily obtain inequality (5.3.43). □

Before continuing with the next theorem, we want to discuss the minimization problem, which occurs in the inequalities (5.3.40) and (5.3.43), and the corresponding convex conjugates $\xi^*_{\varphi,h}$ and $\zeta^*_{\varphi,h}$ in some more detail. Therefore, let us suppose that the function φ is additionally outer second order regular and twice directionally epidifferentiable at $\bar{p} = \mathrm{prox}^{\Lambda}_{\varphi}(\bar{u})$ in some direction $h \in \mathcal{C}^{\Lambda}_{\bar{u}}(\bar{p})$. Furthermore, let $(t_k)_k$, $t_k \downarrow 0$, be an arbitrary sequence. Then, due to the epi-convergence of the first order difference quotients $\Delta_t\, \varphi(\bar{p})$, there exists $(h^k)_k$, $h^k \to h$, such that

$$\limsup_{k\to\infty} \frac{\varphi(\bar{p} + t_k h^k) - \varphi(\bar{p})}{t_k} \le \varphi^{\downarrow}(\bar{p}; h) \in \mathbb{R}.$$

Next, let us define

$$w^k := 2t_k^{-1}[h^k - h], \quad \tau^k := 2t_k^{-2}[\varphi(\bar{p} + t_k h + \tfrac{1}{2}t_k^2 w^k) - \varphi(\bar{p}) - t_k\varphi^{\downarrow}(\bar{p}; h)].$$

Clearly, this yields $t_k w^k \to 0$ and

$$\varphi(\bar{p} + t_k h + \tfrac{1}{2}t_k^2 w^k) = \varphi(\bar{p}) + t_k\varphi^{\downarrow}(\bar{p}; h) + \tfrac{1}{2}t_k^2\tau^k.$$

Using Lemma 2.5.5, and the positive homogeneity and lower semicontinuity of the epiderivative $\varphi^{\downarrow}(\bar{p}; \cdot)$, we obtain

$$\begin{aligned}\liminf_{k\to\infty}\ t_k\tau^k &\ge \liminf_{k\to\infty}\ \frac{\varphi^{\downarrow}(\bar{p}; t_k h + \frac{1}{2}t_k^2 w^k) - t_k\varphi^{\downarrow}(\bar{p}; h)}{\frac{1}{2}t_k} \\ &= \liminf_{k\to\infty}\ 2(\varphi^{\downarrow}(\bar{p}; h + \tfrac{1}{2}t_k w^k) - \varphi^{\downarrow}(\bar{p}; h)) \ge 0.\end{aligned}$$

On the other hand, we have

$$\limsup_{k\to\infty}\ t_k\tau^k = \limsup_{k\to\infty}\ \frac{\varphi(\bar{p} + t_k h^k) - \varphi(\bar{p})}{\frac{1}{2}t_k} - 2\varphi^{\downarrow}(\bar{p}; h) \le 2\varphi^{\downarrow}(\bar{p}; h) - 2\varphi^{\downarrow}(\bar{p}; h) = 0.$$

Thus, it follows $t_k\tau^k \to 0$ and, due to the twice epidifferentiability and outer second order regularity of φ, there exist $(\tilde{w}^k)_k$, $\tilde{w}^k - w^k \to 0$, and $(\tilde{\tau}^k)_k$, $\tilde{\tau}^k - \tau^k \to 0$, such that

$$\varphi^{\downarrow\downarrow}(\bar{p}; h, \tilde{w}^k) = \varphi^{\downarrow\downarrow}_{+}(\bar{p}; h, \tilde{w}^k) = \varphi^{\downarrow\downarrow}_{-}(\bar{p}; h, \tilde{w}^k) \le \tilde{\tau}^k.$$

Consequently, in this case, we have

$$-\zeta^*_{\varphi,h}(\nabla\mathrm{env}^{\Lambda}_{\varphi}(\bar{u})) \le \tilde{\tau}^k - \langle\nabla\mathrm{env}^{\Lambda}_{\varphi}(\bar{u}), \tilde{w}^k\rangle < +\infty.$$

Moreover, as in Remark 5.2.9, it follows

$$\varphi^{\downarrow\downarrow}_{-}(\bar{p}; h, w) \ge \liminf_{t\downarrow 0,\, \tilde{w}\to w} \frac{\langle\nabla\mathrm{env}^{\Lambda}_{\varphi}(\bar{u}), th + \frac{1}{2}t^2\tilde{w}\rangle - t\varphi^{\downarrow}(\bar{p}; h)}{\frac{1}{2}t^2} = \langle\nabla\mathrm{env}^{\Lambda}_{\varphi}(\bar{u}), w\rangle,$$

and

$$-\xi^*_{\varphi,h}(\nabla \mathrm{env}^\Lambda_\varphi(\bar{u})) = \inf_w \ \varphi^{\downarrow\downarrow}_-(\bar{p}; h, w) - \langle \nabla \mathrm{env}^\Lambda_\varphi(\bar{u}), w\rangle \geq 0.$$

Altogether, this finally implies

$$-\xi^*_{\varphi,h}(\nabla \mathrm{env}^\Lambda_\varphi(\bar{u})) = -\zeta^*_{\varphi,h}(\nabla \mathrm{env}^\Lambda_\varphi(\bar{u})) \in [0, +\infty).$$

Hence, in summary, if φ is outer second order regular and twice directionally epidifferentiable at $\bar{p}$ in every direction $h \in \mathcal{C}^\Lambda_{\bar{u}}(\bar{p})$, then the objective function of the minimization problem

$$\min_{h \in \mathcal{C}^\Lambda_{\bar{u}}(\bar{p})} \ \|d-h\|^2_\Lambda - \xi^*_{\varphi,h}(\nabla \mathrm{env}^\Lambda_\varphi(\bar{u})), \quad \xi_{\varphi,h}(w) := \varphi^{\downarrow\downarrow}(\bar{p}; h, w)$$

is real valued on the critical cone $\mathcal{C}^\Lambda_{\bar{u}}(\bar{p})$. Now, by combining Lemma 5.3.29 and 5.3.30 and our latter observations, we obtain the following theorem.

Theorem 5.3.31. *Let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be convex, proper, and lower semicontinuous. Furthermore, suppose that φ is outer second order regular and twice directionally epidifferentiable at $\bar{p} = \mathrm{prox}^\Lambda_\varphi(\bar{u})$ in all directions $h \in \mathcal{C}^\Lambda_{\bar{u}}(\bar{p})$. Then, the second order directional derivative $(\mathrm{env}^\Lambda_\varphi)''(\bar{u}; d, r)$ exists and it holds*

$$(5.3.44) \qquad (\mathrm{env}^\Lambda_\varphi)''(\bar{u}; d, r) = \langle \nabla \mathrm{env}^\Lambda_\varphi(\bar{u}), r\rangle + \min_{h \in \mathcal{C}^\Lambda_{\bar{u}}(\bar{p})} \ \left\{ \|d-h\|^2_\Lambda - \xi^*_{\varphi,h}(\nabla \mathrm{env}^\Lambda_\varphi(\bar{u})) \right\},$$

where $\xi_{\varphi,h}(\cdot) = \varphi^{\downarrow\downarrow}(\bar{p}; h, \cdot)$. Moreover, the proximity operator $\mathrm{prox}^\Lambda_\varphi$ is directionally differentiable at $\bar{u}$ and its derivative satisfies

$$(\mathrm{prox}^\Lambda_\varphi)'(\bar{u}; d) = \operatorname*{arg\,min}_{h \in \mathcal{C}^\Lambda_{\bar{u}}(\bar{p})} \ \left\{ \|d-h\|^2_\Lambda - \xi^*_{\varphi,h}(\nabla \mathrm{env}^\Lambda_\varphi(\bar{u})) \right\}.$$

Proof. At first, we want to note that the following proof is based on the proofs of [24, Theorem 4.1 and Corollary 4.1] and [27, Theorem 4.101]. Again, we will tailor the abstract and general results in [24, 27] to our specific situation.

The second order directional differentiability of the Moreau envelope $\mathrm{env}^\Lambda_\varphi$ and formula (5.3.44) follow directly from Lemma 5.3.29 and 5.3.30, the continuity of $\mathrm{env}^\Lambda_\varphi$, and the fact that the objective function

$$h \mapsto \Gamma_d(h) := \|d-h\|^2_\Lambda - \xi^*_{\varphi,h}(\nabla \mathrm{env}^\Lambda_\varphi(\bar{u}))$$

is real valued on $\mathcal{C}^\Lambda_{\bar{u}}(\bar{p})$. (In particular, the term on the right side of equation (5.3.44) is a real number). Furthermore, Lemma 5.2.12 implies that the mapping $h \mapsto -\xi^*_{\varphi,h}(\nabla \mathrm{env}^\Lambda_\varphi(\bar{u}))$ is convex on the critical cone $\mathcal{C}^\Lambda_{\bar{u}}(\bar{p}) \subset \mathrm{dom}\ \varphi^\downarrow(\bar{p}; \cdot)$. Hence, the function Γ_d is strongly convex on $\mathcal{C}^\Lambda_{\bar{u}}(\bar{p})$ and the optimization problem

$$(5.3.45) \qquad \min_{h \in \mathcal{C}^\Lambda_{\bar{u}}(\bar{p})} \ \Gamma_d(h)$$

has a unique solution $\hat{h} \in \mathcal{C}^\Lambda_{\bar{u}}(\bar{p})$. Now, let $(t_k)_k$, $t_k \downarrow 0$, be an arbitrary sequence. Then, by combining Lemma 5.3.29 and 5.3.30 and by reconsidering the proof of Lemma 5.3.29, we see that any accumulation point $\bar{h} \in \mathbb{R}^n$ of the sequence $(q^k)_k$,

$$q^k := \frac{\text{prox}_\varphi^\Lambda(u(t_k)) - \text{prox}_\varphi^\Lambda(\bar{u})}{t_k}, \quad k \in \mathbb{N},$$

satisfies the following inequality

$$\Gamma_d(\bar{h}) \le \limsup_{k\to\infty} \frac{\text{env}_\varphi^\Lambda(u(t_k)) - \text{env}_\varphi^\Lambda(\bar{u}) - t_k \nabla \text{env}_\varphi^\Lambda(\bar{u})^\top d}{\frac{1}{2}t_k^2} - \nabla\text{env}_\varphi^\Lambda(\bar{u})^\top r \le \min_{h\in\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})} \Gamma_d(h).$$

Thus, the point $\bar{h}$ is a solution of the minimization problem (5.3.45) and it immediately follows $\bar{h} = \hat{h}$. Moreover, since $(q^k)_k$ is bounded and $\bar{h}$ was an arbitrary accumulation point, the sequence $(q_k)_k$ has to converge to the unique solution $\hat{h}$ of problem (5.3.45). Finally, since the proximity operator is a Lipschitz continuous function and the sequence $(t_k)_k$ was also arbitrarily chosen, we obtain

$$(\text{prox}_\varphi^\Lambda)'(\bar{u}; d) = \lim_{t\downarrow 0} \frac{\text{prox}_\varphi^\Lambda(u(t)) - \text{prox}_\varphi^\Lambda(\bar{u})}{t} = \hat{h} = \operatorname*{arg\,min}_{h\in\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})} \Gamma_d(h),$$

as desired. □

Next, we are going to combine the results of Theorem 5.3.31, the full decomposability of φ, and the strict complementarity condition to establish Fréchet differentiability of the proximity operator $\text{prox}_\varphi^\Lambda$. Let us mention that a similar result was already proven by Shapiro, [212, Proposition 3.1], for projections onto cone-reducible sets. The proof of the following Lemma is motivated by the ideas in [212].

Lemma 5.3.32. *Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping and let $\bar{u} \in \mathbb{R}^n$ and $\Lambda \in \mathbb{S}_{++}^n$ be arbitrary. Moreover, suppose that φ is C^2-fully decomposable at $\bar{p} := \text{prox}_\varphi^\Lambda(\bar{u})$. Then, the proximity operator $\text{prox}_\varphi^\Lambda$ is Fréchet differentiable at $\bar{u}$ if and only if the following strict complementarity condition is satisfied*

$$\nabla\text{env}_\varphi^\Lambda(\bar{u}) \in \text{ri } \partial\varphi(\bar{p}).$$

Proof. Let (φ_d, F) be a corresponding decomposition pair of φ. Since φ is C^2-fully decomposable at $\bar{p}$, Lemma 5.3.5 implies that φ is twice directionally epidifferentiable and outer second order regular at $\bar{p}$ in all directions $h \in \mathbb{R}^n$ with $DF(\bar{p})h \in \text{dom } \varphi_d$. Now, let $h \in \mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$ be arbitrary, then it follows

$$0 = \varphi^\downarrow(\bar{p}; h) - \langle \nabla\text{env}_\varphi^\Lambda(\bar{u}), h\rangle = \varphi_d(DF(\bar{p})h) - \langle \nabla\text{env}_\varphi^\Lambda(\bar{u}), h\rangle.$$

Clearly, this shows $DF(\bar{p})\mathcal{C}_{\bar{u}}^\Lambda(\bar{p}) \subset \text{dom } \varphi_d$ and, consequently, Theorem 5.3.31 is applicable and we can infer that the proximity operator $\text{prox}_\varphi^\Lambda$ is directionally differentiable at $\bar{u}$. Moreover, since $\text{prox}_\varphi^\Lambda$ is a Lipschitz continuous function, it is also directionally differentiable in the *Hadamard sense* and its directional derivative $(\text{prox}_\varphi^\Lambda)'(\bar{u}; \cdot)$ is Lipschitz continuous.

Hence, the proximity operator $\text{prox}_\varphi^\Lambda$ is Fréchet differentiable at $\bar{u}$ if and only if its directional derivative $(\mathrm{prox}_\varphi^\Lambda)'(\bar{u};\cdot)$ is a linear mapping. (Let us refer to [27, Section 2.2.1] for more details on Hadamard and Fréchet differentiability). Furthermore, via identifying $f(x) \equiv \frac{1}{2}\|x-\bar{u}\|_\Lambda^2$, $\bar{x} \equiv \bar{p}$, $-\nabla f(\bar{x}) \equiv \nabla \mathrm{env}_\varphi^\Lambda(\bar{u})$, and $\mathcal{C}(\bar{x}) \equiv \mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$, Lemma 5.3.9 implies

$$-\xi^*_{\varphi,h}(\nabla \mathrm{env}_\varphi^\Lambda(\bar{u})) = \langle \bar{\lambda}, D^2F(\bar{p})[h,h] \rangle, \quad \bar{\lambda} \in \mathcal{M}(\bar{p}), \quad \forall\, h \in \mathcal{C}_{\bar{u}}^\Lambda(\bar{p}).$$

Thus, by using Theorem 5.3.31, the directional derivative $(\mathrm{prox}_\varphi^\Lambda)'(\bar{u};d)$ is the unique, optimal solution of the following, strongly convex and quadratic program

$$\text{(5.3.46)} \qquad \min_{h \in \mathcal{C}_{\bar{u}}^\Lambda(\bar{p})} \ \|d-h\|_\Lambda^2 + h^\top \mathcal{H}_\varphi(\bar{p})h, \qquad \mathcal{H}_\varphi(\bar{p}) := \sum_{i=1}^m \bar{\lambda}_i \nabla^2 F_i(\bar{p}).$$

We can now proceed as in the proof of [212, Proposition 3.1]. Since the critical cone $\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$ coincides with the normal cone $N_{\partial\varphi(\bar{p})}(\nabla \mathrm{env}_\varphi^\Lambda(\bar{u}))$, we can utilize Lemma 5.1.10 and consequently, we only need to verify that the linearity of the mapping $(\mathrm{prox}_\varphi^\Lambda)'(\bar{u};\cdot)$ is equivalent to $\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$ being a subspace. Additionally, since the set $\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$ is a convex, nonempty, and closed cone and by defining $\varphi(h) := \iota_{\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})}(h)$, the optimization problem (5.3.46) is of the form $(\mathcal{P})$ and the general first order optimality theory of section 4.1 is applicable. Therefore, let us first suppose that the critical cone $\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$ is a subspace and let us set $\hat{h}_i := (\mathrm{prox}_\varphi^\Lambda)'(\bar{u};d_i)$, $d_i \in \mathbb{R}^n$, $i=1,2$. Then, due to

$$N_{\mathcal{C}_{\bar{u}}^\Lambda(\bar{u})}(\hat{h}_i) = [\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})]^\circ \cap \{\hat{h}_i\}^\perp = \mathcal{C}_{\bar{u}}^\Lambda(\bar{p})^\perp,$$

the corresponding first order optimality conditions for problem (5.3.46) reduce to

$$\langle \Lambda(\hat{h}_i - d_i) + \mathcal{H}_\varphi(\bar{p})\hat{h}_i, h \rangle = 0, \quad \forall\, h \in \mathcal{C}_{\bar{u}}^\Lambda(\bar{p}), \quad \forall\, i,$$

where we used Example 2.1.5, Example 2.5.16, Lemma 4.1.2, and the fact that $\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$ is subspace. Obviously, this establishes

$$(\mathrm{prox}_\varphi^\Lambda)'(\bar{u};\alpha d_1 + \beta d_2) = \alpha \hat{h}_1 + \beta \hat{h}_2, \quad \forall\, \alpha, \beta \in \mathbb{R},$$

which in turn implies the linearity of $(\mathrm{prox}_\varphi^\Lambda)'(\bar{u};\cdot)$. On the other hand, suppose that the directional derivative $(\mathrm{prox}_\varphi^\Lambda)'(\bar{u};\cdot)$ is a linear mapping and let $\bar{h} \in \mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$ be arbitrary. Setting $\bar{d} := \bar{h} + \Lambda^{-1}\mathcal{H}_\varphi(\bar{p})\bar{h}$, it holds

$$\Lambda(\bar{h} - \bar{d}) + \mathcal{H}_\varphi(\bar{p})\bar{h} = 0 \in N_{\mathcal{C}_{\bar{u}}^\Lambda(\bar{u})}(\bar{h}) = [\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})]^\circ \cap \{\bar{h}\}^\perp$$

and, thus, it follows $(\mathrm{prox}_\varphi^\Lambda)'(\bar{u};\bar{d}) = \bar{h}$. Now, formula (5.3.46), the linearity of $(\mathrm{prox}_\varphi^\Lambda)'(\bar{u};\cdot)$, and $0 \in \mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$ immediately imply that the critical cone $\mathcal{C}_{\bar{u}}^\Lambda(\bar{p})$ is a subspace. □

Let us briefly reconsider our initial problem

$$\min_x \ f(x) + \varphi(x)$$

and let $\bar{x} \in \mathrm{dom}\ \varphi$ be a stationary point of the latter problem. Moreover, let us suppose that the function φ is C^2-fully decomposable at $\bar{x}$ and let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary parameter matrix. The corresponding first order optimality conditions can be represented as follows

$$\bar{x} = \operatorname{prox}^{\Lambda}_{\varphi}(\bar{u}), \quad \bar{u} := \bar{x} - \Lambda^{-1}\nabla f(\bar{x}).$$

Consequently, φ is "also" C^2-fully decomposable at $\bar{p} := \operatorname{prox}^{\Lambda}_{\varphi}(\bar{u}) = \bar{x}$ and, due to

$$\nabla \operatorname{env}^{\Lambda}_{\varphi}(\bar{u}) = \Lambda(\bar{u} - \bar{p}) = -\nabla f(\bar{x}) + \Lambda(\bar{x} - \operatorname{prox}^{\Lambda}_{\varphi}(\bar{u})) = -\nabla f(\bar{x}),$$

we can formulate the following corollary, which apparently does not need a proof.

Corollary 5.3.33. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be convex, proper, and lower semicontinuous. Suppose that $\bar{x} \in \operatorname{dom} \varphi$ is a stationary point of problem $(\mathcal{P})$ and let φ be C^2-fully decomposable at $\bar{x}$. Then, for every parameter matrix $\Lambda \in \mathbb{S}^n_{++}$, the proximity operator $\operatorname{prox}^{\Lambda}_{\varphi}$ is Fréchet differentiable at $\bar{u} := \bar{x} - \Lambda^{-1}\nabla f(\bar{x})$ if and only if the strict complementarity condition*

$$-\nabla f(\bar{x}) \in \operatorname{ri} \partial\varphi(\bar{x})$$

holds at $\bar{x}$.

Thus, in summary, if the convex, proper, and lower semicontinuous mapping φ is C^2-fully decomposable at every point $x \in \operatorname{dom} \varphi$, then our analysis shows that the strict complementarity condition completely characterizes the Fréchet differentiability of the corresponding proximity operator. Once more, this illustrates the abundance and the structural advantages of the concept of full decomposability. We conclude this paragraph with an exemplary application.

Example 5.3.34 (Semidefinite programming). We consider the matrix optimization problem

$$\min_{X \in \mathbb{S}^n} \; f(X) \quad \text{s.t.} \quad X \in \mathbb{S}^n_+, \tag{5.3.47}$$

where $f : \mathbb{S}^n \to \mathbb{R}$ is a twice continuously differentiable function. Since the cone of positive semidefinite, symmetric matrices is C^∞-cone reducible at every point $X \in \mathbb{S}^n_+$, (see, e.g., [27, Example 3.140]), it immediately follows that the indicator function $\varphi : \mathbb{S}^n \to (-\infty, +\infty]$, $\varphi(X) := \iota_{\mathbb{S}^n_+}(X)$ is C^∞-fully decomposable at $X \in \mathbb{S}^n_+$. Hence, our abstract second order framework can also be applied to general semidefinite programs of the form (5.3.47).

In the following, we briefly want to analyze the differentiability properties of the proximity operator $\operatorname{prox}^{I}_{\varphi}(\cdot) = \mathcal{P}_{\mathbb{S}^n_+}(\cdot)$. Therefore, let $\bar{U} \in \mathbb{S}^n$ be arbitrary and let us consider the spectral decomposition of $\bar{U}$,

$$\bar{U} = P\Sigma P^\top, \quad \Sigma = \operatorname{diag}(\sigma) \in \mathbb{S}^n,$$

where each σ_i, $i = 1, ..., n$, denotes a corresponding eigenvalue of $\bar{U}$ and $P \in \mathbb{R}^{n\times n}$ is an orthogonal matrix. Then, by [107], the proximity operator $\operatorname{prox}^{I}_{\varphi}(\bar{U})$ can be computed as follows

$$\bar{U}_+ := \mathrm{prox}_\varphi^I(\bar{U}) = \mathcal{P}_{\mathbb{S}^n_+}(\bar{U}) = \operatorname*{arg\,min}_{Y \in \mathbb{S}^n_+} \frac{1}{2}\|\bar{U} - Y\|_F^2 = P \operatorname{diag}(\max\{\sigma, 0\})P^\top.$$

Next, let us define the index sets

$$\alpha := \{i : \sigma_i > 0\}, \quad \beta := \{i : \sigma_i = 0\}, \quad \gamma := \{i : \sigma_i < 0\}.$$

Then, as in [27, 183, 226, 43] and by setting $\bar{\alpha} = \{1, ..., n\} \setminus \alpha$, the tangent cone $T_{\mathbb{S}^n_+}(\bar{U}_+)$ has the following explicit representation

$$T_{\mathbb{S}^n_+}(\bar{U}_+) = \{H \in \mathbb{S}^n : P_{[\cdot \bar{\alpha}]}^\top H P_{[\cdot \bar{\alpha}]} \succeq 0\}$$

and, due to $\varphi^\downarrow(\bar{U}_+; H) = \iota_{T_{\mathbb{S}^n_+}(\bar{U}_+)}(H)$, it can be shown that the critical cone reduces to

$$\begin{aligned} \mathcal{C}_{\bar{U}}(\bar{U}_+) &= \{H \in T_{\mathbb{S}^n_+}(\bar{U}_+) : \langle \bar{U} - \bar{U}_+, H \rangle = 0\} \\ &= \{H \in \mathbb{S}^n : P_{[\cdot\beta]}^\top H P_{[\cdot\beta]} \succeq 0,\ P_{[\cdot\beta]}^\top H P_{[\cdot\gamma]} = 0,\ P_{[\cdot\gamma]}^\top H P_{[\cdot\gamma]} = 0\}. \end{aligned}$$

Consequently, Lemma 5.3.32 implies that the projection $\mathcal{P}_{\mathbb{S}^n_+}(\cdot)$ is Fréchet differentiable at $\bar{U}$ if and only the index set β is empty, i.e., if and only if the matrix $\bar{U}$ is invertible. Of course, this result is already well-known; see, e.g., [183, Corollary 10] or [226] for more details.

The $\mathcal{VU}$-concept

The *$\mathcal{VU}$-concept* was introduced by Lemaréchal, Oustry, and Sagastizábal in [130] to analyze and express second order differentiability properties of real valued, convex and possibly nonsmooth functions. Basically, the idea is to decompose the space

$$\mathbb{R}^n = \mathcal{U} \oplus \mathcal{V}$$

into two perpendicular subspaces $\mathcal{U}$ and $\mathcal{V}$ and to study the behavior of φ along this subspaces. Typically, the gully-shaped space $\mathcal{U}$ is chosen such that the restriction of φ to the set $\mathcal{U}$ is differentiable in the classical sense. On the other hand, the narrow, V-shaped space $\mathcal{V}$ is parallel to the affine hull of the subdifferential of φ and captures the nonsmoothness of the function φ. Moreover, Lemaréchal et al. developed the *$\mathcal{U}$-Lagrangian*, $L_\mathcal{U} : \mathcal{U} \to \mathbb{R}$, of φ that, in contrast to the convex function φ, is solely defined on the $\mathcal{U}$-space and can be shown to be Fréchet differentiable at a certain point of interest. This enables the investigation of second order properties of $L_\mathcal{U}$ and leads to the concept of the $\mathcal{U}$-Hessian of φ. In [103, 155], these ideas were extended to general, real extended valued and *prox-regular* functions by introducing a regularized version of the $\mathcal{U}$-Lagrangian – the so-called *quadratic sub-Lagrangian*. Based on the $\mathcal{VU}$-concept, Mifflin and Sagastizábal [154] proposed an algorithm for convex, real valued, and unconstrained minimization. It uses a $\mathcal{VU}$-space decomposition, bundle techniques and generates a proximal point sequence that follows a smooth trajectory in the $\mathcal{V}$-space. For more details on $\mathcal{VU}$-related algorithms and applications we refer to [60, 140, 102] and [179, 112], respectively.

Here, we are motivated by the following facts:

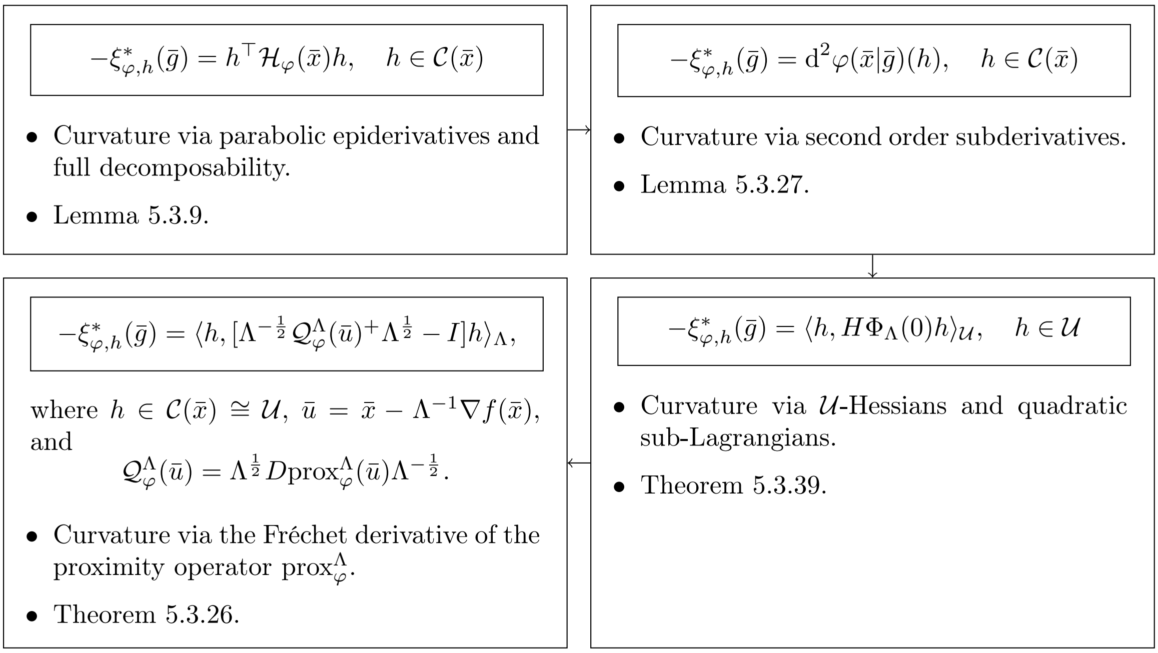
- In [155] Mifflin and Sagastizábal showed that, if the quadratic sub-Lagrangian has a generalized Hessian at 0 and the strict complementarity condition is satisfied, then the second order subderivative of φ exists (at a certain point) and its value coincides with the quadratic form induced by the $\mathcal{U}$-Hessian on the subspace $\mathcal{U}$.
- In [150, 152, 153] Mifflin and Sagastizábal analyzed the second order behavior of max-type functions of the form
$$f : \mathbb{R}^n \to \mathbb{R}, \quad f(x) := \max\{f_i(x) : i = 1, ..., m\}, \quad f_i \in C^2,$$
and of a class of functions with primal-dual gradient structure (pdg structure). As in [155], they established a connection between the $\mathcal{U}$-Hessian, the second order subderivative and the second order parabolic epiderivative of φ. Moreover, under a certain index set-based regularity condition, they derive and provide explicit formulae for the $\mathcal{U}$-Hessian and the epiderivatives.
- Finally, in [132, 130, 151], a connection between the $\mathcal{U}$-Hessian of φ and the Hessian of the corresponding Moreau envelope $\mathrm{env}_\varphi^\Lambda$ is presented. We will utilize this connection to complete our theoretical "detour" and to derive an intrinsic characterization of the curvature $\xi^*_{\varphi,h}(-\nabla f(\bar{x}))$, $h \in \mathcal{U}$, in terms of the Fréchet derivative of the proximity operator $\mathrm{prox}_\varphi^\Lambda$ – just as presented in Theorem 5.3.26.

Let us briefly recall our current situation to clarify the motivational aspects of the latter observations. Let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function and let $\bar{x} \in \mathrm{dom}\, \varphi$ be a stationary point of problem $(\mathcal{P})$. Furthermore, let $\Lambda \in \mathbb{S}^n_{++}$ be given and suppose that φ is C^2-fully decomposable at $\bar{x}$. Then, by setting $\bar{g} := -\nabla f(\bar{x})$ and by using the notation of the proof of Lemma 5.3.32 and Lemma 5.3.27, it follows

$$\mathrm{d}^2\varphi(\bar{x}|\bar{g})(h) = -\xi^*_{\varphi,h}(\bar{g}) = h^\top \mathcal{H}_\varphi(\bar{x})h, \quad \forall\ h \in \mathcal{C}(\bar{x}). \tag{5.3.48}$$

Thus, the second order subderivative has obviously a Hessian-like structure. Moreover, inspired by the mentioned results for max-type and pdg-structured functions, this also indicates that a similar connection between second order parabolic epiderivatives, second order subderivatives, and $\mathcal{U}$-Hessians does also exist for the class of fully decomposable functions. However, at this point and in contrast to [150, 155, 153, 152], we cannot directly infer that φ has a $\mathcal{U}$-Hessian at $\bar{x}$.

Our task is now as follows. First, we will introduce the $\mathcal{U}$-, and $\mathcal{V}$-space, the quadratic sub-Lagrangian, and several helpful and necessary $\mathcal{VU}$-tools. Then, we extend the computational and theoretical results of Mifflin and Sagastizábal to the class of fully decomposable functions. In particular, by mimicking the proof of [155, Theorem 3.2], we will show that the second order subderivative $\mathrm{d}^2\varphi(\bar{x}|\bar{g})$ coincides with the second order subderivative of the quadratic sub-Lagrangian. Invoking equation (5.3.48) and Theorem 2.4.4, and using the convexity of the second order difference quotients, this finally implies that $\mathcal{H}_\varphi(\bar{x})$ is actually the $\mathcal{U}$-Hessian of φ at $\bar{x}$, as expected. Applying the results of Lemaréchal et al. [132, 130], we are then able to conclude the proof of Theorem 5.3.26. An overview of the various representations of the



$$-\xi^*_{\varphi,h}(\bar g) = h^\top \mathcal{H}_\varphi(\bar x)h, \quad h \in \mathcal{C}(\bar x)$$

- Curvature via parabolic epiderivatives and full decomposability.
- Lemma 5.3.9.

$$-\xi^*_{\varphi,h}(\bar g) = \mathrm{d}^2\varphi(\bar x|\bar g)(h), \quad h \in \mathcal{C}(\bar x)$$

- Curvature via second order subderivatives.
- Lemma 5.3.27.

$$-\xi^*_{\varphi,h}(\bar g) = \langle h, H\Phi_\Lambda(0)h\rangle_{\mathcal{U}}, \quad h \in \mathcal{U}$$

- Curvature via $\mathcal{U}$-Hessians and quadratic sub-Lagrangians.
- Theorem 5.3.39.

$$-\xi^*_{\varphi,h}(\bar g) = \langle h, [\Lambda^{-\frac{1}{2}}\mathcal{Q}^\Lambda_\varphi(\bar u)^+\Lambda^{\frac{1}{2}} - I]h\rangle_\Lambda,$$

where $h \in \mathcal{C}(\bar x) \cong \mathcal{U}$, $\bar u = \bar x - \Lambda^{-1}\nabla f(\bar x)$, and

$$\mathcal{Q}^\Lambda_\varphi(\bar u) = \Lambda^{\frac{1}{2}} D\mathrm{prox}^\Lambda_\varphi(\bar u)\Lambda^{-\frac{1}{2}}.$$

- Curvature via the Fréchet derivative of the proximity operator $\mathrm{prox}^\Lambda_\varphi$.
- Theorem 5.3.26.

Figure 5.3.: Different expressions of the curvature term $-\xi^*_{\varphi,h}(\bar g)$ and illustration of concept of the proof of Theorem 5.3.26.

curvature term $-\xi^*_{\varphi,h}(\bar g)$ and of the different steps of the proof of Theorem 5.3.26 is given in Figure 5.3.

As in the last subsection, we start with a slightly more general setting that includes the stationary case as a special case (see, e.g., Corollary 5.3.33 and the preceding discussion). Therefore, let $\bar u \in \mathbb{R}^n$, $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary and let us suppose that φ is C^2-fully decomposable at the point $\bar p := \mathrm{prox}^\Lambda_\varphi(\bar u)$. Moreover, let us set $\bar g := \nabla\mathrm{env}^\Lambda_\varphi(\bar u)$. Here, we consider the following subspaces

$$\mathcal{U} := \mathrm{lin}\ N_{\partial\varphi(\bar p)}(\bar g), \quad \mathcal{U}^\perp = \mathrm{aff}\ \partial\varphi(\bar p) - \bar g, \quad \mathcal{V} := \mathcal{U}^{\perp,\Lambda}, \tag{5.3.49}$$

where the latter operation is defined via

$$v \in \mathcal{V} = \mathcal{U}^{\perp,\Lambda} \quad :\Longleftrightarrow \quad \langle u, v\rangle_\Lambda = 0, \quad \forall\ u \in \mathcal{U}.$$

Since the bilinear form $\langle\cdot,\cdot\rangle_\Lambda : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ represents a scalar product on $\mathbb{R}^n \times \mathbb{R}^n$ and $\mathcal{U}$ is a closed subspace, it immediately follows $\mathcal{U} = \mathcal{V}^{\perp,\Lambda} = [\mathcal{U}^{\perp,\Lambda}]^{\perp,\Lambda}$. Let us note, that the space $\mathcal{U}$ does not depend on the specific choice of the subgradient $\bar g$, i.e., we have $\mathcal{U} = \mathrm{lin}\ N_{\partial\varphi(\bar p)}(g)$ for every $g \in \partial\varphi(\bar p)$. Furthermore, if the strict complementarity condition $\bar g \in \mathrm{ri}\ \partial\varphi(\bar p)$ is satisfied, then the lineality space operation "lin" is superfluous.

Now, let us define $n_{\mathcal{U}} := \dim\ \mathcal{U}$, $n_{\mathcal{V}} := \dim\ \mathcal{V}$ and let $\bar U \in \mathbb{R}^{n\times n_{\mathcal{U}}}$, and $\bar V \in \mathbb{R}^{n\times n_{\mathcal{V}}}$ be two basis matrices of the subspaces $\mathcal{U}$ and $\mathcal{V}$, respectively. Then, the projection of a vector

$x \in \mathbb{R}^n$ onto the sets $\mathcal{U}$ and $\mathcal{V}$ can be calculated via

$$\mathcal{P}_{\mathcal{U}}(x) = \bar{U}[\bar{U}^\top \Lambda \bar{U}]^{-1}\bar{U}^\top \Lambda x =: \bar{U}x_{\mathcal{U}}, \quad \mathcal{P}_{\mathcal{V}}(x) = \bar{V}[\bar{V}^\top \Lambda \bar{V}]^{-1}\bar{V}^\top \Lambda x =: \bar{V}x_{\mathcal{V}}$$

and we have the following space decomposition formula

$$x = \bar{U}[\bar{U}^\top \Lambda \bar{U}]^{-1}\bar{U}^\top \Lambda x + \bar{V}[\bar{V}^\top \Lambda \bar{V}]^{-1}\bar{V}^\top \Lambda x = \bar{U}x_{\mathcal{U}} + \bar{V}x_{\mathcal{V}} =: x_{\mathcal{U}} \oplus_\Lambda x_{\mathcal{V}} \in \mathbb{R}^{n_{\mathcal{U}}} \times \mathbb{R}^{n_{\mathcal{V}}}.$$

Due to $\iota_{\mathcal{U}}^* = \iota_{\mathcal{U}^\perp}$, the latter formula is just a simple application of Moreau's decomposition principle, Theorem 3.2.5,

$$x = \mathcal{P}_{\mathcal{U}}^\Lambda(x) + \Lambda^{-1}\mathcal{P}_{\mathcal{U}^\perp}^{\Lambda^{-1}}(\Lambda x) = \mathcal{P}_{\mathcal{U}}^\Lambda(x) + \mathcal{P}_{\mathcal{V}}^\Lambda(x).$$

This terminology can also be naturally extended to scalar products and norms. It holds

- $\langle x, y\rangle_\Lambda = \langle \bar{U}x_{\mathcal{U}} + \bar{V}x_{\mathcal{V}}, \Lambda\bar{U}y_{\mathcal{U}} + \Lambda\bar{V}y_{\mathcal{V}}\rangle = \langle \bar{U}x_{\mathcal{U}}, \bar{U}y_{\mathcal{U}}\rangle_\Lambda + \langle \bar{V}x_{\mathcal{V}}, \bar{V}y_{\mathcal{V}}\rangle_\Lambda$
  $=: \langle x_{\mathcal{U}}, y_{\mathcal{U}}\rangle_{\mathcal{U},\Lambda} + \langle x_{\mathcal{V}}, y_{\mathcal{V}}\rangle_{\mathcal{V},\Lambda}$,
- $\|x\|_\Lambda^2 = \begin{pmatrix} x_{\mathcal{U}}^\top & x_{\mathcal{V}}^\top \end{pmatrix} \begin{pmatrix} \bar{U}^\top\Lambda\bar{U} & \bar{U}^\top\Lambda\bar{V} \\ \bar{V}^\top\Lambda\bar{U} & \bar{V}^\top\Lambda\bar{V} \end{pmatrix} \begin{pmatrix} x_{\mathcal{U}} \\ x_{\mathcal{V}} \end{pmatrix} =: \|x_{\mathcal{U}}\|_{\mathcal{U},\Lambda}^2 + \|x_{\mathcal{V}}\|_{\mathcal{V},\Lambda}^2$,

where we used the Λ-orthogonality of the basis matrices $\bar{U}$ and $\bar{V}$. Let us note that, besides the original definition (5.3.49), the subspaces $\mathcal{U}$ and $\mathcal{V}$ are also often identified with the subspaces $\mathbb{R}^{n_{\mathcal{U}}}$ and $\mathbb{R}^{n_{\mathcal{V}}}$, respectively. In particular, we will work with the following notions:

- $u \in \mathcal{U} \subset \mathbb{R}^n \iff u = u_{\mathcal{U}} \oplus_\Lambda 0 \iff \bar{V}^\top \Lambda u = 0.$
- $u \in \mathcal{U} \cong \mathbb{R}^{n_{\mathcal{U}}} \iff \exists\ \tilde{u} \in \mathbb{R}^n : \bar{U}u = \tilde{u}.$

Additionally, if necessary, we will also use the terms

$$\langle\cdot,\cdot\rangle_{\mathcal{U}} : \mathbb{R}^{n_{\mathcal{U}}} \times \mathbb{R}^{n_{\mathcal{U}}} \to \mathbb{R}, \quad \|x\|_{\mathcal{U}}^2 = \langle x, x\rangle_{\mathcal{U}}, \quad \langle\cdot,\cdot\rangle_{\mathcal{V}} : \mathbb{R}^{n_{\mathcal{V}}} \times \mathbb{R}^{n_{\mathcal{V}}} \to \mathbb{R}, \quad \|y\|_{\mathcal{V}}^2 = \langle y, y\rangle_{\mathcal{V}}$$

to denote the scalar product and the corresponding, induced norm on $\mathbb{R}^{n_{\mathcal{U}}}$ and $\mathbb{R}^{n_{\mathcal{V}}}$.

Next, for $\hat{g} := \Lambda^{-1}\bar{g}$, we define the so-called *quadratic sub-Lagrangian* of φ at $\bar{p}$

$$\Phi_\Lambda : \mathcal{U} \to [-\infty, +\infty], \quad \Phi_\Lambda(u) := \inf_{v\in\mathcal{V}}\ \varphi(\bar{p} + u \oplus_\Lambda v) - \langle \hat{g}_{\mathcal{V}}, v\rangle_{\mathcal{V},\Lambda} + \frac{1}{2}\|v\|_{\mathcal{V},\Lambda}^2$$

and the associated multi-valued mapping

$$W_\Lambda : \mathcal{U} \rightrightarrows \mathcal{V}, \quad W_\Lambda(u) := \operatorname*{arg\,min}_{v\in\mathcal{V}}\ \varphi(\bar{p} + u \oplus_\Lambda v) - \langle \hat{g}_{\mathcal{V}}, v\rangle_{\mathcal{V},\Lambda} + \frac{1}{2}\|v\|_{\mathcal{V},\Lambda}^2.$$

These two definitions are essentially based on the work of Hare and Poliquin [103], but, in contrast to [103, 130], do also take account of the parameter matrix Λ and of the induced $\mathcal{V}$-geometry. In comparison, the original $\mathcal{U}$-Lagrangian, introduced by Lemaréchal, Oustry, and Sagastizábal, [130], is only well-defined for convex, real valued functions $\varphi : \mathbb{R}^n \to \mathbb{R}$

and has the following form

$$L_{\mathcal{U}}(u) := \inf_{v\in\mathcal{V}} \varphi(\bar{p} + u \oplus v) - \langle \bar{g}_{\mathcal{V}}, v\rangle_{\mathcal{V}}, \quad (\Lambda = I).$$

Hence, by adding the quadratic term $\frac{1}{2}\|v\|^2_{\mathcal{V},\Lambda}$, Hare and Poliquin successfully extended the concept of the $\mathcal{U}$-Lagrangian to real extended valued and possibly nonconvex functions.

In the following, we present several important properties of the quadratic sub-Lagrangian. More precisely, we will show that the quadratic sub-Lagrangian is convex, Lipschitz continuous in a neighborhood of 0, and Fréchet differentiable at 0. Let us note that the convexity of φ significantly simplifies the discussion of the quadratic sub-Lagrangian and that most of our proofs are easy extensions of the corresponding proofs for the original $\mathcal{U}$-Lagrangian.

Lemma 5.3.35. *Let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. The functions Φ_Λ and W_Λ have the following properties:*

(i) *The function Φ_Λ is convex, proper, and lower semicontinuous. In particular, it holds $\Phi_\Lambda(0) = \varphi(\bar{p})$.*

(ii) *It holds $W_\Lambda(u) \neq \emptyset$ for all $u \in \operatorname{dom} \Phi_\Lambda$ and $W_\Lambda(0) = \{0\}$.*

Proof. In [103, Theorem 5 and Proposition 6], a proof of these two statements is provided for the general nonconvex and prox-regular setting. Here, similar to [130, Theorem 3.2], we will explicitly exploit the convexity of φ. First, let us note that the quadratic sub-Lagrangian can be written as a marginal function

$$\Phi_\Lambda(u) = \inf_{v\in\mathcal{V}} \theta(u,v), \quad \theta(u,v) := \varphi(\bar{p} + u \oplus_\Lambda v) - \langle \hat{g}_{\mathcal{V}}, v\rangle_{\mathcal{V},\Lambda} + \frac{1}{2}\|v\|^2_{\mathcal{V},\Lambda}.$$

Since the function $\theta : \mathbb{R}^{n_{\mathcal{U}}} \times \mathbb{R}^{n_{\mathcal{V}}} \to (-\infty, +\infty]$ is obviously convex, we can utilize [11, Proposition 8.26]. This establishes convexity of the quadratic sub-Lagrangian Φ_Λ. Furthermore, using $\bar{g} \in \partial\varphi(\bar{p})$, it follows

$$\theta(u,v) - \theta(0,0) \geq \langle \bar{g}, u \oplus_\Lambda v\rangle - \langle \hat{g}_{\mathcal{V}}, v\rangle_{\mathcal{V},\Lambda} + \frac{1}{2}\|v\|^2_{\mathcal{V},\Lambda} = \langle \hat{g}_{\mathcal{U}}, u\rangle_{\mathcal{U},\Lambda} + \frac{1}{2}\|v\|^2_{\mathcal{V},\Lambda} \tag{5.3.50}$$

and consequently, we have

$$\theta(u,0) \geq \Phi_\Lambda(u) = \inf_{v\in\mathcal{V}} \theta(u,v) \geq \varphi(\bar{p}) + \langle \hat{g}_{\mathcal{U}}, u\rangle_{\mathcal{U},\Lambda}, \quad \forall\, u \in \mathcal{U}.$$

Clearly, this implies $\Phi_\Lambda(0) = \varphi(\bar{p})$ and shows that the quadratic sub-Lagrangian is a proper function. The lower semicontinuity of Φ_Λ follows from [103, Theorem 7] and will not be discussed here. Let us continue with the verification of the second part. Apparently, inequality (5.3.50) also implies that for every fixed $u \in \mathcal{U}$, the function $\theta(u,\cdot)$ is coercive on the subspace $\mathcal{V}$. Moreover, for all $u \in \operatorname{dom} \Phi_\Lambda$, we can further deduce that $\theta(u,\cdot) : \mathcal{V} \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous mapping. Thus, by Lemma 4.2.3, all level sets of $\theta(u,\cdot)$ are bounded and $W_\Lambda(u)$ has to be nonempty. The formula $W_\Lambda(0) = \{0\}$ immediately follows from (5.3.50). □

Lemma 5.3.36. *Let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. Then, the quadratic sub-Lagrangian Φ_Λ is locally Lipschitz continuous and Fréchet differentiable at* 0 *and its gradient satisfies $\nabla\Phi_\Lambda(0) = \bar{U}^\top \bar{g}$.*

Proof. Since Φ_Λ is a convex marginal function

$$\Phi_\Lambda(u) = \inf_{v\in\mathcal{V}} \theta(u,v), \quad \theta(u,v) := \varphi(\bar{p} + \bar{U}u + \bar{V}v) - \langle \bar{V}\hat{g}, \bar{V}v\rangle_\Lambda + \frac{1}{2}\|v\|^2_{\bar{V}^\top\Lambda\bar{V}},$$

we can apply [11, Proposition 16.46] to characterize the subdifferential $\partial\Phi_\Lambda(0)$. In particular, it holds

$$\begin{aligned} w \in \partial\Phi_\Lambda(0) \quad &\iff \quad \begin{pmatrix} w \\ 0 \end{pmatrix} \in \partial\theta(0, W_\Lambda(0)) = \begin{pmatrix} \bar{U}^\top \\ \bar{V}^\top \end{pmatrix} \partial\varphi(\bar{p}) - \begin{pmatrix} 0 \\ \bar{V}^\top\Lambda\bar{V}\hat{g} \end{pmatrix} \\ &\iff \quad \bar{U}(\bar{U}^\top\Lambda\bar{U})^{-1}w \in \Lambda^{-1}\partial\varphi(\bar{p}) - \bar{V}\hat{g} \\ &\iff \quad \Lambda\bar{U}(\bar{U}^\top\Lambda\bar{U})^{-1}w \in \partial\varphi(\bar{p}) - \Lambda\bar{V}\hat{g} \subset \mathcal{U}^\perp + \bar{g} - \Lambda\bar{V}\hat{g}, \end{aligned}$$

where we used $W_\Lambda(0) = \{0\}$, Lemma 2.5.15, and (5.3.49). Multiplying the latter inclusion with $\bar{U}^\top$ yields

$$w \in \bar{U}^\top\mathcal{U}^\perp + \bar{U}^\top\bar{g} - \bar{U}^\top\Lambda\bar{V}\hat{g} = \{\bar{U}^\top\bar{g}\}.$$

Now, by utilizing Lemma 2.5.12 and Theorem 2.2.1, it follows that Φ_Λ is locally Lipschitz continuous near 0. Moreover, in this case, since the subdifferential $\partial\Phi_\Lambda(0) = \{\bar{U}^\top\bar{g}\}$ is a singleton, [11, Proposition 17.26] is applicable and we obtain $\nabla\Phi_\Lambda(0) = \bar{U}^\top\bar{g}$. Hence, the quadratic sub-Lagrangian Φ_Λ is Fréchet differentiable at 0; see also [11, Proposition 17.36]. □

Consequently, the quadratic sub-Lagrangian Φ_Λ has a higher regularity than the convex base-function φ which allows to analyze and characterize second order properties of Φ_Λ via classical tools. Now, due to

$$\nabla\Phi_\Lambda(0)^\top h = \langle \bar{U}(\bar{U}^\top\Lambda\bar{U})^{-1}\bar{U}^\top\Lambda\hat{g}, \bar{U}h\rangle_\Lambda = \langle \bar{U}\hat{g}_\mathcal{U}, \bar{U}h\rangle_\Lambda = \langle \hat{g}_\mathcal{U}, h\rangle_{\mathcal{U},\Lambda}, \quad h \in \mathcal{U},$$

the vector $\hat{g}_\mathcal{U}$ is called *$\mathcal{U}$-gradient* of φ at $\bar{p}$. Moreover, we say that Φ_Λ has a *generalized Hessian* at 0 if and only if there exists a symmetric, positive semidefinite operator $H\Phi_\Lambda(0) \in \mathbb{R}^{n_\mathcal{U}\times n_\mathcal{U}}$ such that

$$\Phi_\Lambda(h) - \Phi_\Lambda(0) - \nabla\Phi_\Lambda(0)^\top h - \frac{1}{2}\langle h, H\Phi_\Lambda(0)h\rangle_\mathcal{U} = o(\|h\|^2_\mathcal{U}), \quad (h \to 0). \tag{5.3.51}$$

If the generalized Hessian $H\Phi_\Lambda(0)$ exists, then we call it a *$\mathcal{U}$-Hessian* for φ at $\bar{x}$. Let us emphasize that the generalized Hessian $H\Phi_\Lambda(0)$ must not be confused with the classical Hessian $\nabla^2\Phi_\Lambda(0)$ of the function Φ_Λ. In particular, the expansion (5.3.51) does not guarantee twice Fréchet differentiability of Φ_Λ at 0 since the quadratic sub-Lagrangian typically need not be differentiable in a neighborhood of 0. The next lemma mathematically clarifies the term "$\mathcal{U}$-Hessian" and combines [130, Corollary 3.5] and [154, Lemma 3.1].

Lemma 5.3.37. *Let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary and suppose that the condition $\bar{g} \in$ ri $\partial\varphi(\bar{p})$ is satisfied. Let us consider a $\mathcal{V}$-space minimizer function $v(h) \in W_\Lambda(h)$, $h \in$ dom Φ_Λ. Then, the following statements hold:*

(i) *It holds* $v(h) = o(\|h\|_{\mathcal{U}})$, $h \to 0$.

(ii) *Additionally, if the quadratic sub-Lagrangian has a generalized Hessian at* 0, *then it holds* $v(h) = O(\|h\|_{\mathcal{U}}^2)$, $h \to 0$, *and we obtain*

$$\varphi(\bar{p} + h \oplus_\Lambda v(h)) = \varphi(\bar{p}) + \langle \hat{g}, h \oplus_\Lambda v(h)\rangle_\Lambda + \frac{1}{2}\langle h, H\Phi_\Lambda(0)h\rangle_{\mathcal{U}} + o(\|h\|_{\mathcal{U}}^2),$$

for all $h \in \mathcal{U}$ *sufficiently small.*

Proof. We only prove the first part of Lemma 5.3.37. A proof of the second part can be found in [154, Lemma 3.1]. The strict complementarity condition $\bar{g} \in \mathrm{ri}\ \partial\varphi(\bar{p})$ and the definition of the subspace $\mathcal{V}$ imply that there exists $\varepsilon > 0$ such that

$$\hat{g} + \frac{\varepsilon}{\|v\|_{\mathcal{V},\Lambda}} \cdot 0 \oplus_\Lambda v \in \Lambda^{-1}\partial\varphi(\bar{p}), \quad \forall\ v \in \mathcal{V} \setminus \{0\}.$$

Hence, as in (5.3.50), we have

$$\theta(u, v) - \theta(0, 0) \geq \langle \hat{g}_{\mathcal{U}}, u\rangle_{\mathcal{U},\Lambda} + \frac{1}{2}\|v\|_{\mathcal{V},\Lambda}^2 + \varepsilon\|v\|_{\mathcal{V},\Lambda},$$

and for $v(h) \in W_\Lambda(h)$, it follows

$$\begin{aligned}\Phi_\Lambda(h) - \Phi_\Lambda(0) - \nabla\Phi_\Lambda(0)^\top h &= \theta(h, v(h)) - \theta(0, 0) - \langle \hat{g}_{\mathcal{U}}, h\rangle_{\mathcal{U},\Lambda} \\ &\geq \frac{1}{2}(\|v(h)\|_{\mathcal{V},\Lambda} + 2\varepsilon)\|v(h)\|_{\mathcal{V},\Lambda} \geq \varepsilon\|v(h)\|_{\mathcal{V},\Lambda}.\end{aligned}$$

Now, the Fréchet differentiability of Φ_Λ immediately establishes $v(h) = o(\|h\|_{\mathcal{U}})$, as $h \to 0$. Moreover, let us note that, due to Theorem 2.2.1, the continuity of Φ_Λ is equivalent to the condition $0 \in \mathrm{int\ dom}\ \Phi_\Lambda$. Thus, by Lemma 5.3.35 (ii), we have $W_\Lambda(h) \neq \emptyset$ for all $h \in \mathcal{U}$ sufficiently small and the $\mathcal{V}$-space minimizer function $v(h)$ is well-defined in a neighborhood of 0. □

Connecting second order subderivatives and $\mathcal{U}$-Hessians

In the following, we show that the second order subderivatives of the quadratic sub-Lagrangian Φ_Λ and of the convex function φ coincide. This result is strongly motivated by [154, Theorem 3.2] and covers the case when existence of the $\mathcal{U}$-Hessian or the generalized Hessian of Φ_Λ cannot be guaranteed in advance.

Lemma 5.3.38. *Let* $\Lambda \in \mathbb{S}_{++}^n$ *be an arbitrary parameter matrix and let the strict complementarity condition* $\bar{g} \in \mathrm{ri}\ \partial\varphi(\bar{p})$ *be satisfied. Furthermore, let us suppose that the function* φ *is twice epi-subdifferentiable at* $\bar{p}$, *relative to* $\bar{g}$. *Then, the quadratic sub-Lagrangian* Φ_Λ *is twice epi-subdifferentiable at* 0, *relative to* $g_0 := \nabla\Phi_\Lambda(0)$ *and the corresponding second order subderivative of* Φ_Λ *at* 0 *is given by*

$$\mathrm{d}^2\Phi_\Lambda(0|g_0)(h_{\mathcal{U}}) = \mathrm{d}^2\varphi(\bar{p}|\bar{g})(h), \quad \forall\ h = h_{\mathcal{U}} \oplus_\Lambda 0 \in \mathcal{U}. \tag{5.3.52}$$

Proof. Let $u \in \operatorname{dom} \Phi_\Lambda$ be arbitrary and let $v(u) \in W_\Lambda(u)$ be a corresponding $\mathcal{V}$-space minimizer function. Then, it holds

$$\Phi_\Lambda(u) = \theta(u, v(u)) \le \varphi(\bar{p} + u \oplus_\Lambda v) - \langle \hat{g}_\mathcal{V}, v\rangle_{\mathcal{V},\Lambda} + \tfrac{1}{2}\|v\|^2_{\mathcal{V},\Lambda}, \quad \forall\, v \in \mathcal{V},$$

and it readily follows

$$\Phi_\Lambda(u) - \Phi_\Lambda(0) - \nabla\Phi_\Lambda(0)^\top u \le \varphi(\bar{p} + u \oplus_\Lambda v) - \varphi(\bar{p}) - \langle \hat{g}, u \oplus_\Lambda v\rangle_\Lambda + \tfrac{1}{2}\|v\|^2_{\mathcal{V},\Lambda}. \tag{5.3.53}$$

Next, let $(t_k)_k$, $t_k \downarrow 0$, be an arbitrary sequence and fix any $h = h_\mathcal{U} \oplus_\Lambda 0 \in \mathcal{U}$. Since φ is twice epi-subdifferentiable at $\bar{p}$ relative to $\bar{g}$, there exists a sequence $(h^k)_k \subset \mathbb{R}^n$, $h^k \to h$, such that

$$\limsup_{k\to\infty}\ \Delta^2_{t_k}\,\varphi(\bar{p}|\bar{g})(h^k) \le \mathrm{d}^2\varphi(\bar{p}|\bar{g})(h).$$

Moreover, using the $\mathcal{VU}$-structure, we obtain

$$h^k = h^k_\mathcal{U} \oplus_\Lambda h^k_\mathcal{V}, \quad \bar{p} + t_k h^k = \bar{p} + (t_k h^k_\mathcal{U}) \oplus_\Lambda (t_k h^k_\mathcal{V}),$$

and $h^k_\mathcal{U} \to h_\mathcal{U}$, $h^k_\mathcal{V} \to 0$. Now, from (5.3.53) and for all k sufficiently large, it follows

$$\frac{\Phi_\Lambda(t_k h^k_\mathcal{U}) - \Phi_\Lambda(0) - t_k\langle \nabla\Phi_\Lambda(0), h^k_\mathcal{U}\rangle_\mathcal{U}}{\frac{1}{2}t_k^2} \le \Delta^2_{t_k}\,\varphi(\bar{p}|\bar{g})(h^k) + \|h^k_\mathcal{V}\|^2_{\mathcal{V},\Lambda}$$

and

$$\limsup_{k\to\infty}\ \Delta^2_{t_k}\,\Phi_\Lambda(0|g_0)(h^k_\mathcal{U}) \le \mathrm{d}^2\varphi(\bar{p}|\bar{g})(h) + \limsup_{k\to\infty}\ \|h^k_\mathcal{V}\|^2_{\mathcal{V},\Lambda} = \mathrm{d}^2\varphi(\bar{p}|\bar{g})(h).$$

On the other hand, let $(h^k_\mathcal{U})_k \subset \mathbb{R}^{n_\mathcal{U}}$ be an arbitrary sequence that converges to $h_\mathcal{U}$ and let us define

$$h^k := h^k_\mathcal{U} \oplus_\Lambda \frac{1}{t_k} v(t_k h^k_\mathcal{U}).$$

Now, by Lemma 5.3.37 (i), we have $v(t_k h^k_\mathcal{U}) = o(\|t_k h^k_\mathcal{U}\|_\mathcal{U})$, $k \to \infty$. Consequently, it follows $h^k \to h_\mathcal{U} \oplus_\Lambda 0 =: h$ and we obtain

$$\frac{\Phi_\Lambda(t_k h^k_\mathcal{U}) - \Phi_\Lambda(0) - t_k\langle \nabla\Phi_\Lambda(0), h^k_\mathcal{U}\rangle_\mathcal{U}}{\frac{1}{2}t_k^2} = \frac{\varphi(\bar{p} + t_k h^k) - \varphi(\bar{p}) - t_k\langle \bar{g}, h^k\rangle}{\frac{1}{2}t_k^2} + \frac{\|v(t_k h^k_\mathcal{U})\|^2_{\mathcal{V},\Lambda}}{t_k^2}.$$

Taking the limes inferior $k \to \infty$ over both sides of the latter equality and using the second order epi-subdifferentiability of φ, this yields

$$\liminf_{k\to\infty}\ \Delta^2_{t_k}\,\Phi_\Lambda(0|g_0)(h^k_\mathcal{U}) \ge \mathrm{d}^2\varphi(\bar{p}|\bar{g})(h) + \liminf_{k\to\infty}\ \frac{\|v(t_k h^k_\mathcal{U})\|^2_{\mathcal{V},\Lambda}}{t_k^2} = \mathrm{d}^2\varphi(\bar{p}|\bar{g})(h).$$

In summary, we have shown, that for any sequence $(t_k)_k$, $t_k \downarrow 0$, it holds

$$\begin{cases} \liminf\limits_{k\to\infty}\ \Delta^2_{t_k}\,\Phi_\Lambda(0|g_0)(h^k_\mathcal{U}) \ \ge \mathrm{d}^2\varphi(\bar{p}|\bar{g})(h) & \text{for every sequence } h^k_\mathcal{U} \to h_\mathcal{U}, \\ \limsup\limits_{k\to\infty}\ \Delta^2_{t_k}\,\Phi_\Lambda(0|g_0)(h^k_\mathcal{U}) \le \mathrm{d}^2\varphi(\bar{p}|\bar{g})(h) & \text{for some sequence } h^k_\mathcal{U} \to h_\mathcal{U}. \end{cases}$$

Thus, due to Lemma 2.4.3, the difference quotients $\Delta_t^2\,\Phi_\Lambda(0|g_0)$ epi-converge to the second order subderivative $\mathrm{d}^2\varphi(\bar{p}|\bar{g})$ on $\mathcal{U}$ and it holds

$$\mathrm{d}^2\Phi_\Lambda(0|g_0)(h_{\mathcal{U}}) = \mathrm{d}^2\varphi(\bar{p}|\bar{g})(h),$$

for all $h = h_{\mathcal{U}} \oplus_\Lambda 0 \in \mathcal{U}$. □

Next, we will merge our different results and observations. Therefore, let $\bar{x} \in \mathrm{dom}\ \varphi$ be a stationary point of problem $(\mathcal{P})$ and suppose that φ is C^2-fully decomposable at $\bar{x}$. Moreover, as usual, let us set $\bar{u} := \bar{x} - \Lambda^{-1}\nabla f(\bar{x})$ and let the strict complementarity condition be satisfied

$$-\nabla f(\bar{x}) \in \mathrm{ri}\ \partial\varphi(\bar{x}).$$

Then, we have the following identifications

$$\bar{x} = \bar{p} = \mathrm{prox}_\varphi^\Lambda(\bar{u}), \quad -\nabla f(\bar{x}) = \bar{g} = \nabla\mathrm{env}_\varphi^\Lambda(\bar{u}), \quad \mathcal{C}(\bar{x}) = N_{\partial\varphi(\bar{x})}(-\nabla f(\bar{x})) = \mathcal{U},$$

where we used the stationarity of $\bar{x}$, Definition 5.1.5 and Lemma 5.1.10. Furthermore, by Lemma 5.3.27, the stationarity of $\bar{x}$ and the full decomposability of φ imply that φ is twice epi-subdifferentiable at $\bar{x}$ relative to $\bar{g}$ and it holds

$$\mathrm{d}^2\varphi(\bar{x}|\bar{g})(h) = -\xi^*_{\varphi,h}(\bar{g}) = h^\top\mathcal{H}_\varphi(\bar{x})h, \quad \forall\ h \in \mathcal{C}(\bar{x}).$$

In particular, if (φ_d, F) is a corresponding decomposition pair of φ, then the symmetric and positive semidefinite matrix $\mathcal{H}_\varphi(\bar{x}) \in \mathbb{R}^{n\times n}$ is given by

$$\mathcal{H}_\varphi(\bar{x}) := \sum_{i=1}^m \bar{\lambda}_i\nabla^2 F_i(\bar{x}), \tag{5.3.54}$$

where $\bar{\lambda} \in \mathcal{M}(\bar{x})$ is the associated, unique Lagrange multiplier of the decomposed problem (5.3.33). Clearly, at this point, Lemma 5.3.38 is applicable and we can infer that the quadratic sub-Lagrangian is twice epi-subdifferentiable at 0, relative to g_0. We obtain

$$\mathrm{d}^2\Phi_\Lambda(0|g_0)(h_{\mathcal{U}}) = \mathrm{d}^2\varphi(\bar{x}|\bar{g})(h) = h^\top\mathcal{H}_\varphi(\bar{x})h, \quad \forall\ h\,(= h_{\mathcal{U}} \oplus_\Lambda 0) \in \mathcal{C}(\bar{x}).$$

Now, rephrasing the second order epi-subdifferentiability of the function Φ_Λ, the latter equation means that for every sequence $(t_k)_k$, $t_k \downarrow 0$, the family of *convex* difference quotients $\Delta_{t_k}^2\,\Phi_\Lambda(0|g_0) : \mathcal{U} \to (-\infty, +\infty]$,

$$\Delta_{t_k}^2\,\Phi_\Lambda(0|g_0)(h) = \frac{\Phi_\Lambda(t_k h) - \Phi_\Lambda(0) - t_k\nabla\Phi_\Lambda(0)^\top h}{\frac{1}{2}t_k^2}, \quad h \in \mathcal{U} \cong \mathbb{R}^{n_{\mathcal{U}}}$$

epi-converges to the convex and real valued function $\Xi_\varphi : \mathcal{U} \to \mathbb{R}$,

$$\Xi_\varphi(h) := \langle h, \bar{U}^\top\mathcal{H}_\varphi(\bar{x})\bar{U}h\rangle_{\mathcal{U}}, \quad h \in \mathcal{U}.$$

Thus, by Theorem 2.4.4, the sequence $(\Delta_{t_k}^2\,\Phi_\Lambda(0|g_0))_k$ converges uniformly to the limit function Ξ_φ on every compact subset $C \subset \mathcal{U}$, i.e., we have

$$\lim_{k\to\infty} \sup_{h\in C} |\Delta^2_{t_k} \Phi_\Lambda(0|g_0)(h) - \Xi_\varphi(h)| = 0. \tag{5.3.55}$$

Let us recall that the convexity and Lipschitz continuity of the quadratic sub-Lagrangian Φ_Λ implies $0 \in \text{int dom } \Phi_\Lambda$. (We refer to the Lemmas 5.3.35, 5.3.36 and to Theorem 2.2.1 for details). Hence, for all k sufficiently large and all $h \in C \subset \mathcal{U}$, the difference quotient $\Delta^2_{t_k} \Phi_\Lambda(0|g_0)(h)$ is well-defined, finite valued, and, consequently, we do not necessarily need to work with ρ-truncations in equation (5.3.55). Specifically, for every sequence $(h^k)_k$, $h^k \to 0$, $h^k \neq 0$, it holds

$$\lim_{k\to\infty} \frac{|\Phi_\Lambda(h^k) - \Phi_\Lambda(0) - \nabla\Phi_\Lambda(0)^\top h^k - \frac{1}{2}\Xi_\varphi(h^k)|}{\frac{1}{2}\|h^k\|^2}$$
$$= \lim_{k\to\infty} |\Delta^2_{\|h^k\|} \Phi_\Lambda(0|g_0)(\tilde{h}^k) - \Xi_\varphi(\tilde{h}^k)| = 0,$$

where $\tilde{h}^k := h^k/\|h^k\|$, $k \in \mathbb{N}$. This shows

$$\Phi_\Lambda(h) - \Phi_\Lambda(0) - \nabla\Phi_\Lambda(0)^\top h - \frac{1}{2}\langle h, \bar{U}^\top \mathcal{H}_\varphi(\bar{x})\bar{U}h\rangle_{\mathcal{U}} = o(\|h\|^2_{\mathcal{U}}), \quad h \to 0,$$

and thus, the symmetric and positive semidefinite matrix $H\Phi_\Lambda(0) := \bar{U}^\top \mathcal{H}_\varphi(\bar{x})\bar{U}$ is a generalized Hessian of Φ_Λ at 0 and a $\mathcal{U}$-Hessian for φ at $\bar{x}$. Let us summarize our latter results in the following theorem.

Theorem 5.3.39. *Let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary, let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable, and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. Furthermore, let $\bar{x} \in \text{dom } \varphi$ be a stationary point of problem $(\mathcal{P})$ and suppose that φ is C^2-fully decomposable at $\bar{x}$. If the strict complementarity condition holds at $\bar{x}$, then the quadratic sub-Lagrangian Φ_Λ has a generalized Hessian $H\Phi_\Lambda(0)$ at 0, which is also a $\mathcal{U}$-Hessian for φ at $\bar{x}$, and it holds*

$$H\Phi_\Lambda(0) = \bar{U}^\top \mathcal{H}_\varphi(\bar{x})\bar{U},$$

where the symmetric, positive semidefinite matrix $\mathcal{H}_\varphi(\bar{x})$ is specified in (5.3.54).

Before proceeding with the next paragraph, let us mention that our proof of Theorem 5.3.39 is strongly motivated by [195] and [196, Theorem 6.7]. In particular, Poliquin and Rockafellar used the concept of epi-subdifferentiability and a similar (but more complex) argumentation to establish existence of second order-type expansions in a much more general context. We are now able to finish the proof of Theorem 5.3.26.

Completion of the proof of Theorem 5.3.26

In this last step of the proof, we will connect the $\mathcal{U}$-Hessian of φ at $\bar{x}$ and the Fréchet derivative of the proximity operator $\text{prox}^\Lambda_\varphi$ at $\bar{u} := \bar{x} - \Lambda^{-1}\nabla f(\bar{x})$. Let us assume that all conditions in Theorem 5.3.39 (or Theorem 5.3.26) are satisfied. Then, as in [130, Proposition

5.2], we obtain

$$\begin{aligned}\mathrm{env}^{\Lambda}_{\varphi}(\bar{u}+h_{\mathcal{U}}\oplus_{\Lambda}0)&=\mathrm{env}^{\Lambda}_{\varphi}(\bar{p}+(\hat{g}_{\mathcal{U}}+h_{\mathcal{U}})\oplus_{\Lambda}\hat{g}_{\mathcal{V}})\\&=\min_{(u,v)\in\mathcal{U}\times\mathcal{V}}\ \varphi(\bar{p}+u\oplus_{\Lambda}v)+\frac{1}{2}\|(\hat{g}_{\mathcal{U}}+h_{\mathcal{U}}-u)\oplus_{\Lambda}(\hat{g}_{\mathcal{V}}-v)\|^2_{\Lambda}\\&=\min_{u\in\mathcal{U}}\left[\min_{v\in\mathcal{V}}\ \varphi(\bar{p}+u\oplus_{\Lambda}v)+\frac{1}{2}\|\hat{g}_{\mathcal{V}}-v\|^2_{\mathcal{V},\Lambda}\right]+\frac{1}{2}\|\hat{g}_{\mathcal{U}}+h_{\mathcal{U}}-u\|^2_{\mathcal{U},\Lambda}\\&=\min_{u\in\mathcal{U}}\ \Phi_{\Lambda}(u)+\frac{1}{2}\|\hat{g}_{\mathcal{V}}\|^2_{\mathcal{V},\Lambda}+\frac{1}{2}\|\hat{g}_{\mathcal{U}}+h_{\mathcal{U}}-u\|^2_{\mathcal{U},\Lambda}\\&=\mathrm{env}^{\bar{U}^\top\Lambda\bar{U}}_{\Phi_{\Lambda}}(\hat{g}_{\mathcal{U}}+h_{\mathcal{U}})+\frac{1}{2}\|\hat{g}_{\mathcal{V}}\|^2_{\mathcal{V},\Lambda},\end{aligned}$$

for all $h=h_{\mathcal{U}}\oplus_{\Lambda}0\in\mathcal{C}(\bar{x})$. (Clearly, since the quadratic sub-Lagrangian is a convex, proper, and lower semicontinuous function and the matrix $\bar{U}^\top\Lambda\bar{U}$ is positive definite, the Moreau envelope $\mathrm{env}^{\bar{U}^\top\Lambda\bar{U}}_{\Phi_{\Lambda}}$ is well-defined). Moreover, due to

$$\mathrm{env}^{\Lambda}_{\varphi}(\bar{u}+0\oplus_{\Lambda}0)=\varphi(\bar{p})+\frac{1}{2}\|\bar{p}-\bar{u}\|^2_{\Lambda}=\varphi(\bar{p}+0\oplus_{\Lambda}0)+\frac{1}{2}\|(\hat{g}_{\mathcal{U}}-0)\oplus_{\Lambda}(\hat{g}_{\mathcal{V}}-0)\|^2_{\Lambda},$$

it immediately follows $\mathrm{prox}^{\bar{U}^\top\Lambda\bar{U}}_{\Phi_{\Lambda}}(\hat{g}_{\mathcal{U}})=0$. Since the quadratic sub-Lagrangian Φ_{Λ} has a generalized Hessian at 0, an important result of Lemaréchal and Sagastizábal [132, Theorem 3.1] implies

$$\nabla^2\mathrm{env}^{\bar{U}^\top\Lambda\bar{U}}_{\Phi_{\Lambda}}(\hat{g}_{\mathcal{U}})=\bar{U}^\top\Lambda\bar{U}-\bar{U}^\top\Lambda\bar{U}\big[H\Phi_{\Lambda}(0)+\bar{U}^\top\Lambda\bar{U}\big]^{-1}\bar{U}^\top\Lambda\bar{U}.$$

Let us note that the proof of [132, Theorem 3.1] strongly relies on [108, Theorem 2.12] and that the results in [108, 132] are only formulated for finite valued, convex functions. However, a careful examination of the different arguments and steps used in the proofs of [132, Theorem 3.1] and [108, Theorem 2.12] shows that finiteness is only required locally in a neighborhood of the point of interest. Thus, since Φ_{Λ} is Lipschitz continuous near 0 and its corresponding subdifferential $\partial\Phi_{\Lambda}$ is nonempty and compact in a neighborhood of 0, the results of Hiriart-Urruty, Lemaréchal, and Sagastizábal are applicable in our situation. Now, on the other hand, invoking Corollary 5.3.33, the proximity operator $\mathrm{prox}^{\Lambda}_{\varphi}$ is Fréchet differentiable at $\bar{u}$ and we obtain $\nabla^2\mathrm{env}^{\Lambda}_{\varphi}(\bar{u})=\Lambda-\Lambda D\mathrm{prox}^{\Lambda}_{\varphi}(\bar{u})$. In particular, this yields

$$\begin{aligned}\bar{U}^\top\Lambda\bar{U}-\bar{U}^\top\Lambda D\mathrm{prox}^{\Lambda}_{\varphi}(\bar{u})\bar{U}=\bar{U}^\top\nabla^2\mathrm{env}^{\Lambda}_{\varphi}(\bar{u})\bar{U}&=\nabla^2\mathrm{env}^{\bar{U}^\top\Lambda\bar{U}}_{\Phi_{\Lambda}}(\hat{g}_{\mathcal{U}})\\&=\bar{U}^\top\Lambda\bar{U}-\bar{U}^\top\Lambda\bar{U}\big[\bar{U}^\top\mathcal{H}_{\varphi}(\bar{x})\bar{U}+\bar{U}^\top\Lambda\bar{U}\big]^{-1}\bar{U}^\top\Lambda\bar{U}.\end{aligned}$$

Consequently, using Lemma 3.3.5 (i), the matrix $\bar{U}^\top\Lambda D\mathrm{prox}^{\Lambda}_{\varphi}(\bar{u})\bar{U}$ has to be positive definite and we can infer

$$\bar{U}^\top\mathcal{H}_{\varphi}(\bar{x})\bar{U}=\bar{U}^\top\Lambda\bar{U}\big[\bar{U}^\top\Lambda D\mathrm{prox}^{\Lambda}_{\varphi}(\bar{u})\bar{U}\big]^{-1}\bar{U}^\top\Lambda\bar{U}-\bar{U}^\top\Lambda\bar{U}.$$

Furthermore, by Lemma 3.3.6, it follows $D\mathrm{prox}^{\Lambda}_{\varphi}(\bar{u})h\in N_{\partial\varphi(\bar{p})}(\bar{g})=\mathcal{C}(\bar{x})=\mathcal{U}$ for all $h\in\mathbb{R}^n$. This readily establishes

$$\bar{V}^\top\Lambda D\mathrm{prox}^{\Lambda}_{\varphi}(\bar{u})=0,$$

and

$$[\Lambda D\mathrm{prox}_\varphi^\Lambda(\bar u)]\bar V = [\Lambda D\mathrm{prox}_\varphi^\Lambda(\bar u)]^\top \bar V = [D\mathrm{prox}_\varphi^\Lambda(\bar u)]^\top \Lambda \bar V = 0,$$

where we used the symmetry of the matrices $\Lambda D\mathrm{prox}_\varphi^\Lambda(\bar u)$ and Λ. Finally, we get the following $\mathcal{U}$-characterization of the Fréchet derivative of $\mathrm{prox}_\varphi^\Lambda$

$$\begin{aligned} D\mathrm{prox}_\varphi^\Lambda(\bar u) &= [\bar U(\bar U^\top \Lambda \bar U)^{-1}\bar U^\top + \bar V(\bar V^\top \Lambda \bar V)^{-1}\bar V^\top]\Lambda D\mathrm{prox}_\varphi^\Lambda(\bar u) \\ &= \bar U(\bar U^\top \Lambda \bar U)^{-1}\bar U^\top \Lambda D\mathrm{prox}_\varphi^\Lambda(\bar u)[\bar U(\bar U^\top \Lambda \bar U)^{-1}\bar U^\top + \bar V(\bar V^\top \Lambda \bar V)^{-1}\bar V^\top]\Lambda \\ &= \bar U(\bar U^\top \Lambda \bar U)^{-1}\big[\bar U^\top \Lambda D\mathrm{prox}_\varphi^\Lambda(\bar u)\bar U\big](\bar U^\top \Lambda \bar U)^{-1}\bar U^\top \Lambda. \end{aligned}$$

Next, setting $A^\top := \Lambda^{\frac12}\bar U(\bar U^\top \Lambda \bar U)^{-1}$ and $B := \bar U^\top \Lambda D\mathrm{prox}_\varphi^\Lambda(\bar u)\bar U$, we obtain

$$\begin{aligned} \big[\Lambda^{\frac12} D\mathrm{prox}_\varphi^\Lambda(\bar u)\Lambda^{-\frac12}\big]^+ &= A^\top B(BAA^\top B)^{-1}(AA^\top)^{-1}A = A^\top B(B[\bar U^\top \Lambda \bar U]^{-1}B)^{-1}[\bar U^\top \Lambda \bar U]A \\ &= A^\top(\bar U^\top \Lambda \bar U)B^{-1}(\bar U^\top \Lambda \bar U)A = \Lambda^{\frac12}\bar U\big[\bar U^\top \Lambda D\mathrm{prox}_\varphi^\Lambda(\bar u)\bar U\big]^{-1}\bar U^\top \Lambda^{\frac12}. \end{aligned}$$

Hence, by defining $\mathcal{Q}_\varphi^\Lambda(\bar u) := \Lambda^{\frac12} D\mathrm{prox}_\varphi^\Lambda(\bar u)\Lambda^{-\frac12}$ and combing our computational results, we have for every $h = h_{\mathcal{U}} \oplus_\Lambda 0 \in \mathcal{C}(\bar x)$

$$\begin{aligned} -\xi_{\varphi,h}^*(\bar g) = \langle h, \mathcal{H}_\varphi(\bar x)h\rangle = \langle h_{\mathcal{U}}, [\bar U^\top \mathcal{H}_\varphi(\bar x)\bar U]h_{\mathcal{U}}\rangle_{\mathcal{U}} &= \langle h_{\mathcal{U}}, \bar U^\top[\Lambda^{\frac12}\mathcal{Q}_\varphi^\Lambda(\bar u)^+\Lambda^{\frac12} - \Lambda]\bar U h_{\mathcal{U}}\rangle_{\mathcal{U}} \\ &= \langle h, [\Lambda^{-\frac12}\mathcal{Q}_\varphi^\Lambda(\bar u)^+\Lambda^{\frac12} - I]h\rangle_\Lambda. \end{aligned}$$

This concludes the proof of Theorem 5.3.26. Although the latter characterization seems to be somewhat complicated, it allows a fully intrinsic description of the curvature of φ in terms of the Fréchet derivative $D\mathrm{prox}_\varphi^\Lambda(\bar u)$. In particular, we can formulate and assess second order necessary and sufficient conditions for problem $(\mathcal{P})$ without knowing a specific decomposition pair (φ_d, F) of φ, the corresponding (unique) Lagrange multiplier $\bar\lambda \in \mathcal{M}(\bar x)$, or the second order derivative of F. Moreover, in contrast to the $\mathcal{U}$-Hessian based formulation, this representation does also not depend on the basis matrices $\bar U$ and $\bar V$. Finally, let us mention that if the parameter matrix Λ satisfies $\Lambda = \lambda^{-1}I$, $\lambda > 0$, then the curvature formula takes the much simpler form

$$-\xi_{\varphi,h}^*(\bar g) = \lambda\langle h, [D\mathrm{prox}_\varphi^{\lambda^{-1}I}(\bar u)^+ - I]h\rangle, \quad \forall\ h \in \mathcal{C}(\bar x).$$

At this point, let us briefly state some additional properties of the matrix $\mathcal{Q}_\varphi^\Lambda(\bar u)$ and of its pseudoinverse, which will be needed in the subsequent section. Clearly, due to Lemma 3.3.5, the matrix $\mathcal{Q}_\varphi^\Lambda(\bar u)$ is symmetric and, for all $h = h_{\mathcal{U}} \oplus_\Lambda 0 \in \mathcal{C}(\bar x) \equiv \mathcal{U}$, it holds

$$\mathcal{Q}_\varphi^\Lambda(\bar u)^+\mathcal{Q}_\varphi^\Lambda(\bar u)\Lambda^{\frac12}h = \Lambda^{\frac12}\bar U B^{-1}\bar U^\top \Lambda^{\frac12}A^\top BA\Lambda^{\frac12}h = \Lambda^{\frac12}\bar U(\bar U^\top \Lambda \bar U)^{-1}\bar U^\top \Lambda h = \Lambda^{\frac12}h.$$

Similarly, we also get $\mathcal{Q}_\varphi^\Lambda(\bar u)\mathcal{Q}_\varphi^\Lambda(\bar u)^+\Lambda^{\frac12}h = \Lambda^{\frac12}h$.

5.4. Nonsingularity conditions

In this section, we combine our latest results and apply the alternative, prox-based representation of the curvature term $-\xi^*_{\varphi,h}(\bar{g})$ to derive nonsingularity conditions for the generalized derivatives of the nonsmooth mapping F^Λ.

We first start with a simple observation that is motivated by a result of Pieper [191, Lemmas 3.14 and 3.15], for a Normal map-related formulation. Let us note that the proof of Lemma 5.4.1 can be seen as a prototype for the subsequent generalizations.

Lemma 5.4.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be convex, proper, and lower semicontinuous. Furthermore, let $x \in \mathbb{R}^n$, $\Lambda \in \mathbb{S}^n_{++}$ be given and let us set $u := x - \Lambda^{-1}\nabla f(x)$. For a fixed element $V \in \partial \mathrm{prox}^\Lambda_\varphi(u)$ suppose that the following second order-type condition is satisfied*

$$h^\top \nabla^2 f(x) h + \langle h, (I - V)h\rangle_\Lambda > 0, \quad \forall\ h \in \mathrm{ran}\ V \setminus \{0\}. \tag{5.4.1}$$

Then, the matrix $W = I - V(I - \Lambda^{-1}\nabla^2 f(x))$ is invertible.

Proof. Let us assume that W is not invertible. Then, there exists $h \in \mathbb{R}^n$, $h \neq 0$, such that $Wh = 0$. Setting $d := (I - \Lambda^{-1}\nabla^2 f(x))h$, this implies $h = Vd$, (i.e., $h \in \mathrm{ran}\ V \setminus \{0\}$), and $Vh = h + V\Lambda^{-1}\nabla^2 f(x)h$. Moreover, we obtain

$$\begin{aligned} 0 = \langle h, Wh\rangle_\Lambda &= \langle h, \Lambda(I - V)h\rangle + \langle Vh, \nabla^2 f(x)h\rangle \\ &= \langle h, \Lambda(I - V)h\rangle + \langle h, \nabla^2 f(x)h\rangle + \langle V\Lambda^{-1}\nabla^2 f(x)h, \nabla^2 f(x)h\rangle \\ &= \langle h, (I - V)h\rangle_\Lambda + \langle h, \nabla^2 f(x)h\rangle + \langle \Lambda V[\Lambda^{-1}\nabla^2 f(x)h], [\Lambda^{-1}\nabla^2 f(x)h]\rangle \\ &\geq h^\top \nabla^2 f(x)h + \langle h, (I - V)h\rangle_\Lambda, \end{aligned}$$

where we used the positive semidefiniteness of the matrix ΛV; see Lemma 3.3.5 (i). However, invoking condition (5.4.1), we deduce $h = 0$ which contradicts our assumption. Hence, the matrix W must be invertible. □

Clearly, if the Hessian $\nabla^2 f(x)$ is positive definite on $\mathrm{ran}\ V \setminus \{0\}$, then the second order-type condition (5.4.1) is satisfied. More specifically, we have the following result.

Lemma 5.4.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be convex, proper, and lower semicontinuous. Furthermore, let $x \in \mathbb{R}^n$, $\Lambda \in \mathbb{S}^n_{++}$ be given and let us set $u := x - \Lambda^{-1}\nabla f(x)$. Suppose that the Hessian $\nabla^2 f(x)$ is positive definite, then every matrix $W \in \mathcal{W}^\Lambda(x)$,*

$$\mathcal{W}^\Lambda(x) := \{W \in \mathbb{R}^{n\times n} : W = I - V(I - \Lambda^{-1}\nabla^2 f(x)),\ V \in \partial \mathrm{prox}^\Lambda_\varphi(u)\},$$

is invertible and there exists $C \in \mathbb{R}$ such that

$$\|W^{-1}\| \leq C, \quad \forall\ W \in \mathcal{W}^\Lambda(x).$$

Proof. Due to Lemma 5.4.1, we only need to verify that the matrices $W \in \mathcal{W}^\Lambda(x)$ are *uniformly boundedly* invertible. Therefore, let $V \in \partial \mathrm{prox}^\Lambda_\varphi(u)$ and $r \in \mathbb{R}^n$ be arbitrary and consider the equation

$$Wh = (I - V(I - \Lambda^{-1}\nabla^2 f(x))) \cdot h = r.$$

Multiplying both sides of the latter equation with $\Lambda^{\frac{1}{2}}$ yields

$$(I - \Lambda^{\frac{1}{2}} V \Lambda^{-\frac{1}{2}}(I - \Lambda^{-\frac{1}{2}}\nabla^2 f(x)\Lambda^{-\frac{1}{2}})) \cdot \Lambda^{\frac{1}{2}} h = \Lambda^{\frac{1}{2}} r. \tag{5.4.2}$$

Now, Lemma 3.3.5 (i) implies that the matrix $\tilde{V} := \Lambda^{\frac{1}{2}} V \Lambda^{-\frac{1}{2}} = \Lambda^{-\frac{1}{2}}[\Lambda V]\Lambda^{-\frac{1}{2}}$ is symmetric, positive semidefinite and its eigenvalues are bounded by 1. Let

$$\Lambda^{-\frac{1}{2}}[\Lambda V]\Lambda^{-\frac{1}{2}} = P\Sigma P^\top, \quad \Sigma = \mathrm{diag}(\sigma_1, ..., \sigma_n),$$

be a corresponding eigenvalue decomposition of $\tilde{V}$ and let us define the index sets

$$\alpha := \{i : \sigma_i = 0\}, \quad \beta := \{i : \sigma_i \in (0, \varepsilon)\}, \quad \gamma := \{i : \sigma_i \in [\varepsilon, 1]\}$$

for some arbitrary but fixed $\varepsilon \in (0,1)$. Then, setting $\tilde{H} := P^\top \Lambda^{-\frac{1}{2}} \nabla^2 f(x) \Lambda^{-\frac{1}{2}} P$, equation (5.4.2) can be equivalently rewritten as follows

$$[I - \Sigma(I - \tilde{H})] \cdot P^\top \Lambda^{\frac{1}{2}} h = P^\top \Lambda^{\frac{1}{2}} r.$$

Moreover, by setting $\tilde{h} := P^\top \Lambda^{\frac{1}{2}} h$, $\tilde{r} := P^\top \Lambda^{\frac{1}{2}} r$, $\mathcal{B} := \mathrm{diag}(\sigma_\beta)$, $\mathcal{G} := \mathrm{diag}(\sigma_\gamma)$, and $\Gamma := \mathcal{G}^{-1}(I - \mathcal{G})$ the latter system can also be discussed w.r.t. the index sets α, β, and γ,

$$\begin{pmatrix} I & & \\ & I & \\ & & \mathcal{G} \end{pmatrix} \left[\begin{pmatrix} 0 & 0 & 0 \\ \mathcal{B}\tilde{H}_{[\beta\alpha]} & \mathcal{B}\tilde{H}_{[\beta\beta]} & \mathcal{B}\tilde{H}_{[\beta\gamma]} \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} I & 0 & 0 \\ 0 & I - \mathcal{B} & 0 \\ \tilde{H}_{[\gamma\alpha]} & \tilde{H}_{[\gamma\beta]} & \tilde{H}_{[\gamma\gamma]} + \Gamma \end{pmatrix} \right] \begin{pmatrix} \tilde{h}_\alpha \\ \tilde{h}_\beta \\ \tilde{h}_\gamma \end{pmatrix} = \begin{pmatrix} \tilde{r}_\alpha \\ \tilde{r}_\beta \\ \tilde{r}_\gamma \end{pmatrix}.$$

Now, the remaining part of the proof utilizes a technique that was applied in [156, Lemma 4.3.2] to prove bounded invertibility in an ℓ_1-setting. Specifically, let $\mathcal{R}$, $\mathcal{S}$, and $\mathcal{T}$ denote the three different matrices occurring in the last equation. Then, the idea is to determine $\varepsilon \in (0,1)$ in a way such that Banach's perturbation lemma is applicable and invertibility of the matrix $\mathcal{R} \cdot (\mathcal{S} + \mathcal{T})$ can be inferred.

Using $\mathcal{G}^{-1}_{[ii]} \in [1, \varepsilon^{-1}]$ and $\Gamma_{[ii]} \in [0, \varepsilon^{-1}(1-\varepsilon)]$, we obtain the following estimates:

- $\|\mathcal{R}^{-1}\| < \varepsilon^{-1}$.
- $\lambda_{\min}(\tilde{H}_{[\gamma\gamma]} + \Gamma) \geq \lambda_{\min}(\tilde{H}_{[\gamma\gamma]}) \geq \lambda_{\min}(\Lambda^{-\frac{1}{2}}\nabla^2 f(x)\Lambda^{-\frac{1}{2}}) \geq \lambda_{\max}(\Lambda)^{-1}\lambda_{\min}(\nabla^2 f(x))$.
- For any arbitrary pair of index sets $\mathcal{I}, \mathcal{J} \subset \{1, ..., n\}$, it holds

$$\begin{aligned} \|\tilde{H}_{[\mathcal{I}\mathcal{J}]}\|_2 = \|I_{[\mathcal{I}\cdot]}\tilde{H}I_{[\cdot\mathcal{J}]}\|_2 &\leq \|I_{[\mathcal{I}\cdot]}\|_2\|\tilde{H}\|_2\|I_{[\cdot\mathcal{J}]}\|_2 \\ &= \|\Lambda^{-\frac{1}{2}}\nabla^2 f(x)\Lambda^{-\frac{1}{2}}\|_2 \leq \lambda_{\min}(\Lambda)^{-1}\lambda_{\max}(\nabla^2 f(x)) =: C_s. \end{aligned}$$

- For all $v = (v_\alpha^\top, v_\beta^\top, v_\gamma^\top)^\top \in \mathbb{R}^n$, it holds

$$\|\mathcal{S}v\|^2 \leq C_s^2\|\mathcal{B}\|^2 \cdot (\|v_\alpha\| + \|v_\beta\| + \|v_\gamma\|)^2 \leq 3C_s^2\varepsilon^2 \cdot \|v\|^2$$

and a simple block elimination yields

$$\|\mathcal{T}^{-1}v\|^2 \leq \|v_\alpha\|^2 + (1-\varepsilon)^{-2}\|v_\beta\|^2 + C_{t,\varepsilon}^2 \cdot (\|v_\alpha\| + \|v_\beta\| + \|v_\gamma\|)^2$$
$$\leq (1 + (1-\varepsilon)^{-2} + 3C_\varepsilon^2) \cdot \|v\|^2,$$

where $C_\varepsilon := \max\left\{\frac{\lambda_{\max}(\Lambda)}{\lambda_{\min}(\nabla^2 f(x))}, \frac{\kappa(\Lambda)\kappa(\nabla^2 f(x))}{1-\varepsilon}\right\}$.

Clearly, for every $0 < \varepsilon \leq \bar{\varepsilon} := 1 - 0.5\sqrt{2}$, the constant C_ε is bounded above by $C_{\bar{\varepsilon}}$ and for the specific choice $\varepsilon := \bar{\varepsilon}\min\{1, (3C_s\sqrt{1+C_{\bar{\varepsilon}}^2})^{-1}\} \leq \bar{\varepsilon}$ it follows

$$\|\mathcal{S}\mathcal{T}^{-1}\| \leq 3C_s\varepsilon\sqrt{1+C_\varepsilon^2} \leq \bar{\varepsilon} < 1.$$

Thus, by Banach's perturbation lemma, we establish

$$\|[\mathcal{R}\cdot(\mathcal{S}+\mathcal{T})]^{-1}\| \leq \|\mathcal{R}^{-1}\| \cdot \frac{\|\mathcal{T}^{-1}\|}{1 - \|\mathcal{S}\mathcal{T}^{-1}\|} \leq \frac{\sqrt{3(1+C_{\bar{\varepsilon}}^2)}}{\varepsilon(1-\bar{\varepsilon})} =: C_{rst}.$$

At this point, let us emphasize that the constants ε and C_{rst} are independent of the index sets α, β, and γ, the matrix P, and of the eigenvalues of V. Consequently, the latter estimate holds uniformly for all $V \in \partial\text{prox}_\varphi^\Lambda(u)$. Now, reconsidering our initial system of equations, we readily obtain

$$\|\tilde{h}\|^2 = \|\tilde{h}_\alpha\|^2 + \|\tilde{h}_\beta\|^2 + \|\tilde{h}_\gamma\|^2 \leq C_{rst}^2(\|\tilde{r}_\alpha\|^2 + \|\tilde{r}_\beta\|^2 + \|\tilde{r}_\gamma\|^2) = C_{rst}^2\|\tilde{r}\|^2.$$

Finally, due to

$$\|\tilde{h}\| = \|h\|_\Lambda \geq \sqrt{\lambda_{\min}(\Lambda)}\|h\|, \quad \|\tilde{r}\| = \|r\|_\Lambda \leq \sqrt{\lambda_{\max}(\Lambda)}\|r\|,$$

we can conclude that the bound in Lemma 5.4.2 holds with $C := \sqrt{\kappa(\Lambda)}C_{rst}$. □

Remark 5.4.3. Suppose that $\mathcal{V}^\Lambda \subset \mathbb{R}^{n\times n}$ is a given subset and that the matrices ΛV and $\Lambda(I-V)$ are symmetric and positive semidefinite for every $V \in \mathcal{V}^\Lambda$. Then, the proof of Lemma 5.4.2 shows that the (possibly larger) collection of matrices

$$\tilde{\mathcal{W}}^\Lambda(x) := \{W \in \mathbb{R}^{n\times n} : W = I - V(I - \Lambda^{-1}\nabla^2 f(x)),\ V \in \mathcal{V}^\Lambda\}$$

is also uniformly boundedly invertible (with the same constant C).

We now present the main result of this section.

Theorem 5.4.4. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable and let $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. Furthermore, let $\bar{x} \in \text{dom}\,\varphi$ be a stationary point of problem $(\mathcal{P})$ and suppose that φ is C^2-fully decomposable at $\bar{x}$ and that the strict complementarity condition,*

$$-\nabla f(\bar{x}) \in \text{ri}\ \partial\varphi(\bar{x}),$$

is satisfied. Then, the proximity operator $\text{prox}_\varphi^\Lambda$ is Fréchet differentiable at $\bar{u} := \bar{x} - \Lambda^{-1}\nabla f(\bar{x})$

and the second order conditions

$$h^\top \nabla^2 f(\bar{x})h + \langle h, [\Lambda^{-\frac{1}{2}} \mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \Lambda^{\frac{1}{2}} - I]h\rangle_\Lambda > 0, \quad \forall\ h \in \mathcal{C}(\bar{x}) \setminus \{0\} \tag{5.4.3}$$

are necessary and sufficient for the quadratic growth condition 5.3.10. Moreover, if the latter second order sufficient condition is satisfied at $\bar{x}$, then the following statements do hold:

(i) *The stationary point $\bar{x}$ is a strict local minimum and an isolated stationary point of problem $(\mathcal{P})$.*

(ii) *If, in addition, the proximity operator* $\mathrm{prox}_\varphi^\Lambda$ *is semismooth at $\bar{u}$, then the function F^Λ is strictly differentiable at $\bar{x}$ and its Fréchet derivative $DF^\Lambda(\bar{x})$ is nonsingular.*

Proof. The Fréchet differentiability of the proximity operator $\mathrm{prox}_\varphi^\Lambda$ follows from Corollary 5.3.33. Now, let (φ_d, F) be an appropriate decomposition pair of the function φ. Then, Theorem 5.3.26 and our preceding discussion implies

$$\langle h, [\Lambda^{-\frac{1}{2}} \mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \Lambda^{\frac{1}{2}} - I]h\rangle_\Lambda = \langle \bar{\lambda}, D^2 F(\bar{x})[h,h]\rangle$$

for all $h \in \mathcal{C}(\bar{x})$. (Here, $\bar{\lambda}$ denotes the unique Lagrange multiplier associated with the decomposed problem (5.3.33)). Consequently, the second order conditions (5.4.3) and (5.3.11) coincide and applying Theorem 5.3.6, we can infer that the conditions (5.4.3) are necessary and sufficient for the quadratic growth condition (5.3.10). Moreover, $\bar{x}$ is also a strict locally optimal solution and an isolated stationary point of problem $(\mathcal{P})$. Let us continue with the proof of the second part.

Since the proximity operator is Fréchet differentiable and semismooth at $\bar{u}$, Theorem 2.6.7 implies that $\mathrm{prox}_\varphi^\Lambda$ is strict differentiable at $\bar{u}$. Hence, as a composition of strictly differentiable functions, the mapping F^Λ is also strictly differentiable at $\bar{x}$. In particular, it follows $\partial F^\Lambda(\bar{x}) = \{DF^\Lambda(\bar{x})\}$. Next, as in Lemma 5.4.1, suppose there exists $h \in \mathbb{R}^n \setminus \{0\}$ such that $DF^\Lambda(\bar{x})h = 0$. Then, it follows

$$DF^\Lambda(\bar{x})h = 0 \quad \Longleftrightarrow \quad h = D\mathrm{prox}_\varphi^\Lambda(\bar{u})(I - \Lambda^{-1}\nabla^2 f(\bar{x}))h. \tag{5.4.4}$$

Consequently, Lemma 3.3.6 implies $h \in \mathcal{C}(\bar{x}) \setminus \{0\}$ and we have

$$\mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \mathcal{Q}_\varphi^\Lambda(\bar{u})[\Lambda^{-\frac{1}{2}} \nabla^2 f(\bar{x})h] = \Lambda^{\frac{1}{2}} h - \mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \Lambda^{\frac{1}{2}} h.$$

Setting $V := D\mathrm{prox}_\varphi^\Lambda(\bar{u})$, we obtain

$$\begin{aligned}
\langle h, DF^\Lambda(\bar{x})h\rangle_\Lambda &= \langle h, \Lambda(I-V)h\rangle + \langle \Lambda V h, \Lambda^{-1}\nabla^2 f(\bar{x})h\rangle \\
&= h^\top \nabla^2 f(\bar{x})h + \langle h, \Lambda(I-V)h\rangle + \langle \mathcal{Q}_\varphi^\Lambda(\bar{u})[\Lambda^{-\frac{1}{2}}\nabla^2 f(\bar{x})h], [\Lambda^{-\frac{1}{2}}\nabla^2 f(\bar{x})h]\rangle \\
&= h^\top \nabla^2 f(\bar{x})h + \langle h, \Lambda(I-V)h\rangle + \langle \Lambda V \Lambda^{-1}\nabla^2 f(\bar{x})h, [I - \Lambda^{-\frac{1}{2}} \mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \Lambda^{\frac{1}{2}}]h\rangle \\
&= h^\top \nabla^2 f(\bar{x})h + \langle (I-V)h, \Lambda^{\frac{1}{2}} \mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \Lambda^{\frac{1}{2}} h\rangle \\
&= h^\top \nabla^2 f(\bar{x})h + \langle h, \Lambda^{-\frac{1}{2}} \mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \Lambda^{\frac{1}{2}} h\rangle_\Lambda - \langle \mathcal{Q}_\varphi^\Lambda(\bar{u})\Lambda^{\frac{1}{2}} h, \mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \Lambda^{\frac{1}{2}} h\rangle \\
&= h^\top \nabla^2 f(\bar{x})h + \langle h, [\Lambda^{-\frac{1}{2}} \mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \Lambda^{\frac{1}{2}} - I]h\rangle_\Lambda,
\end{aligned}$$

where we used (5.4.4), the properties of the Moore-Penrose inverse, i.e.,

$$\mathcal{Q}^\Lambda_\varphi(\bar u) = \mathcal{Q}^\Lambda_\varphi(\bar u)\mathcal{Q}^\Lambda_\varphi(\bar u)^+\mathcal{Q}^\Lambda_\varphi(\bar u),$$

the symmetry of the matrices ΛV, $\mathcal{Q}^\Lambda_\varphi(\bar u)$, and $h \in \mathcal{C}(\bar x) \equiv \mathcal{U}$. Clearly, the second order sufficient conditions (5.4.3) now imply $h = 0$, which contradicts our assumption. Consequently, the matrix $DF^\Lambda(\bar x)$ has to be nonsingular. This finishes the proof of Theorem 5.4.4. □

In the following, we briefly summarize our previous results and connect our observations to the convergence theory of the semismooth Newton method. Therefore, let the functions $f : \mathbb{R}^n \to \mathbb{R}$, $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ be given and suppose that the assumptions (A.2)–(A.3) are satisfied. Let the sequences $(x^k)_k$ and $(\Lambda_k)_k$ be generated by Algorithm 2 and let $x^* \in \mathbb{R}^n$ and $\Lambda_* \in \mathbb{S}^n_{++}$ be accumulation points of the sequences $(x^k)_k$ and $(\Lambda_k)_k$, respectively. If the parameter matrices Λ_k, $k \in \mathbb{N}$, remain in a bounded set $\mathcal{K} \subset \mathbb{S}^n_{++}$, i.e., if assumption (B) is satisfied, then Theorem 4.3.2 implies that x^* is a stationary point of problem $(\mathcal{P})$. Moreover, if φ is C^2-fully decomposable at x^* with decomposition pair (φ_d, F) and if the second order sufficient conditions

$$(5.4.5) \qquad h^\top \nabla^2 f(x^*)h + \langle \bar\lambda, D^2F(x^*)[h,h]\rangle > 0, \quad \forall\, h \in \mathcal{C}(x^*) \setminus \{0\}, \quad \bar\lambda \in \mathcal{M}(x^*),$$

hold at x^*, then Theorem 5.3.11 shows that condition (D.5) is fulfilled. Additionally, if the accumulation points x^* and Λ_* satisfy the conditions (D.1)–(D.3) and if the strict complementarity condition

$$(5.4.6) \qquad -\nabla f(x^*) \in \mathrm{ri}\ \partial\varphi(x^*)$$

holds at x^*, the latter second order sufficient conditions can be equivalently represented as follows

$$(5.4.7) \qquad h^\top \nabla^2 f(x^*)h + \langle h, [\Lambda_*^{-\frac{1}{2}}\mathcal{Q}^{\Lambda_*}_\varphi(u^*)^+\Lambda_*^{\frac{1}{2}} - I]h\rangle_{\Lambda_*} > 0, \quad \forall\, h \in \mathcal{C}(x^*) \setminus \{0\},$$

where $u^* = x^* - \Lambda_*^{-1}\nabla f(x^*)$ and $\mathcal{Q}^{\Lambda_*}_\varphi(u^*) = \Lambda_*^{\frac{1}{2}} D\mathrm{prox}^{\Lambda_*}_\varphi(u^*)\Lambda_*^{-\frac{1}{2}}$. Furthermore, in this situation, the second order conditions (5.4.5) or (5.4.7) imply that the elements of Clarke's subdifferential $\partial F^{\Lambda_*}(x^*)$ are uniformly boundedly invertible and consequently, by Remark 4.3.7, assumption (D.4) is satisfied.

Thus, in short, if the nonsmooth function is C^2-fully decomposable and if the strict complementarity condition is fulfilled, then the second order sufficient conditions (5.4.5) or (5.4.7) essentially yield fast local convergence of Algorithm 2.

We conclude this subsection with two illustrating examples.

Example 5.4.5 (Group sparsity). Let us reconsider the group-sparse optimization problem

$$\min_x \ f(x) + \varphi(x), \quad \varphi(x) := \sum_{i=1}^s \omega_i \|x_{g_i}\|_2,$$

and let the groups g_i, $i = 1, ..., s$, form a disjoint partitioning of the set $\{1, ..., n\}$. Moreover, let $\bar{x} \in \mathbb{R}^n$ be a stationary point of the latter problem and let us set

$$\Lambda_{[g_i g_j]} = \begin{cases} \Lambda^i := \lambda_i^{-1} I & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases} \qquad \lambda_i > 0, \quad 1 \le i, j \le s.$$

As usual, we then define $\bar{u} := \bar{x} - \Lambda^{-1} \nabla f(\bar{x})$ and consider the index sets $\mathcal{A}(\bar{x}) := \{i : \bar{x}_{g_i} = 0\}$ and $\mathcal{I}(\bar{x}) := \{i : \bar{x}_{g_i} \neq 0\}$. As in Example 4.2.18, the proximity operator $\mathrm{prox}_\varphi^\Lambda(\bar{u})$ has the following group-wise representation:

$$\mathrm{prox}_\varphi^\Lambda(\bar{u})_{g_i} = \mathrm{prox}_{\omega_i \|\cdot\|_2}^{\lambda_i^{-1} I}(\bar{u}_{g_i}) = \bar{u}_{g_i} - \mathcal{P}_{B_{\|\cdot\|_2}(0;\omega_i\lambda_i)}(\bar{u}_{g_i}) = \frac{\bar{u}_{g_i}}{\|\bar{u}_{g_i}\|_2} \max\{\|\bar{u}_{g_i}\|_2 - \omega_i \lambda_i, 0\}.$$

Now, the stationarity of $\bar{x}$ implies

$$\begin{cases} i \in \mathcal{A}(\bar{x}) & \implies & \mathrm{prox}_\varphi^\Lambda(\bar{u})_{g_i} = 0 & \wedge & \|\bar{u}_{g_i}\|_2 \le \omega_i \lambda_i, \\ i \in \mathcal{I}(\bar{x}) & \implies & \mathrm{prox}_\varphi^\Lambda(\bar{u})_{g_i} \neq 0 & \wedge & \|\bar{u}_{g_i}\|_2 > \omega_i \lambda_i \end{cases}$$

and, by Corollary 5.3.33, the proximity operator is Fréchet differentiable at $\bar{u}$ if and only if the index set

$$\mathcal{A}_\pm(\bar{x}) = \{i \in \mathcal{A}(\bar{x}) : \|\bar{u}_{g_i}\|_2 = \omega_i \lambda_i\} = \{i \in \mathcal{A}(\bar{x}) : \|\nabla f(\bar{x})_{g_i}\|_2 = \omega_i\}$$

is empty. In this case, the Fréchet derivative of $\mathrm{prox}_\varphi^\Lambda$ at $\bar{u}$ is given by

$$D\mathrm{prox}_\varphi^\Lambda(\bar{u})_{[g_i g_j]} = \begin{cases} \Xi^i & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \qquad \Xi^i := \begin{cases} 0 & \text{if } i \in \mathcal{A}(\bar{x}), \\ \frac{\|\bar{u}_{g_i}\|_2 - \omega_i\lambda_i}{\|\bar{u}_{g_i}\|_2} I + \frac{\omega_i \lambda_i}{\|\bar{u}_{g_i}\|_2^3} \bar{u}_{g_i} \bar{u}_{g_i}^\top & \text{if } i \in \mathcal{I}(\bar{x}) \end{cases}$$

and it holds

$$\left[\Lambda^{\frac{1}{2}} D\mathrm{prox}_\varphi^\Lambda(\bar{u}) \Lambda^{-\frac{1}{2}}\right]^+_{[g_i g_j]} = \begin{cases} \hat{\Xi}^i & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \qquad \hat{\Xi}^i = \begin{cases} 0 & \text{if } i \in \mathcal{A}(\bar{x}), \\ [\Xi^i]^{-1} & \text{if } i \in \mathcal{I}(\bar{x}). \end{cases}$$

Moreover, for $i \in \mathcal{I}(\bar{x})$, the matrix $\hat{\Xi}^i$ can be calculated explicitly by using the Sherman-Morrison-Woodbury formula:

$$\hat{\Xi}^i = \left(1 + \frac{\omega_i \lambda_i}{\|\bar{u}_{g_i}\|_2 - \omega_i \lambda_i}\right) I - \frac{\omega_i \lambda_i}{\|\bar{u}_{g_i}\|_2 - \omega_i \lambda_i} \cdot \frac{\bar{u}_{g_i} \bar{u}_{g_i}^\top}{\|\bar{u}_{g_i}\|_2^2}.$$

Again, from the stationarity of $\bar{x}$, we deduce $\|\bar{x}_{g_i}\|_2 = |\|\bar{u}_{g_i}\|_2 - \omega_i \lambda_i| = \|\bar{u}_{g_i}\|_2 - \omega_i \lambda_i$, and $\|\bar{u}_{g_i}\|_2 \cdot \bar{x}_{g_i} = \|\bar{x}_{g_i}\|_2 \cdot \bar{u}_{g_i}$ for all $i \in \mathcal{I}(\bar{x})$. Finally, this shows

$$\begin{aligned} &\langle h, [\Lambda^{-\frac{1}{2}} \mathcal{Q}_\varphi^\Lambda(\bar{u})^+ \Lambda^{\frac{1}{2}} - I] h \rangle_\Lambda \\ &\quad = \sum_{i=1}^s \lambda_i^{-1} h_{g_i}^\top [\hat{\Xi}^i - I] h_{g_i} = \sum_{i \in \mathcal{I}(\bar{x})} h_{g_i}^\top \left[\frac{\omega_i}{\|\bar{x}_{g_i}\|} I - \frac{\omega_i}{\|\bar{x}_{g_i}\|_2^3} \bar{x}_{g_i} \bar{x}_{g_i}^\top\right] h_{g_i} - \sum_{i \in \mathcal{A}(\bar{x})} \frac{1}{\lambda_i} \|h_{g_i}\|_2^2. \end{aligned}$$

For $h \in \mathcal{C}(\bar{x})$, the latter formula reduces to

$$\langle h, [\Lambda^{-\frac{1}{2}}\mathcal{Q}_{\varphi}^{\Lambda}(\bar{u})^{+}\Lambda^{\frac{1}{2}} - I]h\rangle_{\Lambda} = \sum_{i \in \mathcal{I}(\bar{x})} h_{g_i}^{\top}\left[\frac{\omega_i}{\|\bar{x}_{g_i}\|}I - \frac{\omega_i}{\|\bar{x}_{g_i}\|_2^3}\bar{x}_{g_i}\bar{x}_{g_i}^{\top}\right]h_{g_i}$$

and coincides with the representation of the curvature term $-\xi_{\varphi,h}^{*}(-\nabla f(\bar{x}))$ that was calculated in Example 5.3.13, as expected. Thus, since the proximity operator $\mathrm{prox}_{\varphi}^{\Lambda}$ is a semismooth function, (see, e.g., Example 4.2.18), the strict complementarity condition,

$$-\nabla f(\bar{x}) \in \mathrm{ri}\ \partial\varphi(\bar{x}) \quad \Longleftrightarrow \quad \mathcal{A}_{\pm}(\bar{x}) = \emptyset,$$

and the second order sufficient conditions

$$h^{\top}\nabla^2 f(\bar{x})h + \langle h, [\Lambda^{-\frac{1}{2}}\mathcal{Q}_{\varphi}^{\Lambda}(\bar{u})^{+}\Lambda^{\frac{1}{2}} - I]h\rangle_{\Lambda} > 0, \quad \forall\ h \in \mathcal{C}(\bar{x}) \setminus \{0\}$$

ensure invertibility of the Fréchet derivative of F^{Λ} at $\bar{x}$ and locally q-superlinear convergence of Algorithm 2.

Example 5.4.6 (Semidefinite programming). Next, we want to apply our nonsingularity results to semidefinite programs of the form

$$\min\ f(X) \quad \mathrm{s.t.} \quad X \in \mathbb{S}_{+}^{n},$$

where $f : \mathbb{S}^n \to \mathbb{R}$ is a twice continuously differentiable function. Let $\bar{X} \in \mathbb{S}_{+}^{n}$ be a stationary point of the latter problem and let us set $\bar{U} := \bar{X} - \lambda\nabla f(\bar{X})$, $\lambda > 0$. As in Example 5.3.34, we consider the following spectral decomposition of $\bar{U}$

$$\bar{U} = P\Sigma P^{\top}, \quad \Sigma = \mathrm{diag}(\sigma) \in \mathbb{S}^n,$$

and the associated index sets

$$\alpha := \{i : \sigma_i > 0\}, \quad \beta := \{i : \sigma_i = 0\}, \quad \gamma := \{i : \sigma_i < 0\}.$$

Moreover, let us set

$$\bar{U}_{+} := \mathrm{prox}_{\varphi}^{\lambda I}(\bar{U}) = \mathcal{P}_{\mathbb{S}_{+}^{n}}(\bar{U}) = P\,\mathrm{diag}(\max\{\sigma, 0\})P^{\top}.$$

In Example 5.3.34, we have seen that the strict complementarity condition is equivalent to the invertibility of the matrix $\bar{U}$. Thus, in this case, we have $\beta = \emptyset$ and the metric projection $\mathcal{P}_{\mathbb{S}_{+}^{n}}$ is Fréchet differentiable at $\bar{U}$. In particular, by using [46, Proposition 4.3], it holds

$$D\mathcal{P}_{\mathbb{S}_{+}^{n}}(\bar{U})[H] = P(\Omega \odot (P^{\top}HP))P^{\top}, \ \Omega_{[ij]} = \begin{cases} \dfrac{\max\{\sigma_i, 0\} - \max\{\sigma_j, 0\}}{\sigma_i - \sigma_j} & \text{if } \sigma_i \neq \sigma_j, \\ 1 & \text{if } \sigma_i = \sigma_j,\ i \in \alpha, \\ 0 & \text{if } \sigma_i = \sigma_j,\ i \in \gamma, \end{cases}$$

for all $H \in \mathbb{S}^n$ and $1 \leq i, j \leq n$. In the following, let us suppose that the eigenvalues of $\bar{U}$

are arranged in decreasing order. This implies $P = (P_{[\cdot\alpha]}, P_{[\cdot\gamma]})$ and setting $\bar{H} := P^\top H P$, the term $D\mathcal{P}_{\mathbb{S}^n_+}(\bar{U})[H]$ can be further simplified to

$$D\mathcal{P}_{\mathbb{S}^n_+}(\bar{U})[H] = P \begin{pmatrix} \bar{H}_{[\alpha\alpha]} & \Omega_{[\alpha\gamma]} \odot \bar{H}_{[\alpha\gamma]} \\ \bar{H}^\top_{[\alpha\gamma]} \odot \Omega^\top_{[\alpha\gamma]} & 0 \end{pmatrix} P^\top.$$

Moreover, due to $\Omega_{[ij]} = \frac{\sigma_i}{\sigma_i - \sigma_j} \neq 0$ for all $i \in \alpha$, $j \in \gamma$, it immediately follows

$$[D\mathcal{P}_{\mathbb{S}^n_+}(\bar{U})]^+[H] = P \begin{pmatrix} \bar{H}_{[\alpha\alpha]} & \bar{H}_{[\alpha\gamma]} \oslash \Omega_{[\alpha\gamma]} \\ \bar{H}^\top_{[\alpha\gamma]} \oslash \Omega^\top_{[\alpha\gamma]} & 0 \end{pmatrix} P^\top.$$

Now, defining

$$\Omega^\oslash \in \mathbb{R}^{n\times n}, \quad \Omega^\oslash_{[ij]} := \begin{cases} -\dfrac{\sigma_j}{\sigma_i} & \text{if } i \in \alpha, j \in \gamma, \\ 0 & \text{otherwise,} \end{cases}$$

we readily obtain

$$\begin{aligned} \langle H, [D\mathcal{P}_{\mathbb{S}^n_+}(\bar{U})]^+[H] - H \rangle &= \mathrm{tr}\left(P\bar{H} \begin{pmatrix} 0 & \bar{H}_{[\alpha\gamma]} \odot \Omega^\oslash_{[\alpha\gamma]} \\ \bar{H}^\top_{[\alpha\gamma]} \odot \Omega^{\oslash,\top}_{[\alpha\gamma]} & -\bar{H}_{[\gamma\gamma]} \end{pmatrix} P^\top \right) \\ &= \mathrm{tr}\left(\begin{pmatrix} \bar{H}_{[\alpha\gamma]}(\bar{H}^\top_{[\alpha\gamma]} \odot \Omega^{\oslash,\top}_{[\alpha\gamma]}) & \star \\ \star & \bar{H}^\top_{[\alpha\gamma]}(\bar{H}_{[\alpha\gamma]} \odot \Omega^\oslash_{[\alpha\gamma]}) - \bar{H}_{[\gamma\gamma]}\bar{H}_{[\gamma\gamma]} \end{pmatrix} \right) \\ &= 2 \cdot \mathrm{tr}(\bar{H}^\top_{[\alpha\gamma]}(\bar{H}_{[\alpha\gamma]} \odot \Omega^\oslash_{[\alpha\gamma]})) - \|\bar{H}_{[\gamma\gamma]}\|_F^2, \end{aligned}$$

where we used the symmetry of H and the invariance of the trace operation under cyclic permutations. Furthermore, an easy calculation yields

$$\begin{aligned} \mathrm{tr}(\bar{H}^\top_{[\alpha\gamma]}(\bar{H}_{[\alpha\gamma]} \odot \Omega^\oslash_{[\alpha\gamma]})) &= \mathrm{tr}\left(\bar{H} \begin{pmatrix} 0 & \\ & -\mathrm{diag}(\sigma_\gamma) \end{pmatrix} \bar{H} \begin{pmatrix} \mathrm{diag}(\mathbb{1} \oslash \sigma_\alpha) & \\ & 0 \end{pmatrix} \right) \\ &= \mathrm{tr}\left(P \begin{pmatrix} 0 & \\ & -\mathrm{diag}(\sigma_\gamma) \end{pmatrix} P^\top H P \begin{pmatrix} \mathrm{diag}(\mathbb{1} \oslash \sigma_\alpha) & \\ & 0 \end{pmatrix} P^\top H \right) \end{aligned}$$

and from the stationarity of $\bar{X}$ we deduce

$$\bar{X} = \bar{U}_+ = P \begin{pmatrix} \mathrm{diag}(\sigma_\alpha) & \\ & 0 \end{pmatrix} P^\top, \quad \lambda \nabla f(\bar{X}) = \bar{X} - \bar{U} = P \begin{pmatrix} 0 & \\ & -\mathrm{diag}(\sigma_\gamma) \end{pmatrix} P^\top.$$

Consequently, we have

$$\mathrm{tr}(\bar{H}^\top_{[\alpha\gamma]}(\bar{H}_{[\alpha\gamma]} \odot \Omega^\oslash_{[\alpha\gamma]})) = \lambda \langle \nabla f(\bar{X}), H\bar{X}^+ H \rangle$$

and

$$\langle H, [D\mathcal{P}_{\mathbb{S}^n_+}(\bar{U})]^+[H] - H \rangle_\Lambda = 2\langle \nabla f(\bar{X}), H\bar{X}^+ H \rangle - \frac{1}{\lambda}\|\bar{H}_{[\gamma\gamma]}\|_F^2, \quad \Lambda = \lambda^{-1} I.$$

Hence, since for all $H \in \mathcal{C}(\bar{X}) \equiv \mathcal{C}_{\bar{U}}(\bar{U}_+)$, it holds $\bar{H}_{[\gamma\gamma]} = P^\top_{[\cdot\gamma]} H P_{[\cdot\gamma]} = 0$ (see Example

5.3.34), the curvature term reduces to the well-known formula

$$\langle H, [D\mathcal{P}_{\mathbb{S}^n_+}(\bar{U})]^+[H] - H\rangle_\Lambda = 2\langle \nabla f(\bar{X}), H\bar{X}^+H\rangle.$$

Let us also refer to [215] and [27, Section 5.3 and 5.3.5] for more details on second order conditions for semidefinite programs. In particular, since the metric projection $\mathcal{P}_{\mathbb{S}^n_+}$ is a semismooth function (see, e.g., [229]), the strict complementarity condition and the second order optimality condition,

$$\nabla^2 f(\bar{X})[H, H] + 2\langle \nabla f(\bar{X}), H\bar{X}^+H\rangle > 0, \quad \forall\ H \in \mathcal{C}(\bar{X}) \setminus \{0\},$$

guarantee invertibility of the Fréchet derivative of $F^\Lambda(X) := X - \mathcal{P}_{\mathbb{S}^n_+}(X - \lambda\nabla f(X))$ at $\bar{X}$. Let us note that, in this situation, fast local convergence of Algorithm 2 can only be expected under the additional and restrictive assumption $\bar{X} \in \mathbb{S}^n_{++}$ (this is exactly condition (D.2) in Assumption 4.3.6). However, the results of this example can be immediately applied to a pure and local semismooth Newton method.

5.5. Extensions

In this section, we demonstrate the broad applicability and the advantages of the concept of decomposability and, based on our second order and nonsingularity results, we derive an analogue second order framework for general convex composite problems of the form

$$(\mathcal{P}_c) \qquad \min_{x\in\mathbb{R}^n} \ \psi_c(x) := f(x) + \phi(G(x)),$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^m$ are given, twice continuously differentiable functions and the mapping $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ is convex, proper, and lower semicontinuous, as usual. In the following, let $\bar{x} \in G^{-1}(\text{dom } \phi)$ be an arbitrary stationary point of the problem $(\mathcal{P}_c)$ and let us suppose that ϕ is C^2-fully decomposable at $G(\bar{x})$ with decomposition pair (ϕ_d, F). Furthermore, let us assume that the nondegeneracy condition

$$(5.5.1) \qquad DG(\bar{x})\mathbb{R}^n + \text{lin } N_{\partial\phi(G(\bar{x}))}(\lambda) = \mathbb{R}^m, \quad \lambda \in \partial\phi(G(\bar{x})),$$

holds at $\bar{x}$. The first order necessary optimality conditions associated with problem $(\mathcal{P}_c)$ can be characterized as follows

$$(5.5.2) \qquad \nabla f(\bar{x}) + DG(\bar{x})^\top\bar{\lambda} = 0, \quad \bar{\lambda} \in \partial\phi(G(\bar{x})).$$

In particular, the nondegeneracy condition implies that the set of Lagrange multipliers,

$$\mathcal{M}(\bar{x}) = \{\lambda \in \partial\phi(G(\bar{x})) : \nabla f(\bar{x}) + DG(\bar{x})^\top\lambda = 0\},$$

is nonempty and reduces to the singleton $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$. Moreover, by Lemma 5.3.23, we can infer that the composite function $\phi \circ G$ is C^2-fully decomposable at $\bar{x}$ with decomposition pair $(\phi_d, F \circ G)$.

Now, let $\Gamma \in \mathbb{S}^m_{++}$ be an arbitrary parameter matrix. Then, the KKT conditions (5.5.2) can be equivalently represented as the following system of equations

$$\Psi^\Gamma : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m, \quad \Psi^\Gamma(x, \lambda) := \begin{pmatrix} \nabla f(x) + DG(x)^\top \lambda \\ G(x) - \operatorname{prox}_\phi^\Gamma(G(x) + \Gamma^{-1}\lambda) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and thus, every KKT pair $(\bar{x}, \bar{\lambda})$ of the problem $(\mathcal{P}_c)$ corresponds to a solution of the latter system and vice versa. Therefore, we can again apply the semismooth Newton method to approximately solve the nonsmooth system of equations

$$\Psi^\Gamma(x, \lambda) = 0 \tag{5.5.3}$$

and to compute a stationary point of the convex composite problem $(\mathcal{P}_c)$. Let us emphasize that the mapping Ψ^Γ plays a similarly important role as the nonsmooth function F^Λ. Clearly, Ψ^Γ can be seen as an adequate generalization of F^Λ taking account of the more general and more difficult structure of the convex composite problem $(\mathcal{P}_c)$. Our goal is now to extend our second order results and to establish appropriate nonsingularity conditions for the generalized derivatives of the nonsmooth mapping Ψ^Γ.

As for the initial problem $(\mathcal{P})$, we first discuss and clarify the different notions of stationarity emerging from the decomposability of problem $(\mathcal{P}_c)$. The first order optimality conditions for the decomposed problem

$$\min_{x \in \mathbb{R}^n} \; f(x) + \phi_d((F \circ G)(x)) + \bar{c}, \quad \bar{c} = \phi(G(\bar{x})), \tag{5.5.4}$$

formally take the following form

$$\exists \; \bar{\mu} \in \mathbb{R}^p : \quad \nabla f(\bar{x}) + D(F \circ G)(\bar{x})^\top \bar{\mu} = 0, \quad \bar{\mu} \in \partial \phi_d(0).$$

(Here, we assume that $F : \mathbb{R}^m \to \mathbb{R}^p$ is a p-dimensional mapping). Due to the nondegeneracy condition and the full decomposability of $\phi \circ G$, the following constraint qualifications are satisfied w.r.t. $\bar{x}$

$$0 \in \operatorname{int}\{G(\bar{x}) + DG(\bar{x})\mathbb{R}^n - \operatorname{dom} \phi\}, \quad 0 \in \operatorname{int}\{F(G(\bar{x})) + D(F \circ G)(\bar{x})\mathbb{R}^n - \operatorname{dom} \phi_d\}.$$

Thus, setting $\mathcal{M}_c(\bar{x}) := \{\mu \in \partial\phi_d(0) : \nabla f(\bar{x}) + D(F \circ G)(\bar{x})^\top \mu = 0\}$, the discussion of the first order necessary condition on page 87f. implies

$$\mathcal{M}_c(\bar{x}) \neq \emptyset \quad \iff \quad (\psi_c)_-^\downarrow(\bar{x}; h) \geq 0, \quad \forall \; h \in \mathbb{R}^n \quad \iff \quad \mathcal{M}(\bar{x}) \neq \emptyset.$$

Consequently, $\bar{x}$ is also a stationary point of the decomposed problem (5.5.4) (in the sense of $\mathcal{M}_c(\bar{x}) \neq \emptyset$) and the set of the corresponding Lagrange multipliers necessarily has to reduce to a singleton $\mathcal{M}_c(\bar{x}) = \{\bar{\mu}\} \subset \mathbb{R}^p$. Moreover, as in Remark 5.3.3, the stability property of Robinson's constraint qualification guarantees that the latter equivalence does also hold for every stationary point in a certain neighborhood of $\bar{x}$. Hence, every isolated stationary point of problem $(\mathcal{P}_c)$ is also an isolated stationary point of the decomposed problem (5.5.4) and vice versa. In addition, if $\bar{x}$ is an isolated stationary point, then the uniqueness of the Lagrange multiplier $\bar{\lambda}$ implies that the pair $(\bar{x}, \bar{\lambda})$ is an isolated, local solution of the nonsmooth equation (5.5.3). On the other hand, if $(\bar{x}, \bar{\lambda})$ is an isolated solution of the system (5.5.3), then there exists $\varepsilon > 0$ such that

$$\mathcal{M}(x) = \emptyset, \quad \forall\, x \in B_\varepsilon(\bar{x}) \setminus \{\bar{x}\}, \quad \mathcal{M}(\bar{x}) \cap B_\varepsilon(\bar{\lambda}) = \{\bar{\lambda}\}, \tag{5.5.5}$$

i.e., $\bar{x}$ is an isolated stationary point of the convex composite problem $(\mathcal{P}_c)$. As a consequence and similar to the discussion of ordinary decomposable problems in section 5.3, it suffices to analyze second order conditions for the decomposed problem (5.5.4) and to pass the stationarity results to the initial convex composite problem $(\mathcal{P}_c)$. Fortunately, since the composite function $\varphi \circ G$ is C^2-fully decomposable at $\bar{x}$, we can reuse and apply our theory and results of section 5.3 to the decomposed problem (5.5.4).

Next, we derive an explicit connection between the Lagrange multipliers $\bar{\lambda}$ and $\bar{\mu}$. The stationarity of $\bar{x}$ implies that the function $\Pi(y) := \phi^{\downarrow}(G(\bar{x}); y)$ is proper and subdifferentiable at 0 and it holds $\partial\Pi(0) = \partial\phi(G(\bar{x}))$. Moreover, since Robinson's constraint qualification

$$0 \in \operatorname{int}\{F(G(\bar{x})) + DF(G(\bar{x}))\mathbb{R}^m - \operatorname{dom}\, \phi_d\} = \operatorname{int}\{DF(G(\bar{x}))\mathbb{R}^m - \operatorname{dom}\, \phi_d\}$$

holds at $G(\bar{x})$ (this follows from the full decomposability of ϕ), we have

$$\Pi(y) = (\phi_d \circ F)^{\downarrow}(G(\bar{x}); y) = \phi_d^{\downarrow}(F(G(\bar{x})); DF(G(\bar{x}))y) = \phi_d(DF(G(\bar{x}))y), \quad \forall\, y \in \mathbb{R}^m,$$

where we used Lemma 2.5.6 and (5.3.5). Applying Lemma 2.5.15, we can infer

$$\partial\Pi(0) = DF(G(\bar{x}))^\top \partial\phi_d(0)$$

and thus, $\partial\phi(G(\bar{x})) = DF(G(\bar{x}))^\top \partial\phi_d(0)$. Consequently, if $\bar{\mu} \in \mathcal{M}_c(\bar{x})$ is a Lagrange multiplier of the decomposed problem, then it follows

$$\nabla f(\bar{x}) - DG(\bar{x})^\top [DF(G(\bar{x}))^\top \bar{\mu}] = 0, \quad DF(G(\bar{x}))^\top \bar{\mu} \in \partial\phi(G(\bar{x})).$$

Hence, $DF(G(\bar{x}))^\top \bar{\mu}$ is also a Lagrange multiplier of the initial problem $(\mathcal{P}_c)$ and the uniqueness of $\bar{\lambda}$ implies

$$\bar{\lambda} = DF(G(\bar{x}))^\top \bar{\mu}.$$

In the following, based on Theorem 5.3.6, we formulate second order necessary and sufficient conditions for the convex composite problem $(\mathcal{P}_c)$. Let us recall that the critical cone $\mathcal{C}(\bar{x})$ associated with problem $(\mathcal{P}_c)$ is given by

$$\begin{aligned} \mathcal{C}(\bar{x}) &= \{h \in \mathbb{R}^n : DG(\bar{x})h \in N_{\partial\phi(G(\bar{x}))}(\bar{\lambda})\} \\ &= \{h \in \mathbb{R}^n : \nabla f(\bar{x})^\top h + \phi^{\downarrow}(G(\bar{x}); DG(\bar{x})h) = 0\}. \end{aligned}$$

Furthermore, we will also need the *Lagrange function*

$$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, \quad (x, \lambda) \mapsto \mathcal{L}(x, \lambda) := f(x) + \langle \lambda, G(x) \rangle.$$

Theorem 5.5.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable and let $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping.*

Let $\bar{x} \in G^{-1}(\mathrm{dom}\ \phi)$ be given and assume that ϕ is C^2-fully decomposable at $G(\bar{x})$ with decomposition pair (ϕ_d, F). Moreover, suppose that the nondegeneracy condition

$$DG(\bar{x})\mathbb{R}^n + \mathrm{lin}\ N_{\partial\phi(G(\bar{x}))}(\lambda) = \mathbb{R}^m, \quad \lambda \in \partial\phi(G(\bar{x}))$$

is satisfied at $\bar{x}$. Then, the following statements hold:

(i) (Second order necessary conditions). *Suppose that $\bar{x}$ is a locally optimal solution of problem $(\mathcal{P}_c)$. Then, for every $h \in \mathcal{C}(\bar{x})$ the following inequality is satisfied*

$$h^\top \nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda})h + \langle \bar{\mu}, D^2F(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle \geq 0.$$

Here, $\bar{\lambda} \in \mathcal{M}(\bar{x})$ and $\bar{\mu} \in \mathcal{M}_c(\bar{x})$ are the unique Lagrange multipliers associated with the problems $(\mathcal{P}_c)$ and (5.5.4), respectively.

(ii) (Second order sufficient conditions). *Let $\bar{x}$ be a stationary point of the initial minimization problem $(\mathcal{P}_c)$. Then, the quadratic growth condition,*

$$f(x) + \phi(G(x)) \geq f(\bar{x}) + \phi(G(\bar{x})) + \alpha\|x - \bar{x}\|^2, \tag{5.5.6}$$

holds for some $\alpha > 0$ and all x in a neighborhood of $\bar{x}$ if and only if the following second order sufficient condition is satisfied,

$$h^\top \nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda})h + \langle \bar{\mu}, D^2F(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle > 0, \quad \forall\ h \in \mathcal{C}(\bar{x}) \setminus \{0\}. \tag{5.5.7}$$

In the latter case, $\bar{x}$ is a (strict) locally optimal solution and an isolated stationary point of problem $(\mathcal{P}_c)$.

Proof. Due to the full decomposability of $\phi \circ G$ and

$$\begin{aligned}\langle \bar{\mu}, D^2(F \circ G)(\bar{x})[h, h]\rangle &= \langle \bar{\mu}, DF(G(\bar{x}))D^2G(\bar{x})[h, h] + D^2F(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle \\ &= \langle \bar{\lambda}, D^2G(\bar{x})[h, h]\rangle + \langle \bar{\mu}, D^2F(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle,\end{aligned}$$

Theorem 5.5.1 simply follows from Theorem 5.3.6 and Remark 5.3.8. □

As in last section, we will now show that the term

$$\langle \bar{\mu}, D^2F(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle$$

exactly represents the possible curvature of the nonsmooth function ϕ. Accordingly, for some arbitrary parameter matrix $\Gamma \in \mathbb{S}^m_{++}$ and by using the strict complementarity condition, we will also derive an additional, alternative characterization in terms of the Fréchet derivative of the proximity operator $\mathrm{prox}^\Gamma_\phi$ at $\bar{u} := G(\bar{x}) + \Gamma^{-1}\bar{\lambda}$ that is independent of the specific decomposition pair (ϕ_d, F).

In this paragraph, we mainly recreate the argumentation in Lemma 5.3.9 for the more general convex composite setting. Therefore, let us define $\mathcal{Y} := DG(\bar{x})\mathcal{C}(\bar{x}) + \mathrm{lin}\ N_{\partial\phi(G(\bar{x}))}(\bar{\lambda})$ and let $y \in \mathcal{Y}$ be arbitrary. Then, there exist $h \in \mathcal{C}(\bar{x})$ and $d \in \mathrm{lin}\ N_{\partial\phi(G(\bar{x}))}(\bar{\lambda})$ such that

$y = DG(\bar{x})h + d$ and we have

$$\begin{aligned}(5.5.8)\qquad \Pi(y) = \phi^{\downarrow}(G(\bar{x}); DG(\bar{x})h + d) &= \sup_{\lambda \in \partial\phi(G(\bar{x}))} \langle \lambda, DG(\bar{x})h + d\rangle \\ &= \phi^{\downarrow}(G(\bar{x}); DG(\bar{x})h) + \langle \bar{\lambda}, d\rangle = -\langle \nabla f(\bar{x}), h\rangle + \langle \bar{\lambda}, d\rangle = \langle \bar{\lambda}, y\rangle,\end{aligned}$$

where we used $h \in \mathcal{C}(\bar{x})$, and the stationarity of $\bar{x}$. (Let us recall, that the stationarity of $\bar{x}$ implies that ϕ is subdifferentiable at $G(\bar{x})$). On the other hand, due to

$$\Pi(y) = \phi_d(DF(G(\bar{x}))y),$$

it follows $DF(G(\bar{x}))\mathcal{Y} \subset \operatorname{dom} \phi_d$. Consequently, by Lemma 5.3.5, the mapping ϕ is twice directionally epidifferentiable at $G(\bar{x})$ on $\mathcal{Y}$ and, using Lemma 5.2.5 and (5.3.6), we obtain

$$\begin{aligned}\phi^{\downarrow\downarrow}(G(\bar{x}); y, w) &= \phi_d^{\downarrow\downarrow}(F(G(\bar{x})); DF(G(\bar{x}))y, DF(G(\bar{x}))w + D^2F(G(\bar{x}))[y,y]) \\ &= \phi_d^{\downarrow}(DF(G(\bar{x}))y; DF(G(\bar{x}))w + D^2F(G(\bar{x}))[y,y])\end{aligned}$$

for all $y \in \mathcal{Y}$. Now, let us fix an arbitrary element $y \in \mathcal{Y}$. Then, setting $\bar{y} := G(\bar{x})$ and $\bar{w} := D^2F(\bar{y})[y,y]$, it holds

$$\begin{aligned}(5.5.9)\qquad -\xi^*_{\phi,y}(\bar{\lambda}) &= \inf_{w\in\mathbb{R}^m} \langle -\bar{\lambda}, w\rangle + \phi^{\downarrow\downarrow}(\bar{y}; y, w) \\ &= \inf_{w\in\mathbb{R}^m} \langle -\bar{\lambda}, w\rangle + \phi_d^{\downarrow}(DF(\bar{y})y, DF(\bar{y})w + \bar{w}),\end{aligned}$$

where $\xi_{\phi,y}(\cdot) := \phi^{\downarrow\downarrow}(\bar{y}; y, \cdot)$. Next, as in Remark 5.2.9, we have

$$\begin{aligned}\phi_d^{\downarrow}(DF(\bar{y})y; \omega) &= \phi_d^{\downarrow\downarrow}(F(\bar{y}); DF(\bar{y})y, \omega) \\ &\geq \liminf_{t\downarrow 0,\, \tilde{\omega}\to\omega} \frac{\langle \bar{\mu}, tDF(\bar{y})y + \frac{1}{2}t^2\tilde{\omega}\rangle - t\cdot \phi_d(DF(\bar{y})y)}{\frac{1}{2}t^2} = \langle \bar{\mu}, \omega\rangle,\end{aligned}$$

where we used $\bar{\mu} \in \partial\phi_d(0)$, $DF(\bar{y})^\top\bar{\mu} = \bar{\lambda}$, and (5.5.8). Together with (5.3.8), this implies

$$\phi_d^{\downarrow}(DF(\bar{y})y; 0) = \phi_d^{\downarrow\downarrow}(F(\bar{y}); DF(\bar{y})y, 0) = 0$$

and thus, the mapping $\Upsilon : \mathbb{R}^p \to (-\infty, +\infty]$, $\Upsilon(\omega) := \phi_d^{\downarrow}(DF(\bar{y})y; \omega + \bar{w})$ is convex, proper, and lower semicontinuous. Hence, the Fenchel-Rockafellar duality framework can again be applied to dualize the problem (5.5.9). Specifically, setting $\varrho(\omega) := \langle -\bar{\lambda}, \omega\rangle$, the dual problem is formally given by

$$\max_v \ \varrho^*(DF(\bar{y})^\top v) - \Upsilon^*(-v)$$

and by repeating the computations in Lemma 5.3.9, we obtain the following representation

$$\max_v \ \langle -v, \bar{w}\rangle \quad \text{s.t.} \quad \begin{cases} -v \in \partial\phi_d(0), \\ \bar{\lambda} + DF(\bar{y})^\top v = 0, \\ \langle v, DF(\bar{y})y\rangle + \phi_d(DF(\bar{y})y) = 0. \end{cases}$$

Since the third condition is automatically satisfied for all $y \in \mathcal{Y}$ and since it holds

$$\{\bar{\mu}\} \subseteq \{v \in \partial\phi_d(0) : DF(\bar{y})^\top v = \bar{\lambda}\} \subseteq \mathcal{M}_c(\bar{x}) = \{\bar{\mu}\},$$

we finally deduce

$$\max_v \ \varrho^*(DF(\bar{y})^\top v) - \Upsilon^*(-v) \ = \ \langle \bar{\mu}, D^2F(\bar{y})[y,y]\rangle. \tag{5.5.10}$$

Now, as in Lemma 5.3.9 and using the full decomposability of ϕ, it can be shown that there is no gap between the primal problem (5.5.9) and the dual problem (5.5.10). Moreover, the infimum that defines the curvature term $-\xi^*_{\phi,y}(\bar{\lambda})$ is attained at some $\hat{w} \in \mathbb{R}^m$, i.e., it holds

$$-\xi^*_{\phi,y}(\bar{\lambda}) = \phi^{\downarrow\downarrow}(\bar{y}; y, \hat{w}) - \langle \bar{\lambda}, \hat{w}\rangle.$$

Next, as in Lemma 5.3.27, using $\bar{\mu} \in \partial\phi_d(0)$, $\bar{\lambda} = DF(\bar{y})^\top \bar{\mu}$, and a second order Taylor expansion of F at $\bar{y}$, we establish the following lower bound for the second order difference quotient

$$\begin{aligned}\Delta^2_{t_k}\,\phi(\bar{y}|\bar{\lambda})(y^k) &= \frac{\phi_d(F(\bar{y}+t_k y^k)) - \phi_d(F(\bar{y})) - t_k \cdot \langle \bar{\lambda}, y^k\rangle}{\frac{1}{2}t_k^2} \\ &\geq \frac{\langle \bar{\mu}, F(\bar{y}+t_k y^k)\rangle - t_k \cdot \langle \bar{\mu}, DF(\bar{y})y^k\rangle}{\frac{1}{2}t_k^2} = \langle \bar{\mu}, D^2F(\bar{y})[y^k, y^k]\rangle + o(1),\end{aligned}$$

where $(t_k)_k$, $t_k \downarrow 0$, and $(y^k)_k$, $y^k \to y \in \mathcal{Y}$, are arbitrarily chosen. Furthermore, let us note that the nondegeneracy condition (5.5.1) and (5.5.8) imply

$$\mathcal{Y} = \{y \in \mathbb{R}^m : \phi^\downarrow(G(\bar{x}); y) = \langle \bar{\lambda}, y\rangle\} = N_{\partial\phi(G(\bar{x}))}(\bar{\lambda}).$$

Thus, by combining the latter facts and reconsidering the proof of Lemma 5.3.27, it can be readily shown that ϕ is twice epi-subdifferentiable at $\bar{y} = G(\bar{x})$ relative to $\bar{\lambda}$ and it follows

$$\mathrm{d}^2\phi(G(\bar{x})|\bar{\lambda})(y) = -\xi^*_{\phi,y}(\bar{\lambda}) = \langle \bar{\mu}, D^2F(G(\bar{x}))[y,y]\rangle$$

for all $y \in \mathcal{Y}$. In particular, this also implies

$$\mathrm{d}^2\phi(G(\bar{x})|\bar{\lambda})(DG(\bar{x})h) = \langle \bar{\mu}, D^2F(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle, \quad \forall\ h \in \mathcal{C}(\bar{x}).$$

Now, let us briefly discuss an appropriate adaption of the $\mathcal{VU}$-theory. Again, let $\Gamma \in \mathbb{S}^m_{++}$ be an arbitrary parameter matrix. Then, by using $\bar{u} = G(\bar{x}) + \Gamma^{-1}\bar{\lambda}$ and the stationarity of $\bar{x}$, we have

$$\bar{p} := \mathrm{prox}^\Gamma_\phi(\bar{u}) = G(\bar{x}), \quad \bar{g} := \nabla\mathrm{env}^\Gamma_\phi(\bar{u}) = \Gamma(G(\bar{x}) + \Gamma^{-1}\bar{\lambda} - \mathrm{prox}^\Gamma_\phi(\bar{u})) = \bar{\lambda}.$$

Hence, in this case, the strict complementarity condition apparently reduces to

$$\bar{\lambda} \in \mathrm{ri}\ \partial\phi(G(\bar{x})) \tag{5.5.11}$$

and the $\mathcal{U}$- and $\mathcal{V}$-space can be defined via

$$\mathcal{U} := \operatorname{lin} N_{\partial\phi(G(\bar{x}))}(\bar{\lambda}) = \operatorname{lin} \mathcal{Y}, \quad \mathcal{V} := [\mathcal{U}]^{\perp,\Gamma}.$$

(If the strict complementarity condition (5.5.11) is satisfied, then the lineality space operation in the definition of the subspace $\mathcal{U}$ is superfluous). Thus, at this point, it is clear that the $\mathcal{VU}$-concept and our theoretical results, which were presented and discussed in the last section, can be naturally extended and applied to the convex composite problem $(\mathcal{P}_c)$. Specifically, the associated quadratic sub-Lagrangian takes the following form

$$\Phi_\Gamma : \mathcal{U} \to [-\infty, +\infty], \quad \Phi_\Gamma(u) := \inf_{v \in \mathcal{V}} \phi(G(\bar{x}) + u \oplus_\Gamma v) - \langle \hat{\lambda}_{\mathcal{V}}, v \rangle_{\mathcal{V},\Gamma} + \frac{1}{2}\|v\|^2_{\mathcal{V},\Gamma},$$

where $\hat{\lambda} := \Gamma^{-1}\bar{\lambda} = \Gamma^{-1}\bar{g}$. Moreover, let us suppose that $\bar{U}$ and $\bar{V}$ are two basis matrices whose columns span the subspaces $\mathcal{U}$ and $\mathcal{V}$, respectively. If the strict complementarity condition (5.5.11) is fulfilled, Lemma 5.3.38 implies that the quadratic sub-Lagrangian Φ_Γ is twice epi-subdifferentiable at 0, relative to $g_0 := \nabla\Phi_\Gamma(0) = \bar{U}^\top\bar{\lambda}$ and it follows

$$\mathrm{d}^2\Phi_\Gamma(0|g_0)(y_{\mathcal{U}}) = \mathrm{d}^2\phi(G(\bar{x})|\bar{\lambda})(y) = y^\top \mathcal{H}_\phi(G(\bar{x}))y, \quad \forall\ y = y_{\mathcal{U}} \oplus_\Gamma 0 \in \mathcal{U} \equiv \mathcal{Y},$$

where the symmetric, positive semidefinite matrix $\mathcal{H}_\phi(G(\bar{x})) \in \mathbb{R}^{m\times m}$ is given by

$$\mathcal{H}_\phi(G(\bar{x})) := \sum_{i=1}^{p} \bar{\mu}_i \nabla^2 F_i(G(\bar{x})).$$

Hence, as in Theorem 5.3.39, we can infer that Φ_Γ has a generalized Hessian at 0 and it holds $H\Phi_\Gamma(0) = \bar{U}^\top\mathcal{H}_\phi(G(\bar{x}))\bar{U}$. Combining our latter calculations, Theorem 5.3.39, the proximal $\mathcal{VU}$-calculus on page 159f., and using

$$y := DG(\bar{x})h = [DG(\bar{x})h]_{\mathcal{U}} \oplus_\Gamma 0 \in \mathcal{U} \equiv \mathcal{Y}, \quad \forall\ h \in \mathcal{C}(\bar{x}),$$

we finally obtain

$$\begin{aligned} -\xi^*_{\phi,h}(\bar{\lambda}) = \langle DG(\bar{x})h, \mathcal{H}_\phi(G(\bar{x}))DG(\bar{x})h\rangle &= \langle y_{\mathcal{U}}, [\bar{U}^\top\mathcal{H}_\phi(G(\bar{x}))\bar{U}]y_{\mathcal{U}}\rangle_{\mathcal{U}} \\ &= \langle DG(\bar{x})h, [\Gamma^{-\frac{1}{2}}\mathcal{Q}^\Gamma_\phi(\bar{u})^+\Gamma^{\frac{1}{2}} - I]DG(\bar{x})h\rangle_\Gamma, \end{aligned}$$

where $\mathcal{Q}^\Gamma_\phi(\bar{u}) := \Gamma^{\frac{1}{2}}D\mathrm{prox}^\Gamma_\phi(\bar{u})\Gamma^{-\frac{1}{2}}$ and $\xi_{\phi,h}(\cdot) := \phi^{\downarrow\downarrow}(G(\bar{x}); DG(\bar{x})h, \cdot)$.

Let us note that although the set of "critical directions" $DG(\bar{x})^{-1}\mathcal{C}(\bar{x})$ is a more intuitive and plausible choice for the $\mathcal{U}$-space, it cannot be used here. In particular, in this case, the second order subderivative $\mathrm{d}^2\Phi_\Gamma(0|g_0)$ is not necessarily finite on the whole subspace $\mathcal{U}$ and we cannot apply Theorem 5.3.39 and the $\mathcal{VU}$-calculus. Thus, our duality argument in (5.5.9) and (5.5.10) is formulated for directions of the more complex set $\mathcal{Y}$ from the start.

We now present our extended invertibility result for convex composite problems.

Theorem 5.5.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^m$ be twice continuously differentiable and let $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous mapping. Furthermore,* *let $\bar{x} \in G^{-1}(\text{dom } \phi)$ be a stationary point of problem $(\mathcal{P}_c)$ and suppose that ϕ is C^2-fully decomposable at $G(\bar{x})$ and that the nondegeneracy condition,*

$$DG(\bar{x})\mathbb{R}^n + \text{lin } N_{\partial\phi(G(\bar{x}))}(\lambda) = \mathbb{R}^m, \quad \lambda \in \partial\phi(G(\bar{x})),$$

is satisfied at $\bar{x}$. Additionally, let $\bar{\lambda}$ denote the unique Lagrange multiplier associated with the stationary point $\bar{x}$ and assume that the strict complementarity condition

$$\bar{\lambda} \in \text{ri } \partial\phi(G(\bar{x}))$$

holds at $\bar{x}$. Then, for every $\Gamma \in \mathbb{S}^m_{++}$, the proximity operator prox^Γ_ϕ is Fréchet differentiable at $\bar{u} := G(\bar{x}) + \Gamma^{-1}\bar{\lambda}$ and, setting $\mathcal{Q}^\Gamma_\phi(\bar{u}) := \Gamma^{\frac{1}{2}} D\text{prox}^\Gamma_\phi(\bar{u})\Gamma^{-\frac{1}{2}}$, the second order conditions

$$h^\top \nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda})h + \langle DG(\bar{x})h, [\Gamma^{-\frac{1}{2}}\mathcal{Q}^\Gamma_\phi(\bar{u})^+\Gamma^{\frac{1}{2}} - I]DG(\bar{x})h\rangle_\Gamma > 0, \quad \forall\ h \in \mathcal{C}(\bar{x}) \setminus \{0\},$$

are sufficient and necessary for the quadratic growth condition (5.5.6). In particular, if the latter second order sufficient condition is satisfied, then the following implications do hold:

(i) *The point $\bar{x}$ is a strict local minimum and an isolated stationary point of the problem $(\mathcal{P}_c)$. Moreover, the pair $(\bar{x}, \bar{\lambda})$ is also an isolated local solution of the equation*

$$\Psi^\Gamma(x, \lambda) = 0.$$

(ii) *If the proximity operator prox^Γ_ϕ is semismooth at $\bar{u}$, then the mapping Ψ^Γ is strictly differentiable at $(\bar{x}, \bar{\lambda})$ and its Fréchet derivative $D\Psi^\Gamma(\bar{x}, \bar{\lambda})$ is nonsingular.*

Proof. By Lemma 5.3.32, the proximity operator prox^Γ_ϕ is Fréchet differentiable at $\bar{u} = G(\bar{x}) + \Gamma^{-1}\bar{\lambda}$ if and only if the strict complementarity condition

$$\bar{\lambda} = \nabla\text{env}^\Gamma_\phi(\bar{u}) \in \text{ri } \partial\phi(G(\bar{x}))$$

is satisfied. Now, let (ϕ_d, F) be a corresponding decomposition pair of the mapping ϕ. Then, our preceding discussion establishes

$$\langle DG(\bar{x})h, [\Gamma^{-\frac{1}{2}}\mathcal{Q}^\Gamma_\phi(\bar{u})^+\Gamma^{\frac{1}{2}} - I]DG(\bar{x})h\rangle_\Gamma = \langle \bar{\mu}, D^2F(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle$$

for all $h \in \mathcal{C}(\bar{x})$. (As usual, $\bar{\mu}$ denotes the unique Lagrange multiplier of the decomposed problem (5.5.4)). Consequently, Theorem 5.5.1 implies that $\bar{x}$ is a (strict) local minimum and an isolated stationary point of problem $(\mathcal{P}_c)$. Moreover, using the uniqueness of $\bar{\lambda}$ and as we have already shown, the pair $(\bar{x}, \bar{\lambda})$ is also an isolated solution of the nonsmooth system of equations

$$\Psi^\Gamma(x, \lambda) = 0.$$

As in Theorem 5.4.4, the semismoothness and Fréchet differentiability of the proximity operator prox^Γ_ϕ establish strict differentiability of the KKT mapping Ψ^Γ at $(\bar{x}, \bar{\lambda})$.

The rest of the proof is strongly motivated by a general nonsingularity result of Sun [226, Proposition 3.2] for nonlinear semidefinite programming. Furthermore, we will also follow the proof of Theorem 5.4.4. The derivative of Ψ^Γ at $(\bar{x}, \bar{\lambda})$ is given by

$$D\Psi^\Gamma(\bar{x}, \bar{\lambda}) = \begin{pmatrix} \nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda}) & DG(\bar{x})^\top \\ (I - V)DG(\bar{x}) & -V\Gamma^{-1} \end{pmatrix}, \quad V := D\mathrm{prox}^\Gamma_\phi(\bar{u}).$$

Now, suppose that there exists $0 \neq h = (h^x, h^\lambda) \in \mathbb{R}^{n+m}$ such that $D\Psi^\Gamma(\bar{x}, \bar{\lambda})h = 0$. Clearly, this implies

$$\nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda})h^x + DG(\bar{x})^\top h^\lambda = 0, \quad (I - V)DG(\bar{x})h^x - V\Gamma^{-1}h^\lambda = 0 \tag{5.5.12}$$

and it follows

$$DG(\bar{x})h^x = V(DG(\bar{x})h^x + \Gamma^{-1}h^\lambda). \tag{5.5.13}$$

Hence, applying Lemma 3.3.6, we deduce $DG(\bar{x})h^x \in N_{\partial\phi(G(\bar{x}))}(\bar{\lambda}) \equiv \mathcal{U}$, i.e., it holds $h^x \in \mathcal{C}(\bar{x})$. Moreover, since the matrix $\mathcal{Q}^\Gamma_\phi(\bar{u})$ obviously has the same structural properties as the matrix $\mathcal{Q}^\Lambda_\varphi(\bar{u})$ from the last section, we also have

$$\mathcal{Q}^\Gamma_\phi(\bar{u})^+\mathcal{Q}^\Gamma_\phi(\bar{u})\Gamma^{-\frac{1}{2}}h^\lambda = [\mathcal{Q}^\Gamma_\phi(\bar{u})^+ - I]\Gamma^{\frac{1}{2}}DG(\bar{x})h^x \tag{5.5.14}$$

and we obtain

$$\begin{aligned} &\langle h^x, \nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda})h^x + DG(\bar{x})^\top h^\lambda\rangle \\ &\qquad = (h^x)^\top\nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda})h^x + \langle V(DG(\bar{x})h^x + \Gamma^{-1}h^\lambda), h^\lambda\rangle \\ &\qquad = (h^x)^\top\nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda})h^x + \langle\Gamma^{\frac{1}{2}}DG(\bar{x})h^x + \Gamma^{\frac{1}{2}}h^\lambda, \mathcal{Q}^\Gamma_\phi(\bar{u})\Gamma^{-\frac{1}{2}}h^\lambda\rangle \\ &\qquad = (h^x)^\top\nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda})h^x + \langle\mathcal{Q}^\Gamma_\phi(\bar{u})[\Gamma^{\frac{1}{2}}DG(\bar{x})h^x + \Gamma^{\frac{1}{2}}h^\lambda], [\mathcal{Q}^\Gamma_\phi(\bar{u})^+ - I]\Gamma^{\frac{1}{2}}DG(\bar{x})h^x\rangle \\ &\qquad = (h^x)^\top\nabla^2_{xx}\mathcal{L}(\bar{x}, \bar{\lambda})h^x + \langle DG(\bar{x})h^x, [\Gamma^{-\frac{1}{2}}\mathcal{Q}^\Gamma_\phi(\bar{u})^+\Gamma^{\frac{1}{2}} - I]DG(\bar{x})h^x\rangle_\Gamma, \end{aligned}$$

where we used (5.5.13), (5.5.14), the basic identities of the Moore-Penrose inverse, and the symmetry of $\mathcal{Q}^\Gamma_\phi(\bar{u})$. Thus, the second order sufficient conditions imply $h^x = 0$ and from (5.5.12), it follows

$$DG(\bar{x})^\top h^\lambda = 0, \quad V\Gamma^{-1}h^\lambda = 0.$$

Next, by applying the nondegeneracy condition, we can infer that there exist $u \in \mathbb{R}^n$ and $v \in N_{\partial\phi(G(\bar{x}))}(\bar{\lambda})$ such that $h^\lambda = DG(\bar{x})u + v$. Furthermore, due to $v \in N_{\partial\phi(G(\bar{x}))}(\bar{\lambda}) \equiv \mathcal{U}$, it holds

$$\mathcal{Q}^\Gamma_\phi(\bar{u})^+\mathcal{Q}^\Gamma_\phi(\bar{u})\Gamma^{\frac{1}{2}}v = \Gamma^{\frac{1}{2}}v$$

and consequently, we finally get

$$\langle h^\lambda, h^\lambda\rangle = \langle h^\lambda, DG(\bar{x})u + v\rangle = \langle\Gamma^{-\frac{1}{2}}h^\lambda, \Gamma^{\frac{1}{2}}v\rangle = \langle\mathcal{Q}^\Gamma_\phi(\bar{u})^+\Gamma^{\frac{1}{2}}V\Gamma^{-1}h^\lambda, \Gamma^{\frac{1}{2}}v\rangle = 0.$$

Altogether, this implies $h = 0$, which contradicts our assumption. Hence, the Fréchet derivative $D\Psi^\Gamma(\bar{x}, \bar{\lambda})$ must be nonsingular, as desired. □

Of course, if the proximity operator $\mathrm{prox}^\Gamma_\phi$ is directionally differentiable in a certain neighborhood of $\bar{u} = G(\bar{x}) + \Gamma^{-1}\bar{\lambda}$, Theorem 2.5.21 (ii) and Theorem 5.5.2 immediately yield that every sequence of iterates $(x^k, \lambda^k)_k$ generated by the semismooth Newton method

$$\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x^k,\lambda^k) & DG(x^k)^\top \\ (I-V_k)DG(x^k) & -V_k\Gamma^{-1} \end{pmatrix} \begin{pmatrix} s^x \\ s^\lambda \end{pmatrix} = -\Psi^\Gamma(x^k,\lambda^k), \quad \begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} + \begin{pmatrix} s^x \\ s^\lambda \end{pmatrix},$$

$V_k \in \partial\text{prox}^\Gamma_\phi(u^k)$, $u^k := G(x^k) + \Gamma^{-1}\lambda^k$, converges q-superlinearly to the local and isolated solution $(\bar{x}, \bar{\lambda})$ if the respective initial point $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^m$ is chosen sufficiently close to $(\bar{x}, \bar{\lambda})$. We conclude this chapter with some final observations and remarks on future research directions.

Further extensions, full equivalence, and remarks

At first, let us mention that it is straightforward to extend our results to equality constrained problems of the form

$$\min_{x\in\mathbb{R}^n} \ f(x) + \phi(G(x)) \quad \text{s.t.} \quad h(x) = 0,$$

where the additional function $h : \mathbb{R}^n \to \mathbb{R}^p$ is supposed to be twice continuously differentiable, as usual. More specifically, if the extended nondegeneracy condition

$$\begin{pmatrix} \nabla h(\bar{x})^\top \\ DG(\bar{x}) \end{pmatrix} \mathbb{R}^n + \begin{pmatrix} \{0\} \\ \text{lin } N_{\partial\phi(G(\bar{x}))}(\lambda) \end{pmatrix} = \begin{pmatrix} \mathbb{R}^p \\ \mathbb{R}^m \end{pmatrix}, \quad \lambda \in \partial\phi(G(\bar{x})),$$

is satisfied, then, by considering the proof of Lemma 5.3.23, it can be readily shown that the function

$$\varrho : \mathbb{R}^n \to (-\infty, +\infty], \quad \varrho(x) = \phi(G(x)) + \iota_{\{0\}}(h(x))$$

is C^2-fully decomposable at $\bar{x}$. Furthermore, in this case, the corresponding KKT-mapping Ψ^Γ takes the following form

$$\Psi^\Gamma(x,\lambda,\mu) := \begin{pmatrix} \nabla f(x) + DG(x)^\top\lambda + \nabla h(x)\mu \\ G(x) - \text{prox}^\Gamma_\phi(G(x) + \Gamma^{-1}\lambda) \\ h(x) \end{pmatrix}$$

and a similar invertibility result as in Theorem 5.5.2 can be established. However, a detailed discussion of this problem is beyond the scope of this thesis.

Another possible extension arises from the quite apparent, but interesting question whether the assertion in Theorem 5.5.2 is also true for the opposite direction. In particular, let us suppose that the strict complementarity condition holds at $\bar{x}$, the proximity operator prox^Γ_ϕ is semismooth at $\bar{u} = G(\bar{x}) + \Gamma^{-1}\bar{\lambda}$, and that the Fréchet derivative $D\Psi^\Gamma(\bar{x}, \bar{\lambda})$ is nonsingular. Then, can it be shown that the nondegeneracy condition (5.5.1) and the second order sufficient condition in Theorem 5.5.2 are satisfied at $\bar{x}$?

In fact, by using the concept of *strongly regular solutions* of generalized equations and the so-called *uniform quadratic growth condition*, the answer to this question seems to be affirmative if the stationary point $\bar{x}$ is additionally assumed to be a local solution of the optimization problem $(\mathcal{P}_c)$. In this situation, we can argue as follows:

- The invertibility of $D\Psi^\Gamma(\bar{x},\bar{\lambda})$ implies that the KKT-pair $(\bar{x},\bar{\lambda})$ is an isolated local solution of the system of equations

$$\Psi^\Gamma(x,\lambda) = 0$$

  and by (5.5.5), this shows that $\bar{\lambda}$ is a locally unique Lagrange multiplier. Moreover, using the convexity of the set $\mathcal{M}(\bar{x})$, this already establishes global uniqueness of the multiplier $\bar{\lambda}$, i.e., we have $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$. Thus, by Lemma 5.1.11 (ii), the nondegeneracy condition must hold at $\bar{x}$.

- Applying Clarke's inverse function theorem [49, Theorem 1], it can be shown that the mapping Ψ^Γ is a locally Lipschitz homeomorphism near $(\bar{x},\bar{\lambda})$, i.e., there exists a neighborhood $\mathcal{O}$ of $(\bar{x},\bar{\lambda})$ such that the restricted function $\Psi^\Gamma|_{\mathcal{O}} : \mathcal{O} \to \Psi^\Gamma(\mathcal{O})$ is Lipschitz continuous and *bijective* on $\mathcal{O}$ and its inverse mapping is also Lipschitz continuous. Then, as in [226, Remark 3.1], it follows that the pair $(\bar{x},\bar{\lambda})$ is a strongly regular solution of the generalized equation

$$0 \in \begin{pmatrix} \nabla_x \mathcal{L}(x,\lambda) \\ -G(x) \end{pmatrix} + \begin{pmatrix} \{0\} \\ \partial\phi^*(\lambda) \end{pmatrix}.$$

  Now, if $\bar{x}$ is a locally optimal solution of problem $(\mathcal{P}_c)$, Theorem 5.20 in [27] is applicable and we can infer that the uniform quadratic growth condition is satisfied. Thus, since the uniform quadratic growth condition implies the classical quadratic growth condition, Theorem 5.5.1 shows that the second order sufficient condition (5.5.7) must hold at $\bar{x}$.

- Additionally, by mimicking the proof of [217, Theorem 5.2] it can be shown that, under the nondegeneracy and the strict complementarity condition, the uniform quadratic growth and the second order sufficient condition are actually equivalent. Moreover, if ϕ is decomposable at $G(\bar{x})$ with decomposition pair (ϕ_d, F) such that $DF(G(\bar{x}))$ is onto, then [27, Theorem 5.20 and 5.24] imply that $(\bar{x},\bar{\lambda})$ is a strongly regular solution of the above generalized equation and Ψ^Γ is a locally Lipschitz homeomorphism. A well-known inverse function theorem by Kummer [128] now yields invertibility of the derivative $D\Psi^\Gamma(\bar{x},\bar{\lambda})$.

Of course, our argumentation is only preliminary and the latter steps have to be verified more carefully. However, this brief discussion clearly demonstrates the deep connection between these different concepts. For more details on strong regularity and the uniform second order growth condition, we refer to Robinson [202], Sun [226], and the sections 5.1.3–5.1.5 in [27].

Let us note that this adumbrated equivalence has already been investigated in a much broader and more general context. In particular, in [202] Robinson analyzed connections between the strong regularity of KKT points, nonsingularity conditions, and a *second order strong sufficient condition* for nonlinear programming. Similar results were obtained by Bonnans, Ramírez [25] and Wang, Zhang [250] for second order cone programming, and by Sun et al. [226, 43] and Ding [63] for nonlinear semidefinite programs and other matrix optimization problems. For these specific problems, explicit representations and formulae for the corresponding Clarke subdifferentials, the critical cones, and the curvature terms are available and it is possible to derive full equivalence between the mentioned concepts and the nonsingularity of all elements in $\partial\Psi^\Gamma(\bar{x},\bar{\lambda})$ *without* the *strict complementarity condition.* Moreover, since these problems all have a fully decomposable structure, the mentioned strong second order sufficient condition can formally be represented in the following form

$$h^\top \nabla^2_{xx}\mathcal{L}(\bar{x},\bar{\lambda})h + \langle \bar{\mu}, D^2F(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle > 0, \quad \forall\, h \in \operatorname{aff}\, \mathcal{C}(\bar{x}) \setminus \{0\}.$$

Thus, in comparison to the original second order condition (5.5.7) and in order to cope with the missing strict complementarity, the larger set aff $\mathcal{C}(\bar{x})$ has to be used and the curvature term $-\xi_{\phi,h}(\bar{\lambda})$ is "substituted" by the quadratic form

$$h \mapsto \langle \bar{\mu}, D^2F(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle,$$

which is also well-defined for $h \notin \mathcal{C}(\bar{x})$. Let us again emphasize that full equivalence results are only available for very specific cone reducible problems. To the best of our knowledge, general results are not yet known.

Finally, strong regularity has also been analyzed in a second order variational context. In particular, by using new computational results for the coderivative mapping, Outrata and Ramírez [180] and Mordukhovich et al. [159] derived a variational-based connection between strong regularity and *strong stability* concepts and the so-called *Aubin property* of the associated critical point mapping for second order cone programs. Again, more general connections and results which yield an equivalent (variational-based) characterization of the strong second order sufficient condition and the invertibility of all elements in Clarke's subdifferential $\partial\Psi^\Gamma(\bar{x},\bar{\lambda})$ are not yet available.

6. Numerical methods for generalized variational inequalities

In this chapter, we consider and investigate numerical methods for *generalized variational inequality problems* (in short GVIPs) of the following form:

find $x \in G^{-1}(\text{dom } \varphi)$ *such that*

$$(\mathcal{P}_{\text{vip}}) \qquad \langle F(x), y - G(x) \rangle + \varphi(y) - \varphi(G(x)) \geq 0, \quad \forall\, y \in \mathbb{R}^n.$$

As usual, we will assume that $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous function. Furthermore, the mappings F, $G : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ are typically supposed to be continuous on an open set $\Omega \subset \mathbb{R}^n$ that contains the domain $G^{-1}(\text{dom } \varphi)$. The generalized variational inequality $(\mathcal{P}_{\text{vip}})$ subsumes many different classes of variational inequalities. In particular, the abstract problem $(\mathcal{P}_{\text{vip}})$ includes the family of *mixed variational inequalties*,

find $x \in \text{dom } \varphi$ *such that*

$$\langle F(x), y - x \rangle + \varphi(y) - \varphi(x) \geq 0, \quad \forall\, y \in \mathbb{R}^n,$$

which are also known as variational inequalities of the *second kind* or *hemivariational inequalities* [173, 181, 167, 175]. Moreover, in the special case $\varphi(\cdot) = \iota_K(\cdot)$, where $K \subset \mathbb{R}^n$ is a convex, nonempty, and closed set, the latter problem reduces to the *classical variational inequality*,

find $x \in K$ *such that*

$$\langle F(x), y - x \rangle \geq 0, \quad \forall\, y \in K,$$

which has been studied extensively during the last decades and is commonly used to model Nash equilibria problems, nonlinear complementarity problems, saddle point problems or problems arising in nonlinear mechanics or economics [113, 91, 77, 75, 126, 172]. Of course, there also exist other types of generalized variational inequalities that are not immediately covered by the problem $(\mathcal{P}_{\text{vip}})$. For instance, the function F can operate as a general multifunction, or the set K in the classical variational inequality may also depend on the solution x which leads to the class of so-called *quasi-variational inequalities*, see [164, 8, 176, 73, 74]. For more details on these different classes of variational inequalities and applications, we refer to the surveys and monographs [104, 77, 75, 76] and the references therein.

Similar to the derivation of the first order necessary optimality conditions in section 4.1, Solodov [223] showed that every solution $\bar{x} \in \mathbb{R}^n$ of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$ can be equivalently characterized by an alternative, proximal-type equation,

$$(\mathcal{E}_{\text{vip}}) \qquad V^{\Lambda}(\bar{x}) := G(\bar{x}) - \text{prox}_{\varphi}^{\Lambda}(G(\bar{x}) - \Lambda^{-1}F(\bar{x})) = 0, \quad \Lambda \in \mathbb{S}^n_{++},$$

that can be interpreted as a *natural residual* for the problem $(\mathcal{P}_{\text{vip}})$ and naturally extends the nonsmooth mapping F^Λ from the last chapter. Additionally, Solodov [223] also introduced a *regularized gap* and a *D-gap function* that generalize existing *merit functions* for classical variational inequality problems. Here, a nonnegative function $\varrho : \mathbb{R}^n \to \mathbb{R}_+$ is called merit function for the problem $(\mathcal{P}_{\text{vip}})$ if and only if the following correspondence is satisfied:

$$\bar{x} \textit{ is a solution of } (\mathcal{P}_{\text{vip}}) \quad \iff \quad \bar{x} \in G^{-1}(\text{dom}\ \varphi) \quad \textit{and} \quad \varrho(\bar{x}) = 0.$$

Thus, in general, merit functions allow to recast the problem $(\mathcal{P}_{\text{vip}})$ as constrained optimization problems of the form

$$(\mathcal{P}_{\text{mer}}) \qquad \min_x \ \varrho(x) \quad \text{s. t.} \quad G(x) \in \text{dom}\ \varphi,$$

such that every *global* solution $\bar{x}$ of the latter problem with $\varrho(\bar{x}) = 0$ is also a solution of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$. Now, based on these observations and motivated by our previous results, our key idea is to apply the semismooth Newton to solve the nonsmooth system $V^\Lambda(x) = 0$ and to combine it with an iterative algorithm for a merit function-based reformulation of the generalized variational inequalities. Here, we choose Solodov's D-gap function as a suitable merit function and a simple descent method for this purpose. Again, we embed these two different algorithmic components in a multidimensional filter framework to control the acceptance of the semismooth Newton steps and to achieve both global and local fast convergence.

Our approach can be seen as an extension of the algorithm presented in [120]. Here, Kanzow and Fukushima proposed a combination of the semismooth Newton method for $(\mathcal{E}_{\text{vip}})$ and a D-gap function-based descent method to solve box constrained variational inequalities. Moreover, a sufficient decrease condition is used to monitor the acceptance of the Newton steps and to establish global convergence. In this respect, let us also refer to von Heusinger et al. [245, 247] and Dreves et al. [68] where a similar approach is investigated in the context of generalized Nash equilibrium problems and using the so-called *Nikaido-Isoda function*. Sun et al. [227] and Kanzow and Fukushima [121] also studied a different type of algorithm that implements a generalized Newton scheme to minimize the D-gap function and to directly solve the corresponding optimization problem $(\mathcal{P}_{\text{mer}})$. Furthermore, for box constrained variational inequalities, Kanzow and Fukushima [121] showed that a stationary point x^* of the D-gap function is a solution of the variational inequality if the derivative $DF(x^*)$ is a P-matrix. In [224], Solodov and Tseng developed a dynamical parameter strategy for the D-gap function and obtained similar stationarity results for more general variational inequalities with a bounded feasible set K. Other D-gap function related approaches comprise the Newton-type methods presented in [189, 190] and are discussed extensively in [76, Section 10.4]. So far, the literature we have mentioned centers on methods that are based on the D-gap function and that utilize higher order information. Clearly, this only covers a small percentage of the different approaches and methodologies available for variational inequalities. However, since a more comprehensive survey is out of the scope of this work, we again refer to the excellent monographs by Facchinei and Pang [75, 76] and the references therein.

Our discussion of the generalized variational inequality problem $(\mathcal{P}_{\text{vip}})$ is strongly motivated by the observation that the D-gap function can be used to define an alternative base algorithm to substitute the proximal gradient method in Algorithm 2. Moreover, such a merit function-based approach may also allow us to globalize the semismooth Newton method in situations where first order objective function-based descent methods fail or are simply not available. In the following, we provide three different examples that connect the generalized variational inequality problem with other nonsmooth problems considered in this thesis.

- *Nonsmooth optimization problems.* In the case $G = I$ and $F = \nabla f$, the generalized variational inequality ($\mathcal{P}_{\text{vip}}$),

$$\langle \nabla f(x), y - x\rangle + \varphi(y) - \varphi(x) \geq 0,$$

  clearly coincides with the first order optimality conditions of the nonsmooth problem ($\mathcal{P}$), which were discussed and analyzed in chapter 4.

- *Convex composite problems.* In the following, let us reconsider the convex composite problem

$$\min_{x\in\mathbb{R}^n} \ \psi_c(x) := f(x) + \phi(F(x)),$$

  where $f : \mathbb{R}^n \to \mathbb{R}$, $F : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable functions and $\phi : \mathbb{R}^m \to (-\infty, +\infty]$ is a convex, proper, and lower semicontinuous mapping. Let $\bar{x} \in F^{-1}(\text{dom } \phi)$ be a local solution of the latter problem and suppose that Robinson's constraint qualification is satisfied at $\bar{x}$. Then, by Theorem 5.1.2, there exists a Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^m$ such that

$$\nabla f(\bar{x}) + DF(\bar{x})^\top \bar{\lambda} = 0, \quad \bar{\lambda} \in \partial\phi(F(\bar{x})).$$

  Moreover, due to Lemma 2.5.14, the inclusion $\bar{\lambda} \in \partial\phi(F(\bar{x}))$ is equivalent to $F(\bar{x}) \in \partial\phi^*(\bar{\lambda})$. Thus, the first order necessary conditions are satisfied if and only if the following two conditions are fulfilled:

$$\begin{aligned} \langle \nabla f(\bar{x}) + DF(\bar{x})^\top \bar{\lambda}, y - \bar{x}\rangle &= 0, \quad \forall\ y \in \mathbb{R}^n, \\ \phi^*(z) - \phi^*(\bar{\lambda}) - \langle F(\bar{x}), z - \bar{\lambda}\rangle &\geq 0, \quad \forall\ z \in \mathbb{R}^m. \end{aligned}$$

  By combining those two conditions, we conclude that $\bar{x}$ is stationary point of the convex composite problem if and only if there exists $\bar{\lambda} \in \mathbb{R}^m$ such that $(\bar{x}, \bar{\lambda})$ is a solution of the generalized variational inequality

  find $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ *such that:*

$$\left\langle \begin{pmatrix} \nabla f(x) + DF(x)^\top \lambda \\ -F(x) \end{pmatrix}, \begin{pmatrix} y \\ z \end{pmatrix} - \begin{pmatrix} x \\ \lambda \end{pmatrix} \right\rangle + \phi^*(z) - \phi^*(\lambda) \geq 0, \quad \forall\ (y, z) \in \mathbb{R}^n \times \mathbb{R}^m.$$

  Clearly, in this case, we have

$$x \equiv (x, \lambda), \quad F(x) \equiv \begin{pmatrix} \nabla f(x) + DF(x)^\top \lambda \\ -F(x) \end{pmatrix}, \quad G(x) \equiv \begin{pmatrix} x \\ \lambda \end{pmatrix},$$

  and $\varphi(x) \equiv \iota_{\mathbb{R}^n}(x) + \phi^*(\lambda)$.

- *Nonlinear saddle point problems.* Suppose that $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable function and let the mappings $\phi : \mathbb{R}^n \to (-\infty, +\infty]$ and $\psi : \mathbb{R}^m \to (-\infty, +\infty]$ be convex, proper, and lower semicontinuous. Furthermore, let us set

  $$\Theta : \mathbb{R}^n \times \mathbb{R}^m \to [-\infty, +\infty], \quad \Theta(x, y) := L(x, y) + \phi(x) - \psi^*(y)$$

  and let us consider the following general *saddle point problem*

  find $(\bar{x}, \bar{y}) \in \operatorname{dom} \varphi \times \operatorname{dom} \psi^*$ *such that:*

  $$\Theta(\bar{x}, y) \leq \Theta(\bar{x}, \bar{y}) \leq \Theta(x, \bar{y}), \quad \forall\ (x, y) \in \mathbb{R}^n \times \mathbb{R}^m.$$

  Then, using the convexity of ϕ, ψ^* and the differentiability of L, it can be shown that the point $(\bar{x}, \bar{y}) \in \operatorname{dom} \varphi \times \operatorname{dom} \psi^*$ is a solution of the latter problem if and only if it satisfies the following two conditions:

  $$\begin{aligned} \langle \nabla_x L(\bar{x}, \bar{y}), x - \bar{x} \rangle \ + \ \varphi(x) - \varphi(\bar{x}) \ &\geq 0, \quad \forall\ x \in \mathbb{R}^n, \\ \langle -\nabla_y L(\bar{x}, \bar{y}), y - \bar{y} \rangle + \psi^*(y) - \psi^*(\bar{y}) &\geq 0, \quad \forall\ y \in \mathbb{R}^m. \end{aligned}$$

  Thus, $(\bar{x}, \bar{y})$ is a saddle point of the function Θ if and only if it is a solution of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$ where

  $$x \equiv (x, y), \quad F(x) \equiv \begin{pmatrix} \nabla_x L(x, y) \\ -\nabla_y L(x, y) \end{pmatrix}, \quad G(x) \equiv \begin{pmatrix} x \\ y \end{pmatrix},$$

  and $\varphi(x) \equiv \phi(x) + \psi^*(y)$. The saddle point problem is also strongly related to the so-called *minimax* and *maximin* problems

  $$\inf_{x \in \mathbb{R}^n} \sup_{y \in \mathbb{R}^m} L(x, y) + \phi(x) - \psi^*(y), \qquad \sup_{y \in \mathbb{R}^m} \inf_{x \in \mathbb{R}^n} L(x, y) + \phi(x) - \psi^*(y).$$

  More specifically, if $(\bar{x}, \bar{y})$ is a saddle point, then it can be easily shown that $\bar{x}$ is a solution of minimax problem and $\bar{y}$ is a solution of the corresponding maximin problem, respectively. Hence, in the special setting

  $$K : \mathbb{R}^n \to \mathbb{R}^m, \qquad L(x, y) := \langle K(x), y \rangle,$$

  and using the identity $\psi^{**} = \psi$, our proposed variational framework can be applied to solve nonlinear and nonsmooth problems of the general form

  $$\min_x \ \varphi(x) + \psi(K(x)).$$

  At this point, let us mention that if the mapping $K : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then the latter problem can be solved by Chambolle and Pock's primal-dual algorithm [42] or via an alternating direction method [91, 71]. Furthermore and very recently, Clason and Valkonen [241, 51] also studied and analyzed an extended (primal-dual-based) approach for the more general nonlinear setting. Clearly, this demonstrates the relevance of the considered class of variational problems.

This chapter is organized as follows. In section 6.1, we derive several reformulations of the generalized variational inequality and based on [75, Chapter 2], we state different conditions that ensure solvability of the problem $(\mathcal{P}_{\text{vip}})$. Moreover, we will also reuse the concept of outer second order regularity and decomposability and present a new second-order type condition yielding local uniqueness of a solution of $(\mathcal{P}_{\text{vip}})$. In section 6.2, based on [76, 223], we discuss various basic properties of the regularized gap and the D-gap function for the generalized variational inequality problem $(\mathcal{P}_{\text{vip}})$. In particular, we will provide generalizations and extensions of different stationarity results that were derived by Facchinei and Pang in [76] for classical variational inequality problems and that can be used to characterize "optimality" of stationary points of the regularized gap function and the D-gap function. In section 6.3, we present an Armijo-type descent method and a globalized semismooth Newton method for the problem $(\mathcal{P}_{\text{vip}})$ and analyze their convergence properties in detail.

From now on, we will always assume that $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ is a conxex, proper, and lower semicontinuous mapping.

6.1. Characterization and existence of solutions

In the following, we briefly introduce monotonicity concepts for the generalized variational inequality and discuss conditions that guarantee existence of a solution of the problem $(\mathcal{P}_{\text{vip}})$. We start with an investigation of alternative representations of the problem $(\mathcal{P}_{\text{vip}})$ resembling the first order optimality conditions that were analyzed in the previous chapters. At first, let us note that the variational inequality $(\mathcal{P}_{\text{vip}})$ can be equivalently rephrased as the following generalized equation:

find $x \in G^{-1}(\text{dom } \varphi)$ such that

$$-F(x) \in \partial\varphi(G(x)).$$

Thus, by (3.1.2), this immediately implies that $x \in G^{-1}(\text{dom } \varphi)$ is a solution of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$, if and only if the point x is a zero of the so-called *natural residual* $V^\Lambda : \mathbb{R}^n \to \mathbb{R}^n$,

$$V^\Lambda(x) := G(x) - \text{prox}_\varphi^\Lambda(G(x) - \Lambda^{-1}F(x)) = 0, \tag{6.1.1}$$

where $\Lambda \in \mathbb{S}^n_{++}$ is an arbitrary parameter matrix. Again, let us notice that the nonsmooth mapping V^Λ plays a similar role as its counterpart $F^\Lambda(x) = x - \text{prox}_\varphi^\Lambda(x - \Lambda^{-1}\nabla f(x))$ from the previous chapters. Now, if $x \in \mathbb{R}^n$ is a solution of the problem $(\mathcal{P}_{\text{vip}})$, then it holds

$$\langle F(x), th\rangle + \varphi(G(x) + th) - \varphi(G(x)) \geq 0, \quad \forall\ t > 0, \quad \forall\ h \in \mathbb{R}^n$$

and hence, we can infer

$$\langle F(x), h\rangle + \varphi^\downarrow(G(x); h) \geq 0, \quad \forall\ h \in \mathbb{R}^n. \tag{6.1.2}$$

On the other hand, by setting $h = y - G(x)$ and using Lemma 2.5.5, the latter inequality also implies

$$\langle F(x), y - G(x)\rangle + \varphi(y) - \varphi(G(x)) \geq 0, \quad \forall\ y \in \mathbb{R}^n.$$

In summary, these simple computations establish the following lemma, which apparently does not need a proof.

Lemma 6.1.1. *Let $F, G : \Omega \to \mathbb{R}^n$ be given functions and suppose that the open set $\Omega \subset \mathbb{R}^n$ contains the domain $G^{-1}(\mathrm{dom}\ \varphi)$. Then, the following conditions are mutually equivalent:*

(i) *The point $\bar{x}$ is a solution of the generalized variational inequality $(\mathcal{P}_{\mathrm{vip}})$.*

(ii) *It holds $-F(\bar{x}) \in \partial\varphi(G(\bar{x}))$.*

(iii) *The point $\bar{x} \in G^{-1}(\mathrm{dom}\ \varphi)$ satisfies the condition*

$$\langle F(\bar{x}), h\rangle + \varphi^{\downarrow}(G(\bar{x}); h) \geq 0, \quad \forall\ h \in \mathbb{R}^n.$$

(iv) *Let $\Lambda \in \mathbb{S}^n_{++}$ be arbitrary. The point $\bar{x}$ is a solution of the fixed point-type equation*

$$V^{\Lambda}(\bar{x}) = G(\bar{x}) - \mathrm{prox}^{\Lambda}_{\varphi}(G(\bar{x}) - \Lambda^{-1}F(\bar{x})) = 0.$$

Similar to our terminology for convex composite problems and to Definition 5.1.5, we now define the *critical cone* associated with the generalized variational inequality $(\mathcal{P}_{\mathrm{vip}})$ via

$$\mathcal{C}(x) := \{h \in \mathbb{R}^n : \langle F(x), h\rangle + \varphi^{\downarrow}(G(x); h) \leq 0\}.$$

Hence, if $\bar{x}$ is a solution of the problem $(\mathcal{P}_{\mathrm{vip}})$, then the critical cone $\mathcal{C}(\bar{x})$ again coincides with the normal cone $N_{\partial\varphi(G(\bar{x}))}(-F(\bar{x}))$. In the subsequent sections, we will also work with the cone

$$\mathcal{C}_G(x) := \{h \in \mathbb{R}^n : DG(x)h \in \mathcal{C}(x)\} = DG(x)^{-1}\mathcal{C}(x),$$

which we will refer to as the *G-critical cone.*

Next, we state several monotonicity concepts for the function F that will be used to derive and formulate existence conditions. The following definition is quite standard, see, e.g., [75, Section 2.3 and Definition 2.3.1] or [223, 176].

Definition 6.1.2. *Let $\Omega \subset \mathbb{R}^n$ be an open set and let the functions $F, G : \Omega \to \mathbb{R}^n$ be given. Then, the mapping F is called*

(i) G-monotone *on Ω if it holds*

$$\langle F(x) - F(y), G(x) - G(y)\rangle \geq 0, \quad \forall\ x, y \in \Omega.$$

(ii) strictly G-monotone *on Ω if it holds*

$$\langle F(x) - F(y), G(x) - G(y)\rangle > 0, \quad \forall\ x, y \in \Omega \quad \textit{and} \quad x \neq y.$$

(iii) (ξ, G)-monotone *on Ω for some $\xi > 1$ if there exists a constant $\mu > 0$ such that*

$$\langle F(x) - F(y), G(x) - G(y)\rangle \geq \mu\|x - y\|^{\xi}, \quad \forall\ x, y \in \Omega.$$

Furthermore, F is called strongly G-monotone *if F is $(2, G)$-monotone on Ω.*

If the function G is the identity mapping then we will drop the "G-" prefix in Definition 6.1.2 and the latter properties coincide with the conventional monotonicity concepts. Similar to [75, Proposition 2.3.2], the monotonicity properties of F and G can also be alternatively characterized via their derivatives and appropriate semidefiniteness assumptions. We will not give a proof here.

Lemma 6.1.3. *Let $\Omega \subset \mathbb{R}^n$ be an open set and let the functions $F, G : \Omega \to \mathbb{R}^n$ be continuously differentiable on Ω. Then, it holds:*

(i) *F is G-*monotone *on Ω if and only if*

$$\langle DF(x)h, DG(x)h\rangle \geq 0, \quad \forall\, h \in \mathbb{R}^n, \quad \forall\, x \in \Omega.$$

(ii) *F is* strictly G-monotone *on Ω if it holds*

$$\langle DF(x)h, DG(x)h\rangle > 0, \quad \forall\, h \in \mathbb{R}^n \setminus \{0\}, \quad \forall\, x \in \Omega.$$

(iii) *F is* $(2, G)$-monotone *on Ω if and only if there exists a constant $\mu > 0$ such that*

$$\langle DF(x)h, DG(x)h\rangle \geq \mu\|h\|^2, \quad \forall\, h \in \mathbb{R}^n, \quad \forall\, x \in \Omega.$$

The next lemma presents several basic existence conditions and results for the generalized and the mixed variational inequality. A more refined and thorough discussion of the existence of solutions for classical variational inequalities can be found in the monograph [75, Sections 2 and 3]. Motivated by [75, Proposition 2.2.3, Theorem 2.3.3, and Exercise 2.9.11], we obtain the following result.

Lemma 6.1.4. *Let $F, G : \Omega \to \mathbb{R}^n$ be given and let $\Omega \subset \mathbb{R}^n$ be an open set that contains the domain $G^{-1}(\mathrm{dom}\ \varphi)$. It holds:*

(i) *If F is strictly G-monotone on Ω, then the generalized variational inequality $(\mathcal{P}_{\mathrm{vip}})$ has at most one solution.*

Now, suppose that G is the identity mapping. Then, the following two statements are valid.

(ii) *Let us further suppose that φ is coercive on Ω and there exist $x^* \in \mathrm{dom}\ \varphi$ and constants $\vartheta > 0$, $\xi \geq 0$ such that*

$$\liminf_{\|x\|\to\infty,\, x\in\Omega} \frac{\langle F(x), x - x^*\rangle}{\|x\|^\xi} \geq \vartheta.$$

Then, the problem $(\mathcal{P}_{\mathrm{vip}})$ has a solution.

(iii) *If F is ξ-monotone for some $\xi > 1$ and if there exists $x^* \in \mathrm{dom}\ \varphi$ such that $\partial\varphi(x^*) \neq \emptyset$, then the generalized variational inequality $(\mathcal{P}_{\mathrm{vip}})$ has a unique solution on Ω.*

Proof. Suppose that F is strictly G-monotone on Ω and let $\bar{x}, \hat{x} \in G^{-1}(\mathrm{dom}\ \varphi)$, $\bar{x} \neq \hat{x}$, be two different solutions of $(\mathcal{P}_{\mathrm{vip}})$. Then, it holds

$$\begin{aligned}
&\langle F(\bar{x}), G(\hat{x}) - G(\bar{x})\rangle + \varphi(G(\hat{x})) - \varphi(G(\bar{x})) \geq 0,\\
&\langle F(\hat{x}), G(\bar{x}) - G(\hat{x})\rangle + \varphi(G(\bar{x})) - \varphi(G(\hat{x})) \geq 0.
\end{aligned}$$

Thus, adding both inequalities and using the strict G-monotonicity of F yields the following contradiction

$$0 \leq \langle F(\bar{x}) - F(\hat{x}), G(\hat{x}) - G(\bar{x}) \rangle < 0.$$

The remaining existence results essentially follow from [75, Exercise 2.9.11]. A more detailed proof is provided in the appendix in section A.3. □

Remark 6.1.5. In the general case $G \neq I$, existence of solutions can be established when the functions F and G are Lipschitz continuous and strongly monotone. In this situation, the mapping

$$x \mapsto x - V^{\Lambda}(x), \quad \Lambda \in \mathbb{S}^n_{++},$$

can be shown to be a contraction and thus, by Banach's famous fixed point theorem, must possess a unique fixed point that is also a solution of the generalized variational inequality ($\mathcal{P}_{\text{vip}}$). Let us emphasize that this result heavily relies on a correct balance of the Lipschitz constants and the monotonicity parameters and is only valid in certain situations. Here, we will not present these dependencies. However, let us refer to Noor et al. [174, 177] for a more detailed examination and similar results.

In the following, based on the second order theory in chapter 5, we introduce second order-type conditions that ensure local uniqueness of solutions. Similar to our observations in chapter 5, the concept of outer second order regularity allows to formulate rather mild conditions. In particular, this theorem generalizes and extends a related result of Facchinei and Pang [75, Proposition 3.3.4], see also [75, Section 3.3].

Theorem 6.1.6. *Suppose that $F, G : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ are continuously differentiable on an open set Ω containing $G^{-1}(\text{dom } \varphi)$ and let $\bar{x} \in G^{-1}(\text{dom } \varphi)$ be a solution of the problem ($\mathcal{P}_{\text{vip}}$). Furthermore, let G be twice continuously differentiable on a neighborhood of $\bar{x}$ and assume that the mapping φ is outer second order regular at $G(\bar{x})$ on $DG(\bar{x})\mathcal{C}_G(\bar{x})$. Then, the second order-type condition*

$$2\langle DF(\bar{x})h, DG(\bar{x})h \rangle - \xi^*_{\varphi,h}(-F(\bar{x})) > 0, \quad \forall\, h \in \mathcal{C}_G(\bar{x}) \setminus \{0\}, \tag{6.1.3}$$

where $\xi_{\varphi,h}(\cdot) := \varphi^{\downarrow\downarrow}_{-}(G(\bar{x}); DG(\bar{x})h; \cdot)$, implies that $\bar{x}$ is an isolated solution of the generalized variational inequality ($\mathcal{P}_{\text{vip}}$).

Proof. The proof of this theorem is similar to the proof of Theorem 5.2.8 (ii), see also section A.2. Let us suppose that $\bar{x}$ is not an isolated solution of the generalized variational inequality ($\mathcal{P}_{\text{vip}}$). Then, there exists a sequence $(x^k)_k$ of solutions of ($\mathcal{P}_{\text{vip}}$) with $x^k \neq \bar{x}$, for all $k \in \mathbb{N}$, that converges to $\bar{x}$, as $k \to \infty$. Furthermore, setting $t_k := \|x^k - \bar{x}\|$ and $h^k := t_k^{-1}(x^k - \bar{x})$, we may assume that h^k converges to some point $h \neq 0$ (after extracting an appropriate subsequence if necessary). Using the optimality of x^k and a first order Taylor expansion of $G(x^k)$ at $\bar{x}$, we readily obtain

$$\langle F(x^k), t_k DG(\bar{x})h^k \rangle + \varphi(G(\bar{x}) + t_k DG(\bar{x})h^k + o(t_k)) - \varphi(G(\bar{x})) + o(t_k) \leq 0.$$

Thus, dividing both sides of the latter inequality by t_k and taking the limes inferior $k \to \infty$, this yields

$$\langle F(\bar{x}), DG(\bar{x})h \rangle + \varphi^{\downarrow}(G(\bar{x}); DG(\bar{x})h) \leq 0.$$

Consequently, by Lemma 6.1.1 (iii) and the definition of the G-critical cone, it follows $h \in \mathcal{C}_G(\bar{x}) \setminus \{0\}$. Next, writing

$$x^k = \bar{x} + t_k h + \tfrac{1}{2} t_k^2 \cdot [2 t_k^{-1}(h^k - h)], \quad w^k := 2 t_k^{-1}(h^k - h),$$

we obviously have $t_k w^k \to 0$ and hence, a second order Taylor expansion of G at $\bar{x}$ yields

$$G(x^k) = G(\bar{x}) + t_k DG(\bar{x})h + \tfrac{1}{2} t_k^2 \nu^k, \quad \nu^k := DG(\bar{x})w^k + D^2G(\bar{x})[h,h] + o(1).$$

Accordingly, we obtain

$$\begin{aligned} \langle F(x^k), G(x^k) - G(\bar{x})\rangle &= \langle F(\bar{x}) + t_k DF(\bar{x})h^k + o(t_k), G(x^k) - G(\bar{x})\rangle \\ &= -t_k \varphi^{\downarrow}(G(\bar{x}); DG(\bar{x})h) + \tfrac{1}{2} t_k^2 \langle F(\bar{x}), \nu^k\rangle + t_k^2 \langle DF(\bar{x})h^k, DG(\bar{x})h\rangle + o(t_k^2). \end{aligned}$$

Combining the latter facts, this immediately establishes

$$\varphi(G(\bar{x}) + t_k DG(\bar{x})h + \tfrac{1}{2} t_k^2 \nu^k) - \varphi(G(\bar{x})) - t_k \varphi^{\downarrow}(G(\bar{x}); DG(\bar{x})h) \le \tfrac{1}{2} t_k^2 \tau^k,$$

where we set $\tau^k := -\langle F(\bar{x}), \nu^k\rangle - 2\langle DF(\bar{x})h^k, DG(\bar{x})h\rangle - o(1)$ and used the fact that x^k is a solution of the generalized variational inequality ($\mathcal{P}_{\mathrm{vip}}$) for all $k \in \mathbb{N}$. Hence, due to $t_k(\nu^k, \tau^k) \to 0$, $k \to \infty$, and $h \in \mathcal{C}_G(\bar{x})$, the outer second order regularity of φ is applicable and there exist sequences $(\tilde{\nu}^k)_k$ and $(\tilde{\tau}^k)_k$ such that

$$\tilde{\tau}^k - \tau^k \to 0, \quad \tilde{\nu}^k - \nu^k \to 0, \quad \tilde{\tau}^k \ge \varphi_-^{\downarrow\downarrow}(G(\bar{x}); DG(\bar{x}); \tilde{\nu}^k)$$

for all $k \in \mathbb{N}$. Altogether, we now obtain

$$\begin{aligned} 0 &\ge \tau^k - \tilde{\tau}^k + \langle F(\bar{x}), \nu^k - \tilde{\nu}^k\rangle + o(1) + 2\langle DF(\bar{x})h^k, DG(\bar{x})h\rangle \\ &\qquad + \langle F(\bar{x}), \tilde{\nu}^k\rangle + \varphi_-^{\downarrow\downarrow}(G(\bar{x}); DG(\bar{x})h, \tilde{\nu}^k) \\ &\ge \tau^k - \tilde{\tau}^k + \langle F(\bar{x}), \nu^k - \tilde{\nu}^k\rangle + o(1) + 2\langle DF(\bar{x})h^k, DG(\bar{x})h\rangle - \xi^*_{\varphi,h}(-F(\bar{x})). \end{aligned}$$

Finally, taking the limit $k \to \infty$, this clearly contradicts condition (6.1.3). □

Remark 6.1.7. Let us note that the following second order-type condition

$$\langle DF(\bar{x})h, DG(\bar{x})h\rangle > 0, \quad \forall\ h \in \mathcal{C}_G(\bar{x}) \setminus \{0\},$$

does also ensure local uniqueness of a solution $\bar{x} \in G^{-1}(\mathrm{dom}\ \varphi)$ of the generalized variational inequality ($\mathcal{P}_{\mathrm{vip}}$) if the function φ is not outer second order regular or if G is not twice differentiable. This result can be derived in a similar (but easier) fashion to Theorem 6.1.6. In particular, reusing the notation and mimicking the first steps of the proof of Theorem 6.1.6, a first order Taylor expansion of $F(x^k)$ and $G(x^k)$ at $\bar{x}$ establishes

$$\begin{aligned} 0 &\ge \langle F(x^k), G(x^k) - G(\bar{x})\rangle + \varphi(G(x^k)) - \varphi(G(\bar{x})) \\ &\ge \langle F(\bar{x}) + t_k DF(\bar{x})h^k + o(t^k), G(x^k) - G(\bar{x})\rangle - \langle F(\bar{x}), G(x^k) - G(\bar{x})\rangle \\ &= t_k^2 \langle DF(\bar{x})h^k, DG(\bar{x})h^k\rangle + o(t_k^2). \end{aligned}$$

Clearly, taking the limit $k \to \infty$, this produces the same contradiction as in the proof of Theorem 6.1.6. Again, let us refer to [75, Proposition 3.3.4] for comparison.

For general mixed variational inequalities of the form ($\mathcal{P}_{\text{vip}}$) the second order-type condition (6.1.3) seems to be new. Let us also mention that Shapiro [219] derived several local uniqueness results for classical variational inequalities that can be associated with a quadratic growth condition for the regularized gap function and that can be established under weaker differentiability assumptions. In the following, we briefly discuss the curvature term $\xi^*_{\varphi,h}(-F(\bar{x}))$ and present some implications in the decomposable setting.

Suppose that $\bar{x} \in G^{-1}(\text{dom } \varphi)$ is a solution of the problem ($\mathcal{P}_{\text{vip}}$) and let $h \in \mathcal{C}_G(\bar{x})$ be arbitrary. Then, using $-F(\bar{x}) \in \partial\varphi(G(\bar{x}))$ and similar to Remark 5.2.9 and the discussion in section 5.5, it holds

$$\xi_{\varphi,h}(w) = \varphi_{-}^{\downarrow\downarrow}(G(\bar{x}); DG(\bar{x})h, w) \geq -\langle F(\bar{x}), w\rangle, \quad \forall\, w \in \mathbb{R}^n.$$

Thus, it follows $-\xi^*_{\varphi,h}(-F(\bar{x})) = \inf_w \langle F(\bar{x}), w\rangle + \xi_{\varphi,h}(w) \geq 0$. Moreover, this shows that the second-order type conditions (6.1.3) are generally weaker than the sufficient conditions presented in [75, Proposition 3.3.3 and Remark 3.3.4] or in Remark 6.1.7.

Now, let us assume that φ is C^2-fully decomposable at $G(\bar{x})$ and let (φ_d, H) be a corresponding decomposition pair. Then, due to Lemma 5.3.5, φ is twice directionally epidifferentiable and outer second order regular at $G(\bar{x})$ in all directions $h \in \mathbb{R}^n$ with $DH(G(\bar{x}))h \in \text{dom } \varphi_d$. In particular, this implies that the function φ is outer second order regular at $G(\bar{x})$ on $DG(\bar{x})\mathcal{C}_G(\bar{x})$ and on $\mathcal{C}(\bar{x}) = N_{\partial\varphi(G(\bar{x}))}(-F(\bar{x}))$. Furthermore, as in section 5.5, we can establish the following equivalence:

$$-F(\bar{x}) \in \partial\varphi(G(\bar{x})) \quad \iff \quad \exists\, \bar{\mu} \in \partial\varphi_d(0): \quad F(\bar{x}) + DH(G(\bar{x}))^\top \bar{\mu} = 0.$$

Next, as shown in Lemma 5.1.11, the nondegeneracy condition,

$$DH(G(\bar{x}))\mathbb{R}^n + \text{lin } N_{\partial\varphi_d(0)}(\bar{\mu}) = \mathbb{R}^m,$$

implies that the vector $\bar{\mu} \in \mathbb{R}^m$ is uniquely determined and hence, analogous to Lemma 5.3.9, it follows

$$-\xi^*_{\varphi,h}(-F(\bar{x})) = \langle \bar{\mu}, D^2H(G(\bar{x}))[DG(\bar{x})h, DG(\bar{x})h]\rangle, \quad \forall\, h \in \mathcal{C}_G(\bar{x}).$$

Consequently, the curvature term $-\xi^*_{\varphi,h}(-F(\bar{x}))$ can again be represented as an appropriate quadratic form. Moreover, under the strict complementarity condition,

$$-F(\bar{x}) \in \text{ri } \partial\varphi(G(\bar{x})),$$

and invoking Lemma 5.3.32, we can infer that the proximity operator $\text{prox}_\varphi^\Lambda$ has to be Fréchet differentiable at the point $\bar{u} := G(\bar{x}) - \Lambda^{-1}F(\bar{x})$. Additionally, in this situation, the critical cone $\mathcal{C}(\bar{x}) = N_{\partial\varphi(G(\bar{x}))}(-F(\bar{x}))$ is also a subspace. Collecting these preparatory components and reconsidering the proof of Theorem 5.3.26 or the different steps in section 5.5, it is possible to derive the following prox-based, intrinsic characterization of the curvature term

$$-\xi^*_{\varphi,h}(-F(\bar{x})) = \langle DG(\bar{x})h, [\Lambda^{-\frac{1}{2}}\mathcal{Q}^\Lambda_\varphi(\bar{u})^+\Lambda^{\frac{1}{2}} - I]DG(\bar{x})h\rangle_\Lambda, \quad \forall\ h \in \mathcal{C}_G(\bar{x}),$$

where $\mathcal{Q}^\Lambda_\varphi(\bar{u}) = \Lambda^{\frac{1}{2}}D\mathrm{prox}^\Lambda_\varphi(\bar{u})\Lambda^{-\frac{1}{2}}$. We will not go into further detail here. However, let us note that this representation will again be advantageous and helpful when analyzing local convergence properties of the semismooth Newton method later on.

6.2. Merit and gap functions for GVIPs

In this section, we introduce the regularized gap function and the D-gap function for the generalized variational inequality problem ($\mathcal{P}_{\mathrm{vip}}$) and discuss their main properties in detail.

6.2.1. The regularized gap function

The regularized gap function was originally proposed by Auchmuty [5] and Fukushima [86] for classical variational inequalities. For a given function $F : \mathbb{R}^n \to \mathbb{R}^n$ and a set $K \subset \mathbb{R}^n$ it takes the form:

$$G_{\mathrm{gap}} : \mathbb{R}^n \to \mathbb{R}, \quad x \mapsto G_{\mathrm{gap}}(x) := \max_{y\in K}\ \langle F(x), x - y\rangle - \frac{1}{2}\|x - y\|^2_\Lambda, \quad \Lambda \in \mathbb{S}^n_{++}.$$

Indeed, Auchmuty and Fukushima were the first who showed that this gap function is an appropriate merit function for the variational inequality problem and that it possesses all the properties described in the beginning of this chapter. Its overall popularity primarily stems from the fact that the regularized gap function is continuously differentiable whenever the function F is continuously differentiable. Thus, first order methods, such as the projected gradient descent method, can be applied to solve the corresponding gap function-based optimization problem ($\mathcal{P}_{\mathrm{mer}}$). Moreover, the regularized gap function is also often utilized to construct a globalization framework for fast, but only locally convergent methods, see, e.g., [231, 269] and [76, Section 10.4.4]. For more information on the regularized gap function and other merit function-based approaches, we refer to [86, 185, 186, 76].

In the following, we present an extended definition of the regularized gap function for generalized variational inequalities that is due to Solodov [223]. Let us also mention that Patriksson [185, 186] considered a similar extension for a specific class of GVIPs with $G = I$.

Definition 6.2.1. *Let the functions $F : \Omega \to \mathbb{R}^n$ and $G : \Omega \to \mathbb{R}^n$ be defined on an open set $\Omega \subset \mathbb{R}^n$ that contains the domain $G^{-1}(\mathrm{dom}\ \varphi)$ and let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary parameter matrix. The* regularized gap function *$G^\Lambda : \Omega \to [-\infty, +\infty]$ is defined as*

$$G^\Lambda(x) := \max_{y\in\mathbb{R}^n} \left\{\langle F(x), G(x) - y\rangle + \varphi(G(x)) - \varphi(y) - \frac{1}{2}\|G(x) - y\|^2_\Lambda\right\},$$

for all $x \in \Omega$.

The following lemma establishes an alternative representation of the regularized gap function G^Λ that will be useful for our further investigation.

Lemma 6.2.2. *Let $F, G : \Omega \to \mathbb{R}^n$ be given and let $\Omega \subset \mathbb{R}^n$ be an open set that contains the domain $G^{-1}(\mathrm{dom}\ \varphi)$. Let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary parameter matrix. Then, it holds*

$$G^\Lambda(x) = \frac{1}{2}\|F(x)\|^2_{\Lambda^{-1}} + \varphi(G(x)) - \mathrm{env}^\Lambda_\varphi(G(x) - \Lambda^{-1}F(x))$$

for all $x \in \Omega$.

Proof. A direct calculation shows

$$\begin{aligned} G^\Lambda(x) &= \varphi(G(x)) - \min_{y\in\mathbb{R}^n} \left\{\varphi(y) + \frac{1}{2}\left(2\langle -\Lambda^{-1}F(x), \Lambda(G(x)-y)\rangle + \|G(x)-y\|^2_\Lambda\right)\right\} \\ &= \varphi(G(x)) + \frac{1}{2}\|\Lambda^{-1}F(x)\|^2_\Lambda - \min_{y\in\mathbb{R}^n}\left\{\varphi(y) + \frac{1}{2}\|G(x) - \Lambda^{-1}F(x) - y\|^2_\Lambda\right\} \\ &= \frac{1}{2}\|F(x)\|^2_{\Lambda^{-1}} + \varphi(G(x)) - \mathrm{env}^\Lambda_\varphi(G(x) - \Lambda^{-1}F(x)), \end{aligned}$$

as desired. □

Unfortunately, Lemma 6.2.2 immediately implies that the regularized gap function cannot be expected to be continuously differentiable in general. Clearly, in the classical setting, the nonsmooth term

$$\varphi(G(x)) = \iota_K(G(x)) = \begin{cases} 0 & \text{if } G(x) \in K, \\ +\infty & \text{otherwise} \end{cases}$$

vanishes whenever the vector $G(x)$ is feasible. However, this is not the case for more general choices of φ. Now, throughout this section, we will use the following abbreviation

$$u(x) := G(x) - \Lambda^{-1}F(x), \quad \Lambda \in \mathbb{S}^n_{++}.$$

Next, we derive several important properties of the regularized gap function and verify that G^Λ is an appropriate merit function for the generalized variational inequality ($\mathcal{P}_{\mathrm{vip}}$). See also [223, Theorem 4] and [76, Theorem 10.2.3, Remark 10.3.8] for similar results.

Lemma 6.2.3. *Let $F : \Omega \to \mathbb{R}^n$ and $G : \Omega \to \mathbb{R}^n$ be two given continuous functions and let $\Omega \subset \mathbb{R}^n$ be an open set containing the domain $G^{-1}(\mathrm{dom}\ \varphi)$. Let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary parameter matrix. Then, the following two statements are valid.*

(i) *The regularized gap function is lower semicontinuous on Ω and it holds*

$$G^\Lambda(x) \geq \frac{1}{2}\|V^\Lambda(x)\|^2_\Lambda \geq 0, \qquad \forall\ x \in \Omega.$$

(ii) *It holds $G^\Lambda(x) = 0$ if and only if x is a solution of the variational inequality ($\mathcal{P}_{\mathrm{vip}}$).*

Proof. Since the Moreau envelope $\mathrm{env}^\Lambda_\varphi$ is continuously differentiable, (see Lemma 3.1.5), the lower semicontinuity of the gap function G^Λ follows from the continuity of F, G, and the lower semicontinuity of φ. Moreover, an easy calculation yields

$$\begin{aligned}G^\Lambda(x) &= \frac{1}{2}\|F(x)\|^2_{\Lambda^{-1}} - \frac{1}{2}\|\mathrm{prox}^\Lambda_\varphi(u(x)) - u(x)\|^2_\Lambda + \varphi(G(x)) - \varphi(\mathrm{prox}^\Lambda_\varphi(u(x))) \\ &\geq \frac{1}{2}\|F(x)\|^2_{\Lambda^{-1}} - \frac{1}{2}\|\Lambda^{-1}F(x) - V^\Lambda(x)\|^2_\Lambda + \langle \nabla \mathrm{env}^\Lambda_\varphi(u(x)), V^\Lambda(x)\rangle \\ &= \langle F(x), V^\Lambda(x)\rangle - \frac{1}{2}\|V^\Lambda(x)\|^2_\Lambda + \langle \Lambda V^\Lambda(x) - F(x), V^\Lambda(x)\rangle = \frac{1}{2}\|V^\Lambda(x)\|^2_\Lambda.\end{aligned}$$

Clearly, this also shows that any zero of the regularized gap function G^Λ is a solution of the problem $(\mathcal{P}_{\mathrm{vip}})$. On the other hand, if x is a solution of the generalized variational inequality, then it holds $G(x) = \mathrm{prox}^\Lambda_\varphi(u(x))$ and it follows

$$G^\Lambda(x) = \frac{1}{2}\|F(x)\|^2_{\Lambda^{-1}} - \frac{1}{2}\|G(x) - u(x)\|^2_\Lambda = 0.$$

This finishes the proof of Lemma 6.2.3. □

The latter results clearly suggest to compute solutions of the generalized variational inequality $(\mathcal{P}_{\mathrm{vip}})$ via minimization of the regularized gap function. More specifically, let us define the so-called *regularized gap program:*

$$\min_x \; G^\Lambda(x), \quad \text{s.t.} \quad G(x) \in \mathrm{dom}\;\varphi. \tag{6.2.1}$$

Now, if $\bar{x}$ is a stationary point of the latter problem satisfying $G^\Lambda(\bar{x}) = 0$, then Lemma 6.2.3 implies that this point is also a solution of the variational inequality $(\mathcal{P}_{\mathrm{vip}})$. In this respect, any stationary point of the regularized gap program will be called *variational optimal* if it is also a solution of the problem $(\mathcal{P}_{\mathrm{vip}})$. Obviously, every variational optimal stationary point is automatically a *global* solution the minimization problem (6.2.1).

In general, however, we cannot expect that stationary points of the regularized gap function or of the regularized gap program (6.2.1) are solutions of the generalized variational inequality $(\mathcal{P}_{\mathrm{vip}})$ without any further assumptions. In the following, based on related results for classical variational inequalities in [76, Section 10.2.1], we want to derive several equivalent conditions that guarantee variational optimality of stationary points of the regularized gap function.

Let us note that such a discussion is particularly important if the regularized gap function G^Λ is used as a merit function for optimization problems of the type $(\mathcal{P})$ or $(\mathcal{P}_c)$, as has already been indicated in the introductory part of this chapter. Here, solutions of the generalized variational inequality $(\mathcal{P}_{\mathrm{vip}})$ or of the nonsmooth equation (6.1.1) correspond to stationary points of the initial minimization problem and thus, variational optimality is an essential requirement for global convergence.

Thus, let $\bar{x}$ be an arbitrary stationary point of problem (6.2.1) and suppose that Robinson's constraint qualification

$$0 \in \mathrm{int}\{G(\bar{x}) + DG(\bar{x})\mathbb{R}^n - \mathrm{dom}\;\varphi\}$$

is satisfied at $\bar{x}$. Then, using

$$\nabla \mathrm{env}^\Lambda_\varphi(u(\bar{x})) = \Lambda(u(\bar{x}) - \mathrm{prox}^\Lambda_\varphi(u(\bar{x}))) = \Lambda V^\Lambda(\bar{x}) - F(\bar{x}),$$

the corresponding first order necessary optimality conditions take the following form

$$\begin{aligned}(G^\Lambda)^\downarrow(\bar{x};h) &= \varphi^\downarrow(G(\bar{x});DG(\bar{x})h) + F(\bar{x})^\top\Lambda^{-1}DF(\bar{x})h \\ &\qquad - \nabla\mathrm{env}^\Lambda_\varphi(u(\bar{x}))^\top(DG(\bar{x}) - \Lambda^{-1}DF(\bar{x}))h \\ &= \varphi^\downarrow(G(\bar{x});DG(\bar{x})h) + \langle V^\Lambda(\bar{x}), DF(\bar{x})h\rangle + \langle F(\bar{x}) - \Lambda V^\Lambda(\bar{x}), DG(\bar{x})h\rangle \geq 0\end{aligned}$$

for all $h \in \mathbb{R}^n$. Moreover, by using Lemma 2.5.5, Lemma 2.5.11 (ii), and the characterization of the proximity operator, we obtain

$$\begin{aligned}\text{(6.2.2)} \qquad &\varphi(\mathrm{prox}^\Lambda_\varphi(u(x)) + h) - \varphi(\mathrm{prox}^\Lambda_\varphi(u(x))) \\ &\qquad \geq \varphi^\downarrow(\mathrm{prox}^\Lambda_\varphi(u(x)), h) \geq \langle\nabla\mathrm{env}^\Lambda_\varphi(u(x)), h\rangle = \langle\Lambda V^\Lambda(x) - F(x), h\rangle,\end{aligned}$$

for any vector $x \in \mathbb{R}^n$ and all $h \in \mathbb{R}^n$. Clearly, this yields

$$\varphi^\downarrow(\mathrm{prox}^\Lambda_\varphi(u(x)), h) > -\infty, \quad \forall\, h \in \mathbb{R}^n,$$

and, if the function φ is additionally continuous at $\mathrm{prox}^\Lambda_\varphi(u(x))$ then we have

$$\mathrm{dom}\ \varphi^\downarrow(\mathrm{prox}^\Lambda_\varphi(u(x)), \cdot) = \mathbb{R}^m.$$

Now, for a vector $x \in G^{-1}(\mathrm{dom}\ \varphi)$ and an arbitrary matrix $\Lambda \in \mathbb{S}^n_{++}$, we define the sets

$$\mathcal{T}^\Lambda(x) := \{h \in \mathbb{R}^n : \varphi^\downarrow(G(x);h) + \varphi^\downarrow(\mathrm{prox}^\Lambda_\varphi(u(x)); -h) \leq 0\}$$

and

$$\mathcal{N}(x) := \{h \in \mathbb{R}^n : \varphi^\downarrow(G(x);h) + \langle F(x), h\rangle = 0\}.$$

Since the directional epiderivatives $\varphi^\downarrow(G(x);\cdot)$ and $\varphi^\downarrow(\mathrm{prox}^\Lambda_\varphi(u(x));\cdot)$ are convex and positively homogeneous, it immediately follows that the sets $\mathcal{T}^\Lambda(x)$ and $\mathcal{N}(x)$ are convex cones. Furthermore, if $\bar{x}$ is a stationary point of the regularized gap function G^Λ and Robinson's constraint qualification is satisfied at $\bar{x}$, then our latter computations and Lemma 2.5.11 (i) imply

$$\text{(6.2.3)} \qquad \varphi^\downarrow(G(\bar{x});0) = 0.$$

Consequently, φ is subdifferentiable at $G(\bar{x})$ and again by applying Lemma 2.5.11 (ii), we have

$$\varphi^\downarrow(G(\bar{x});h) > -\infty, \quad \forall\, h \in \mathbb{R}^n.$$

In this case, the cones $\mathcal{T}^\Lambda(\bar{x})$ and $\mathcal{N}(\bar{x})$ are nonempty and, due to the lower semicontinuity of the directional epiderivatives, they are also closed sets.

In the following, we discuss an important special case. Let $K \subset \mathbb{R}^n$ be a convex, nonempty, and closed set and let us consider the indicator function $\varphi(x) := \iota_K(x)$. Then, due to Example 2.5.9, it holds

$$\varphi^\downarrow(x;h) = \iota_{T_K(x)}(h), \quad \forall\, x \in K, \quad \forall\, h \in \mathbb{R}^n.$$

Hence, for any $x \in G^{-1}(\mathrm{dom}\ \varphi) = G^{-1}(K)$, we obtain

$$\mathcal{T}^{\Lambda}(x) = T_K(G(x)) \cap (-T_K(\mathcal{P}_K^{\Lambda}(u(x)))$$

and the critical cone reduces to

$$\mathcal{C}(x) = \{h \in T_K(G(x)) : \langle F(x), h \rangle \le 0\} = T_K(G(x)) \cap \{F(x)\}^{\circ} \subset \{F(x)\}^{\circ}.$$

Moreover, if the function G is the identity mapping, then it follows

$$\mathcal{T}^{\Lambda}(x) = T_K(x) \cap (-T_K(\mathcal{P}_K^{\Lambda}(x - \Lambda^{-1}F(x))))$$

and the cones $\mathcal{T}^{\Lambda}(x)$ and $\mathcal{T}^{\Lambda}(x) \cap \mathcal{C}(x)$ coincide with the cones "$T_c(x; K)$" and "$T_c(x; K, F)$" that were introduced and discussed by Facchinei and Pang in [76, Section 10.2.1]. Let us also note that Robinson's constraint qualification is automatically fulfilled when G is the identity mapping. In what follows, we will show that the cones $\mathcal{T}^{\Lambda}(x)$ and $\mathcal{T}^{\Lambda}(x) \cap \mathcal{C}(x)$ represent the correct generalizations of the objects "$T_c(x; K)$" and "$T_c(x; K, F)$". This will then allow us to extend the variational optimality result in [76] to the generalized variational setting.

Before we present the full theorem, let us assume that $\bar{x}$ is a solution of the generalized variational inequality ($\mathcal{P}_{\mathrm{vip}}$) and that Robinson's constraint qualification is satisfied. Then, it holds $V^{\Lambda}(\bar{x}) = 0$ and thus, we can infer $G(\bar{x}) = \mathrm{prox}_{\varphi}^{\Lambda}(u(\bar{x}))$. Therefore, the cone $\mathcal{T}^{\Lambda}(\bar{x})$ reduces to the following set

$$\mathcal{T}^{\Lambda}(\bar{x}) = \{h \in \mathbb{R}^n : \varphi^{\downarrow}(G(\bar{x}); h) + \varphi^{\downarrow}(G(\bar{x}); -h) = 0\} = \mathrm{lin}\ \varphi^{\downarrow}(G(\bar{x}); \cdot),$$

where we used the subadditivity of the epiderivative $\varphi^{\downarrow}(G(\bar{x}); \cdot)$ and (6.2.3). Moreover, as we have already seen, the critical cone $\mathcal{C}(\bar{x})$ admits the following representation:

$$\mathcal{C}(\bar{x}) = \{h \in \mathbb{R}^n : \langle F(\bar{x}), h \rangle + \varphi^{\downarrow}(G(\bar{x}); h) = 0\} = N_{\partial\varphi(G(\bar{x}))}(-F(\bar{x})) = \mathcal{N}(\bar{x}).$$

Consequently, if $\bar{x}$ solves the generalized variational inequality ($\mathcal{P}_{\mathrm{vip}}$), then the sets $\mathcal{T}^{\Lambda}(\bar{x})$ and $\mathcal{T}^{\Lambda}(\bar{x}) \cap \mathcal{C}(\bar{x})$ are equal to the lineality space of the normal cone $N_{\partial\varphi(G(\bar{x}))}(-F(\bar{x}))$ and the cone $\mathcal{N}(\bar{x})$ coincides with the critical cone; let us refer to the discussion in section 5.1.3 for further details.

In the following result and similar to [76, Theorem 10.2.5], we need the cones $\mathcal{T}^{\Lambda}(\bar{x})$, $\mathcal{T}^{\Lambda}(\bar{x}) \cap \mathcal{C}(\bar{x})$, and $\mathcal{N}(\bar{x})$ for a stationary point $\bar{x}$ of (6.2.1) that is not (yet) known to be a solution of the problem ($\mathcal{P}_{\mathrm{vip}}$).

Theorem 6.2.4. *Let the mappings F, $G : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable on the open set $\Omega \subset \mathbb{R}^n$ and suppose that Ω contains the domain $G^{-1}(\mathrm{dom}\ \varphi)$. Let $\Lambda \in \mathbb{S}^n_{++}$ be an arbitrary parameter matrix and let $\bar{x} \in G^{-1}(\mathrm{dom}\ \varphi)$ be a stationary point of the regularized gap function G^{Λ}. Furthermore, let us assume that Robinson's constraint qualification is satisfied at $\bar{x}$. Then, the following three statements are equivalent:*

(i) *$\bar{x}$ solves the generalized variational inequality ($\mathcal{P}_{\mathrm{vip}}$).*

(ii) *The cone $\mathcal{T}^{\Lambda}(\bar{x}) \cap \mathcal{C}(\bar{x})$ is contained in $\mathcal{N}(\bar{x})$.*

(iii) *The implication below holds:*

$$\left.\begin{aligned} d \in \mathcal{T}^\Lambda(\bar{x}) \cap \mathcal{C}(\bar{x}) \\ DF(\bar{x})^\top d \in [DG(\bar{x})^{-1}\mathcal{T}^\Lambda(\bar{x})]^\circ \end{aligned}\right\} \quad \Longrightarrow \quad d \in \mathcal{N}(\bar{x}). \tag{6.2.4}$$

Proof. We proceed as in [76, Theorem 10.2.5]. If $\bar{x}$ is a solution of the variational inequality $(\mathcal{P}_{\text{vip}})$, then we have already shown, that the cone $\mathcal{T}^\Lambda(\bar{x})$ coincides with the linear subspace lin $N_{\partial\varphi(G(\bar{x}))}(-F(\bar{x}))$. Consequently, statement (i) implies (ii). Since the implication "(ii) ⇒ (iii)" is obvious, it remains to be shown that part (iii) implies (i). We start with a quite easy observation. Due to (2.5.1), we have

$$\begin{aligned} &\varphi^\downarrow(G(\bar{x}); -V^\Lambda(\bar{x})) + \varphi^\downarrow(\text{prox}_\varphi^\Lambda(u(\bar{x})); V^\Lambda(\bar{x})) \\ &\qquad \le \varphi(G(\bar{x}) - V^\Lambda(\bar{x})) - \varphi(G(\bar{x})) + \varphi(\text{prox}_\varphi^\Lambda(u(\bar{x})) + V^\Lambda(\bar{x})) - \varphi(\text{prox}_\varphi^\Lambda(u(\bar{x}))) \\ &\qquad = \varphi(\text{prox}_\varphi^\Lambda(u(\bar{x}))) - \varphi(G(\bar{x})) + \varphi(G(\bar{x})) - \varphi(\text{prox}_\varphi^\Lambda(u(\bar{x}))) = 0. \end{aligned}$$

Thus, it follows $d := -V^\Lambda(\bar{x}) \in \mathcal{T}^\Lambda(\bar{x})$. Now, by using the optimality conditions (6.2.2) and setting $h = V^\Lambda(\bar{x})$, we obtain

$$\varphi^\downarrow(\text{prox}_\varphi^\Lambda(u(\bar{x})); V^\Lambda(\bar{x})) + \langle F(\bar{x}), V^\Lambda(\bar{x})\rangle \ge \|V^\Lambda(\bar{x})\|_\Lambda^2 \tag{6.2.5}$$

and hence, we readily establish

$$\varphi^\downarrow(G(\bar{x}); d) + \langle F(\bar{x}), d\rangle \le -\varphi^\downarrow(\text{prox}_\varphi^\Lambda(u(\bar{x})); -d) + \langle F(\bar{x}), d\rangle \le 0.$$

This shows that the vector d must be an element of the cone $\mathcal{T}^\Lambda(\bar{x}) \cap \mathcal{C}(\bar{x})$. Next, we will verify the second condition in (6.2.4). Again, by using (6.2.2), we get

$$\varphi^\downarrow(\text{prox}_\varphi^\Lambda(u(\bar{x})); -DG(\bar{x})h) - \langle \Lambda V^\Lambda(\bar{x}) - F(\bar{x}), -DG(\bar{x})h\rangle \ge 0$$

for all $h \in \mathbb{R}^n$. Now, adding the latter inequality and the stationarity condition

$$(G^\Lambda)^\downarrow(\bar{x}; h) = \varphi^\downarrow(G(\bar{x}); DG(\bar{x})h) + \langle V^\Lambda(\bar{x}), DF(\bar{x})h\rangle + \langle F(\bar{x}) - \Lambda V^\Lambda(\bar{x}), DG(\bar{x})h\rangle \ge 0,$$

we infer

$$\langle -V^\Lambda(\bar{x}), DF(\bar{x})h\rangle \le \varphi^\downarrow(G(\bar{x}); DG(\bar{x})h) + \varphi^\downarrow(\text{prox}_\varphi^\Lambda(u(\bar{x})); -DG(\bar{x})h), \quad \forall\ h \in \mathbb{R}^n.$$

Thus, for all $h \in DG(\bar{x})^{-1}\mathcal{T}^\Lambda(\bar{x})$, it follows

$$\langle DF(\bar{x})^\top d, h\rangle = \langle -V^\Lambda(\bar{x}), DF(\bar{x})h\rangle \le 0$$

and we can conclude $DF(\bar{x})^\top d \in [DG(\bar{x})^{-1}\mathcal{T}^\Lambda(\bar{x})]^\circ$. Consequently, by (6.2.4), we deduce $d \in \mathcal{N}(\bar{x})$ and hence, using (6.2.5), this implies

$$\|V^\Lambda(\bar{x})\|_\Lambda^2 \le \varphi^\downarrow(G(\bar{x}); d) + \varphi^\downarrow(\text{prox}_\varphi^\Lambda(u(\bar{x})); -d) \le 0.$$

This shows that $\bar{x}$ is a solution of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$ and finishes the proof of Theorem 6.2.4. □

In the following, we briefly consider the special case of a mixed variational inequality for which an easy characterization of the conditions in Theorem 6.2.4 is available. Therefore, let us suppose that the function G is the identity mapping and let $\bar{x} \in \mathrm{dom}\ \varphi$ be a stationary point of the regularized gap function. Then, our argumentation in the proof of Theorem 6.2.4 showed

$$(6.2.6) \qquad -V^{\Lambda}(\bar{x}) \in \mathcal{T}^{\Lambda}(\bar{x}) \cap \mathcal{C}(\bar{x}), \quad -DF(\bar{x})^{\top} V^{\Lambda}(\bar{x}) \in \mathcal{T}^{\Lambda}(\bar{x})^{\circ} \subset [\mathcal{T}^{\Lambda}(\bar{x}) \cap \mathcal{C}(\bar{x})]^{\circ}.$$

Consequently, if the matrix $DF(\bar{x})$ is *strictly copositive* on the cone $\mathcal{T}^{\Lambda}(\bar{x}) \cap \mathcal{C}(\bar{x})$, i.e., if it holds

$$\langle h, DF(\bar{x})h \rangle > 0, \quad \forall\ h \in \mathcal{T}^{\Lambda}(\bar{x}) \cap \mathcal{C}(\bar{x}) \setminus \{0\},$$

then (6.2.6) immediately implies that $\bar{x}$ is a solution of the generalized variational inequality $(\mathcal{P}_{\mathrm{vip}})$. Moreover, if $DF(\bar{x})$ is strictly copositive on the critical cone $\mathcal{C}(\bar{x})$, then by Remark 6.1.7, we can even infer that $\bar{x}$ is an isolated solution of the problem $(\mathcal{P}_{\mathrm{vip}})$. This observation is analogous to the corresponding result for classical variational inequalities, see [76, Corollary 10.2.7].

6.2.2. The D-Gap function

The D-gap function was first introduced by Peng [188] and Yamashita et al. [257] and is defined as the *difference* of two regularized gap functions. This elegant approach formally eliminates the nonsmooth term "$\varphi(G(x))$" and thus resolves several disadvantages of the regularized gap function. In particular, Peng and Yamashita et al. showed that the D-gap function is a merit function and that the corresponding *D-gap program* $(\mathcal{P}_{\mathrm{mer}})$ allows to reformulate the classical variational inequality as an *unconstrained* optimization problem. In the last decade, the D-gap function has been analyzed by various authors and in different contexts [227, 257, 121, 120, 224, 76, 223, 245, 246]. In the following, we will consider an extended version of the D-gap function for generalized variational inequalities and discuss its properties.

First, we give a precise definition of the D-gap function that is again due to Solodov [223].

Definition 6.2.5. *Let $A, B \in \mathbb{S}^n_{++}$, $B \succ A$, be arbitrary parameter matrices and let $F, G : \mathbb{R}^n \to \mathbb{R}^n$ be given functions. The* D-gap function *of the generalized variational inequality problem $(\mathcal{P}_{\mathrm{vip}})$ is defined as*

$$H^{A,B} : \mathbb{R}^n \to \mathbb{R}, \quad H^{A,B}(x) := G^A(x) - G^B(x)$$

for all $x \in \mathbb{R}^n$.

Using Lemma 6.2.2, we immediately obtain the following alternative representation of the D-gap function

$$H^{A,B}(x) = \frac{1}{2}\|F(x)\|^2_{A^{-1}-B^{-1}} + \mathrm{env}^B_{\varphi}(G(x) - B^{-1}F(x)) - \mathrm{env}^A_{\varphi}(G(x) - A^{-1}F(x)).$$

Thus, by applying Lemma 3.1.5, we deduce the next differentiability result.

Lemma 6.2.6. *Let $A, B \in \mathbb{S}^n_{++}$, $B \succ A$, be arbitrary parameter matrices and suppose that the mappings $F, G : \mathbb{R}^n \to \mathbb{R}^n$ are continuously differentiable on $\mathbb{R}^n$. Then, the D-gap function $H^{A,B}$ is also continuously differentiable on $\mathbb{R}^n$ and it holds*

$$\nabla H^{A,B}(x) = DG(x)^\top \left(BV^B(x) - AV^A(x)\right) - DF(x)^\top (V^B(x) - V^A(x)).$$

Now, similar to Solodov [223, Theorem 6], we derive two important growth conditions for the D-gap function. In particular, the estimates below will show that the D-gap function is in fact a merit function for the problem $(\mathcal{P}_{\text{vip}})$.

Lemma 6.2.7. *Let $A, B \in \mathbb{S}^n_{++}$, $B \succ A$, be arbitrary matrices and let $F, G : \mathbb{R}^n \to \mathbb{R}^n$ be given functions. Then, it holds*

$$\frac{1}{2}\|V^B(x)\|^2_{B-A} \le H^{A,B}(x) \le \frac{1}{2}\|V^A(x)\|^2_{B-A}, \quad \forall\, x \in \mathbb{R}^n.$$

Hence, $H^{A,B}$ is nonnegative on $\mathbb{R}^n$ and a vector $x \in \mathbb{R}^n$ is a solution of the generalized variational inequality problem $(\mathcal{P}_{\text{vip}})$ if and only if $H^{A,B}(x) = 0$.

Proof. Setting $p_a := \text{prox}^A_\varphi(G(x) - A^{-1}F(x))$ and $p_b := \text{prox}^B_\varphi(G(x) - B^{-1}F(x))$ and using the optimality principle (6.2.2), we readily get

$$\begin{aligned} H^{A,B}(x) &= \frac{1}{2}\|F(x)\|^2_{A^{-1}-B^{-1}} + \frac{1}{2}\|V^B(x) - B^{-1}F(x)\|^2_B - \frac{1}{2}\|V^A(x) - A^{-1}F(x)\|^2_A \\ &\quad + \varphi(p_b) - \varphi(p_a) \\ &\ge \frac{1}{2}\|V^B(x)\|^2_B - \langle F(x), V^B(x) - V^A(x)\rangle - \frac{1}{2}\|V^A(x)\|^2_A \\ &\quad + \langle AV^A(x) - F(x), V^A(x) - V^B(x)\rangle \\ &= \frac{1}{2}\|V^B(x)\|^2_B - \langle V^B(x), V^A(x)\rangle_A + \frac{1}{2}\|V^A(x)\|^2_A \ge \frac{1}{2}\|V^B(x)\|^2_{B-A}. \end{aligned}$$

On the other hand, due to

$$\varphi(p_b) - \varphi(p_a) \le \langle BV^B(x) - F(x), V^A(x) - V^B(x)\rangle,$$

the upper estimate can be established in a similar fashion. The last claim obviously follows from Lemma 6.1.1 (iv). □

Next, we show that the norm of the natural residual $\|V^A(x)\|$ does not grow to much with respect to the parameter matrix A. This result is a straightforward extension of Lemma 4.1.3 where we considered the case $F \equiv \nabla f$ and $G \equiv I$.

Lemma 6.2.8. *Let $F, G : \mathbb{R}^n \to \mathbb{R}^n$ be given functions and let $A, B \in \mathbb{S}^n_{++}$ be two arbitrary symmetric and positive definite matrices. Then, for all $x \in \mathbb{R}^n$ and for $W := B^{-\frac{1}{2}}AB^{-\frac{1}{2}}$, it follows*

$$\|V^A(x)\| \le \frac{1 + \lambda_{\max}(W) + \sqrt{1 - 2\lambda_{\min}(W) + \lambda_{\max}(W)^2}}{2} \frac{\lambda_{\max}(B)}{\lambda_{\min}(A)}\|V^B(x)\|.$$

Proof. The proof is almost identical to the proof presented in [236]. More precisely, by using the optimality principle (6.2.2) separately for $\mathrm{prox}_\varphi^A(G(x) - A^{-1}F(x))$ and $\mathrm{prox}_\varphi^B(G(x) - B^{-1}F(x))$ and by adding the resulting inequalities, we readily obtain

$$\langle BV^B(x) - AV^A(x), V^B(x) - V^A(x)\rangle \leq 0.$$

(The same inequalities were also used in the proof of the last lemma). From this point, we can proceed as in [236, Lemma 3]. □

Remark 6.2.9. Let $A, B \in \mathbb{S}^n_{++}$, $B \succ A$, be given and let $(A_k)_k, (B_k)_k \subset \mathbb{S}^n_{++}$ be two families of symmetric, positive definite matrices. Suppose that there exist matrices $\Lambda^b_M \succeq \Lambda^b_m \succ \Lambda^a_M \succeq \Lambda^a_m \succ 0$ such that

$$\Lambda^b_M \succeq B_k \succeq \Lambda^b_m, \quad \Lambda^a_M \succeq A_k \succeq \Lambda^a_m, \quad \forall\ k \in \mathbb{N}.$$

Then, similar to Remark 4.1.4 and by combining Lemma 6.2.7 and 6.2.8, it is possible to derive the bounds

$$\underline{\lambda} \cdot \|V^A(x)\| \leq \|V^{A_k}(x)\| \leq \overline{\lambda} \cdot \|V^A(x)\|$$

and

$$\underline{\lambda} \cdot H^{A,B}(x) \leq H^{A_k,B_k}(x) \leq \overline{\lambda} \cdot H^{A,B}(x)$$

for all $k \in \mathbb{N}$, $x \in \mathbb{R}^n$ and some constants $\underline{\lambda}, \overline{\lambda} > 0$ which do not depend on k, A_k or B_k. Hence, if the parameter matrices $(A_k)_k, (B_k)_k$ remain in bounded (and separated) sets, then the latter inequalities imply

$$H^{A_k,B_k}(x^k) \to 0 \iff H^{A,B}(x^k) \to 0, \quad \text{and} \quad V^{A_k}(x^k) \to 0 \iff V^A(x^k) \to 0,$$

as $k \to \infty$. Again, if the functions $H^{A,B}$ or V^A are used within an iterative procedure, this shows that the parameter matrices A and B are allowed to change in each iteration.

Similar to the previous section, we will now derive several equivalent conditions that guarantee *variational optimality* of a stationary point $\bar{x} \in G^{-1}(\mathrm{dom}\ \varphi)$ of the D-gap function. (A stationary point of the D-gap function will again be called *variational optimal* if it is a solution of the problem $(\mathcal{P}_{\mathrm{vip}})$). Again, our methodology is strongly motivated by related results of Facchinei and Pang that were established for classical variational inequalities, see [76, Section 10.3]. Now, let A, $B \in \mathbb{R}^{n\times n}$ be two arbitrary parameter matrices satisfying

$$A := \alpha^{-1}I, \quad B := \beta^{-1}I, \quad \alpha > \beta > 0$$

and let $x \in G^{-1}(\mathrm{dom}\ \varphi)$ be given. Let us mention that the following definitions and results can also be formulated for more general parameter matrices $A, B \in \mathbb{S}^n_{++}$. However, since in the fully general setting, the intuition behind the subsequent objects becomes less clear, we decided to present a "streamlined" version for the simpler, one-dimensional parametrizations. Thus, for $u_a(x) := G(x) - A^{-1}F(x)$ and $u_b(x) := G(x) - B^{-1}F(x)$, let us define

$$\Pi_a(h) := \varphi^\downarrow(\mathrm{prox}_\varphi^A(u_a(x)); h), \quad \Pi_b(h) := \varphi^\downarrow(\mathrm{prox}_\varphi^B(u_b(x)); h).$$

Moreover, let us consider the following sets

$$\mathcal{T}^{A,B}(x) := \{h \in \mathbb{R}^n : \Pi_b(h) + \Pi_a(-h) \leq 0\},$$

$$\mathcal{C}^{A,B}(x) := \{h \in \mathbb{R}^n : \langle F(x), h\rangle \leq \Pi_b((B-A)^{-1}Ah) + \Pi_a((A-B)^{-1}Bh)\},$$

and

$$\mathcal{N}^{A,B}(x) := \{h \in \mathbb{R}^n : \langle F(x), h\rangle = \Pi_b((B-A)^{-1}Ah) + \Pi_a((A-B)^{-1}Bh)\}.$$

Again, as in the last section, the properties of the epiderivatives Π_a and Π_b imply that the sets $\mathcal{T}^{A,B}(x)$, $\mathcal{C}^{A,B}(x)$, and $\mathcal{N}^{A,B}(x)$ are convex, nonempty, and closed cones. Furthermore, if $\bar{x} \in G^{-1}(\text{dom } \varphi)$ is a solution of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$, then it holds

$$G(\bar{x}) = \text{prox}_\varphi^A(u_a(\bar{x})), \quad G(\bar{x}) = \text{prox}_\varphi^B(u_b(\bar{x}))$$

and the cone $\mathcal{T}^{A,B}(\bar{x})$ coincides with the lineality space of the normal cone $N_{\partial\varphi(G(\bar{x}))}(-F(\bar{x}))$. Now, by using the positive homogeneity of the epiderivative $\varphi^{\downarrow}(G(\bar{x}); \cdot)$, we obtain

$$\Pi_b(B-A)^{-1}Ah) + \Pi_a((A-B)^{-1}Bh) = \frac{\beta}{\alpha-\beta}\varphi^{\downarrow}(G(\bar{x}); h) + \frac{\alpha}{\alpha-\beta}\varphi^{\downarrow}(G(\bar{x}); -h).$$

Hence, in this situation, it holds

$$\text{lin } \varphi^{\downarrow}(G(\bar{x}); \cdot) = \mathcal{T}^{A,B}(\bar{x}) \subset \mathcal{T}^{A,B}(\bar{x}) \cap \mathcal{C}^{A,B}(\bar{x}) \subset \mathcal{N}^{A,B}(\bar{x}).$$

Before stating the main result of this section, let us note that the cones $\mathcal{C}^{A,B}(\bar{x})$ and $\mathcal{N}^{A,B}(\bar{x})$ can also be simplified as follows:

$$\begin{aligned} \mathcal{C}^{A,B}(x) &:= \{h \in \mathbb{R}^n : (\alpha-\beta)\langle F(x), h\rangle \leq \alpha\Pi_a(-h) + \beta\Pi_b(h)\}, \\ \mathcal{N}^{A,B}(x) &:= \{h \in \mathbb{R}^n : (\alpha-\beta)\langle F(x), h\rangle = \alpha\Pi_a(-h) + \beta\Pi_b(h)\}. \end{aligned}$$

Theorem 6.2.10. *Let the mappings F, $G : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable and let $A = \alpha^{-1}I$, $B = \beta^{-1}I$ be arbitrary parameter matrices with $\alpha > \beta > 0$. Furthermore, let $\bar{x} \in G^{-1}(\text{dom } \varphi)$ be a stationary point of the D-gap function $H^{A,B}$ and suppose that the matrix $DG(\bar{x})$ is invertible. Then, the following three statements are equivalent:*

(i) *$\bar{x}$ solves the generalized variational problem $(\mathcal{P}_{\text{vip}})$.*

(ii) *The set $\mathcal{T}^{A,B}(\bar{x}) \cap \mathcal{C}^{A,B}(\bar{x})$ is contained in $\mathcal{N}^{A,B}(\bar{x})$.*

(iii) *The implication below holds:*

$$\left.\begin{array}{r} d \in \mathcal{T}^{A,B}(\bar{x}) \cap \mathcal{C}^{A,B}(\bar{x}) \\ DF(\bar{x})^\top d \in DG(\bar{x})^\top [\mathcal{T}^{A,B}(\bar{x})]^\circ \end{array}\right\} \quad \Longrightarrow \quad d \in \mathcal{N}^{A,B}(\bar{x}). \tag{6.2.7}$$

Proof. Since the first implication "(i) ⇒ (ii)" follows from our preceding discussion and the second implication "(ii) ⇒ (iii)" is rather obvious, we directly start with the verification of the remaining direction "(iii) ⇒ (i)". Let us define $d := V^B(\bar{x}) - V^A(\bar{x}) = \mathrm{prox}_\varphi^A(u_a(\bar{x})) - \mathrm{prox}_\varphi^B(u_b(\bar{x}))$. Then, by using the optimality principle (6.2.2), it follows

$$\begin{aligned}\Pi_b(d) + \Pi_a(-d) \le \varphi(\mathrm{prox}_\varphi^B(u_b(\bar{x})) + d) - \varphi(\mathrm{prox}_\varphi^B(u_b(\bar{x})))\\ + \varphi(\mathrm{prox}_\varphi^A(u_a(\bar{x})) - d) - \varphi(\mathrm{prox}_\varphi^A(u_a(\bar{x}))) = 0.\end{aligned}$$

This obviously implies $d \in \mathcal{T}^{A,B}(\bar{x})$. Moreover, using the second inequality in (6.2.2), we readily establish the following estimates:

- $\Pi_b((B-A)^{-1}Ad) \ge \langle BV^B(\bar{x}) - F(\bar{x}), (B-A)^{-1}A[V^B(\bar{x}) - V^A(\bar{x})]\rangle$,
- $\Pi_a((A-B)^{-1}Bd) \ge \langle F(\bar{x}) - AV^A(\bar{x}), (B-A)^{-1}B[V^B(\bar{x}) - V^A(\bar{x})]\rangle$.

Now, due to $(B-A)^{-1}B = I + (B-A)^{-1}A$ and by adding the latter inequalities, we obtain

$$\begin{aligned}(6.2.8)\qquad & \Pi_b((B-A)^{-1}Ad) + \Pi_a((A-B)^{-1}Bd)\\ &\ge \langle BV^B(\bar{x}) - AV^A(\bar{x}) - (B-A)V^A(\bar{x}), (B-A)^{-1}Ad\rangle + \langle F(\bar{x}), d\rangle\\ &= \|d\|^2_{A(B-A)^{-1}B} + \langle F(\bar{x}), d\rangle\end{aligned}$$

and hence, it holds $d \in \mathcal{C}^{A,B}(\bar{x})$. Similarly, the optimality principle (6.2.2) also implies

$$\Pi_b(DG(\bar{x})h) \ge \langle BV^B(\bar{x}) - F(\bar{x}), DG(\bar{x})h\rangle, \quad \Pi_a(-DG(\bar{x})h) \ge \langle F(\bar{x}) - AV^A(\bar{x}), DG(\bar{x})h\rangle$$

for all $h \in \mathbb{R}^n$. Next, by summing the latter inequalities, we get

$$\langle DG(\bar{x})^\top [BV^B(\bar{x}) - AV^A(\bar{x})], h\rangle \le \Pi_b(DG(\bar{x})h) + \Pi_a(-DG(\bar{x})h) \le 0$$

for all $h \in \mathbb{R}^n$ with $DG(\bar{x})h \in \mathcal{T}^{A,B}(\bar{x})$. Hence, this shows $DG(\bar{x})^\top[BV^B(\bar{x}) - AV^A(\bar{x})] \in [DG(\bar{x})^{-1}\mathcal{T}^{A,B}(\bar{x})]^\circ$. Now, due to the invertibility of the matrix $DG(\bar{x})$, a general, computational result of Bonnans and Shapiro, [27, Lemma 3.27], is applicable and it follows

$$[DG(\bar{x})^{-1}\mathcal{T}^{A,B}(\bar{x})]^\circ = DG(\bar{x})^\top[\mathcal{T}^{A,B}(\bar{x})]^\circ.$$

Consequently, since the stationarity of $\bar{x}$ implies

$$(6.2.9)\qquad \nabla H^{A,B}(\bar{x}) = DG(\bar{x})^\top[BV^B(\bar{x}) - AV^A(\bar{x})] - DF(\bar{x})^\top d = 0,$$

we immediately establish $DF(\bar{x})^\top d \in DG(\bar{x})^\top[\mathcal{T}^{A,B}(\bar{x})]^\circ$ and from (6.2.7) and (6.2.8) we deduce $d = 0$. Finally, by combining the invertibility of the matrix $DG(\bar{x})$, $V^A(\bar{x}) = V^B(\bar{x})$, and (6.2.9), we obtain

$$V^A(\bar{x}) = V^B(\bar{x}) = 0.$$

This completes the proof of Theorem 6.2.10. □

Remark 6.2.11. A careful examination of the proof of Theorem 6.2.10 shows that the implication "(iii) ⇒ (i)" does also hold for more general parameter matrices $A, B \in \mathbb{S}^n_{++}$. However, in this case, the connection between the sets $\mathcal{T}^{A,B}(\bar{x}) \cap \mathcal{C}^{A,B}(\bar{x})$ and $\mathcal{N}^{A,B}(\bar{x})$ is not clear and thus, full equivalence as in Theorem 6.2.10 cannot be directly inferred.

Now, let $A, B \in \mathbb{S}^n_{++}$, $B \succ A$, be given parameter matrices and suppose that the function G is the identity mapping. Furthermore, let $\bar{x} \in \operatorname{dom} \varphi$ be an arbitrary stationary point of the D-gap function and let us set $d := V^B(\bar{x}) - V^A(\bar{x})$. Then, the proof of Theorem 6.2.10 implies

$$d \in \mathcal{T}^{A,B}(\bar{x}) \cap C^{A,B}(\bar{x}), \quad DF(\bar{x})^\top d \in \mathcal{T}^{A,B}(\bar{x})^\circ \subset [\mathcal{T}^{A,B}(\bar{x}) \cap C^{A,B}(\bar{x})]^\circ.$$

Consequently, similar to our observations in the previous subsection, if the matrix $DF(\bar{x})$ is strictly copositive on the cone $\mathcal{T}^{A,B}(\bar{x}) \cap \mathcal{C}^{A,B}(\bar{x})$, then it follows $d = 0$ and thus, by (6.2.9), $\bar{x}$ must be a solution of the generalized variational inequality ($\mathcal{P}_{\text{vip}}$).

The next result concludes this subsection and establishes a sufficient condition for coercivity of the natural residual mapping and of the D-gap function. Lemma 6.2.12 is an extension of [76, Proposition 10.3.9] and is based on a monotonocity-type assumption (as in, e.g., Lemma 6.1.4).

Lemma 6.2.12. *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ and $G : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz continuous with moduli L_F and L_G, respectively. Moreover, suppose that there exist $x^* \in G^{-1}(\operatorname{dom} \varphi)$ and constants $\vartheta > 0$, $\xi > 1$ such that φ is subdifferentiable at $G(x^*)$ and it holds*

$$\lim_{\|x\| \to \infty} \frac{\langle F(x) - F(x^*), G(x) - G(x^*) \rangle}{\|x - x^*\|^\xi} \geq \vartheta. \tag{6.2.10}$$

Then, the functions $\|V^A\|$ and $H^{A,B}$ are coercive on $\mathbb{R}^n$ for every arbitrary choice of $A, B \in \mathbb{S}^n_{++}$ with $B \succ A$.

Proof. Let us write $\vartheta^* := \|F(x^*)\|$ and $\operatorname{prox}^A_\varphi(G(x) - A^{-1}F(x)) = G(x) - V^A(x)$. Then, by using (6.2.2) and the Lipschitz continuity of the functions F and G, it holds

$$\begin{aligned}
&\varphi(\operatorname{prox}^A_\varphi(G(x) - A^{-1}F(x))) - \varphi(G(x^*)) \\
&\qquad \leq \langle AV^A(x) - F(x), G(x) - V^A(x) - G(x^*) \rangle \\
&\qquad = \langle F(x^*) - F(x), G(x) - G(x^*) \rangle - \langle AV^A(x) - F(x), V^A(x) \rangle \\
&\qquad\qquad + \langle AV^A(x) - F(x^*), G(x) - G(x^*) \rangle \\
&\qquad \leq \langle F(x^*) - F(x), G(x) - G(x^*) \rangle - \|V^A(x)\|^2_A \\
&\qquad\qquad + ((L_F + L_G\|A\|)\|x - x^*\| + \vartheta^*)\|V^A(x)\| + L_G\vartheta^*\|x - x^*\|.
\end{aligned}$$

Now, since φ is subdifferentiable at $G(x^*)$, it follows

$$\begin{aligned}
\varphi(\operatorname{prox}^A_\varphi(G(x) - A^{-1}F(x))) - \varphi(G(x^*)) &\geq \langle \lambda^*, G(x) - G(x^*) - V^A(x) \rangle \\
&\geq -L_G\|\lambda^*\|\|x - x^*\| - \|\lambda^*\|\|F^A(x)\|,
\end{aligned}$$

where $\lambda^* \in \partial\varphi(G(x^*))$ is an arbitrary subgradient. Next, by combining the last inequalities and setting $C(x) := (L_F + L_G\|A\|)\|x - x^*\| + \vartheta^* + \|\lambda^*\|$, we obtain

$$\|V^A(x)\| \geq \frac{\|x - x^*\|^\xi}{C(x)} \left\{ \frac{\langle F(x) - F(x^*), G(x) - G(x^*) \rangle}{\|x - x^*\|^\xi} - \frac{L_G(\vartheta^* + \|\lambda^*\|)}{\|x - x^*\|^{\xi - 1}} \right\}.$$

Clearly, due to $\xi > 1$ and (6.2.10), the term within the curly brackets is strictly positive as $\|x\|$ tends to infinity, while the outer factor $\|x - x^*\|^\xi / C(x)$ diverges to $+\infty$. Using Lemma 6.2.7, this establishes the coercivity of $\|V^A\|$ and $H^{A,B}$. □

As mentioned in Remark 4.3.4, this result can be utilized to ensure boundedness of the iterates of the semismooth Newton method that was discussed in chapter 4. Let us finally note that condition (6.2.10) is obviously satisfied if the function F is (ξ, G)-monotone on $\mathbb{R}^n$.

6.3. Numerical algorithms for GVIPs

In this section, we propose and analyze different numerical algorithms for the solution of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$. Our investigation focuses on a globalized semismooth Newton method that uses semismooth Newton steps for the nonsmooth equation

$$V^\Lambda(x) = G(x) - \text{prox}_\varphi^\Lambda(G(x) - \Lambda^{-1}F(x)) = 0, \quad \Lambda \in \mathbb{S}^n_{++},$$

to augment a merit function-based descent method for the problem $(\mathcal{P}_{\text{vip}})$. Similar to Algorithm 2, we utilize the abstract multidimensional filter mechanism presented in section 4.2.3 to connect these two different algorithmic components.

In the following, as a consequence of our preceding discussions, we will only consider D-gap and regularized gap function-based approaches. In particular, since the D-gap function was shown to be continuously differentiable, a simple gradient descent method can be chosen as an underlying base algorithm to globalize the semismooth Newton method. Moreover, if the function G is the identity mapping, then the regularized gap program,

$$\min_x \; G^\Lambda(x) = \left\{ \tfrac{1}{2}\|F(x)\|^2_{\Lambda^{-1}} - \text{env}_\varphi^\Lambda(x - \Lambda^{-1}F(x)) \right\} + \varphi(x) \quad \text{s.t.} \quad x \in \text{dom } \varphi,$$

reduces to an optimization problem of the form $(\mathcal{P})$ and thus, a specialized version of the proximal gradient descent method represents another suitable base algorithm in this case. At this point, let us clarify that an algorithm for the generalized variational inequality $(\mathcal{P}_{\text{vip}})$ is said to be *globally convergent* if and only if every accumulation point of a generated sequence of iterates is also a stationary point of an appropriate and associated merit function. Since stationary points of the regularized gap function do not necessarily correspond to stationary points of the D-gap function and vice versa, we want to emphasize that this terminology obviously depends on the chosen merit function.

In contrast to the D-gap function based approach, the minimization of the regularized gap function G^Λ may again result in a constrained minimization problem. Here, similar to Algorithm 2, feasibility of the Newton iterates has to be enforced in order to guarantee global and local convergence. Since one of our initial motivations for studying generalized variational inequalities was to circumvent this additional restriction, we will mainly concentrate on a D-gap function-based globalization that is also well-defined for infeasible input vectors from now on. Nevertheless, let us note that the stationarity results in Theorem 6.2.4, which were derived for the regularized gap function, have a much simpler and more natural form than the corresponding conditions for the D-gap function. Furthermore, the computation of the gap function G^Λ also only requires the evaluation of a single proximity operator while the

Algorithm 3: D-Gap Function-Based Gradient Descent Method

```
S0 Initialization: Choose $x^0 \in \mathbb{R}^n$, $B_0 \succ A_0 \succ 0$, $\beta, \gamma \in (0,1)$. Set iteration $k := 0$.
   while $\nabla H^{A_k,B_k}(x^k) \neq 0$ do
S1     Compute a new direction $d^k = -\nabla H^{A_k,B_k}(x^k)$.
S2     Choose a maximal Armijo stepsize $\sigma_k \in \{1, \beta, \beta^2, \beta^3, ...\} \subset (0,1]$ satisfying
           $H^{A_k,B_k}(x^k + \sigma_k d^k) \leq H^{A_k,B_k}(x^k) + \sigma_k \gamma \cdot \nabla H^{A_k,B_k}(x^k)^\top d^k.$
S3     Set $x^{k+1} = x^k + \sigma_k d^k$ and choose $A_{k+1}, B_{k+1} \in \mathbb{S}^n_{++}$ with $B_{k+1} \succ A_{k+1}$.
       $k \leftarrow k + 1$.
```

application of the D-gap function incorporates the calculation of two different proximity operators.

The D-gap function-based descent method is summarized in the next subsection. The proposed semismooth Newton method will be presented in subsection 6.3.2 in detail.

6.3.1. A D-gap function-based descent method

In the following, we consider a basic gradient descent method with an Armijo-type linesearch technique to solve the D-gap program

$$\min_x \; H^{A,B}(x), \quad A, B \in \mathbb{S}^n_{++}, \quad B \succ A.$$

Let us emphasize that our approach to minimize the D-gap function is well-known in nonlinear programming. Moreover, it can be seen as a special case of the proximal gradient method with $f \equiv H^{A,B}$, $\varphi \equiv 0$, and $\Lambda \equiv I$. The details are formulated in Algorithm 3. Again, the parameter matrices $A, B \in \mathbb{S}^n_{++}$ are allowed to change adaptively.

The following theorem is an immediate consequence of Theorem 4.2.10. For various related results and specific parameter strategies for A_k and B_k we refer to [257, 224] and [76, Section 10.4.1].

Theorem 6.3.1 (Global convergence). *Let the functions $F, G : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable and let the sequences $(x^k)_k$, $(A_k)_k$, and $(B_k)_k$ be generated by Algorithm 3. Furthermore, suppose that there exist $k^* \in \mathbb{N}$ and $A_*, B_* \in \mathbb{S}^n_{++}$, $B_* \succ A_*$, such that*

$$A_k = A_* \quad \text{and} \quad B_k = B_*, \quad \forall \; k \geq k^*.$$

Then, every accumulation point x^* of $(x^k)_k$ satisfies $\nabla H^{A_*,B_*}(x^*) = 0$ and is thus a stationary point of the D-gap function H^{A_*,B_*}.

6.3.2. A semismooth Newton method for generalized variational inequalities

The globalized semismooth Newton method we have described so far can be clearly seen as an adaption of Algorithm 2 for generalized variational inequalities. Similar to Algorithm 2,

```
Algorithm 4: Globalized Semismooth Newton Method for GVIPs
S0 Initialization: Choose an initial point $x^0 \in \mathbb{R}^n$, $\Lambda_0 \in \mathbb{S}^n_{++}$, $B_0 \succ A_0 \succ 0$, $\beta, \gamma, \gamma_{\mathcal{F}} \in (0,1)$,
   and $\mathcal{F}_{-1} = \emptyset$. Set iteration $k := 0$.
   while $\nabla H^{A_k,B_k}(x^k) \neq 0$ do
S1     If $k = 0$ or $x^k$ was obtained in step S3, add $\theta(x^k)$ to the filter: $\mathcal{F}_k = \mathcal{F}_{k-1} \cup \{\theta(x^k)\}$.
       Otherwise, set $\mathcal{F}_k = \mathcal{F}_{k-1}$. Choose $A_{k+1}, B_{k+1}, \Lambda_{k+1} \in \mathbb{S}^n_{++}$ with $B_{k+1} \succ A_{k+1}$.
S2     Compute the semismooth Newton step $s^k$ via $M(x^k)s^k = -V^{\Lambda_k}(x^k)$. If this is not
       possible go to step S4.
S3     Set $x^{k+1} = x^k + s^k$ and check if $x^{k+1}$ is acceptable for the filter $\mathcal{F}_k$:
           $\max_{1 \le j \le p} \left(q_j - \theta_j(x^{k+1})\right) \ge \gamma_{\mathcal{F}} \max_{1 \le j \le p} \theta_j(x^{k+1}), \quad \forall\, q \in \mathcal{F}_k.$
       If $x^{k+1}$ is acceptable for $\mathcal{F}_k$ skip step S4 and S5.
S4     Compute the direction $d^k = -\nabla H^{A_k,B_k}(x^k)$ and choose a maximal Armijo step
       $\sigma_k \in \{1, \beta, \beta^2, \beta^3, ...\} \subset (0,1]$ satisfying
           $H^{A_k,B_k}(x^k + \sigma_k d^k) \le H^{A_k,B_k}(x^k) + \sigma_k \gamma \cdot \nabla H^{A_k,B_k}(x^k)^\top d^k.$
S5     Set $x^{k+1} = x^k + \sigma_k d^k$.
       $k \leftarrow k + 1$.
```

to obtain a new Newton step, we have to solve the linear system of equations

$$M(x^k)s^k = -V^{\Lambda_k}(x^k),$$

where $M(x^k)$ is a generalized derivative of the natural residual V^{Λ_k} at x^k and Λ_k is the current parameter matrix. Again, the trial point $x^k + s^k$ is accepted as a new iterate if it is acceptable for the current filter $\mathcal{F}_k$, i.e., whenever the filter value $\theta(x^k+s^k) \equiv \theta(x^k+s^k, \Lambda_{k+1})$ fulfills the acceptance test (4.2.11). More specifically, if the trial point $x^k + s^k$ satisfies the filter conditions, then we set $x^{k+1} = x^k + s^k$, update the filter $\mathcal{F}_{k+1} = \mathcal{F}_k \cup \{\theta(x^k)\}$ and start the next iteration. Otherwise, the Newton step is rejected and we perform a step of the D-gap function-based descent method.

In contrast to Algorithm 2, the feasibility condition $G(x^k+s^k) \in \operatorname{dom} \varphi$ and the additional growth conditions (4.2.18) and (4.2.19) are not required to establish global convergence. The details of the method are summarized in Algorithm 4.

Next, we state several assumptions that are essential for our convergence analysis.

Assumption 6.3.2. *Let the sequences $(x^k)_k$, $(A_k)_k$, $(B_k)_k$, and $(\Lambda_k)_k$ be generated by Algorithm 4 and suppose that $x^* \in \mathbb{R}^n$ and $A_*, B_*, \Lambda_* \in \mathbb{S}^n_{++}$ are accumulation points of $(x^k)_k$, $(A_k)_k$, $(B_k)_k$, and $(\Lambda_k)_k$, respectively. Let us consider the following conditions:*

(F.1) *There exists $k^* \in \mathbb{N}$ such that $A_k = A_*$, $B_k = B_*$, and $\Lambda_k = \Lambda_*$ for all $k \ge k^*$.*

(F.2) *The proximity operator* $\operatorname{prox}_{\varphi}^{\Lambda_*} : \mathbb{R}^n \to \mathbb{R}^n$ *is semismooth at* $u^* := G(x^*) - \Lambda_*^{-1} F(x^*)$.

(F.3) *There exist constants* $k_* \in \mathbb{N}$ *and* $C > 0$ *such that for all* $k \geq k_*$*, every matrix* $M_k := M(x^k) \in \partial V^{\Lambda_k}(x^k)$ *is nonsingular with* $\|M_k^{-1}\| \leq C$.

If, in addition, the accumulation point x^* *is a solution of the generalized variational inequality* $(\mathcal{P}_{\text{vip}})$*, then we assume:*

(F.4) *The accumulation point* x^* *is an isolated solution of problem* $(\mathcal{P}_{\text{vip}})$.

In the following, we discuss global and local convergence properties of Algorithm 4 in detail. In particular, by utilizing the conditions presented in Assumption 6.3.2, our analysis allows to extend well-known convergence results for classical variational inequalities, see, e.g., [120, Theorem 3.5] or [76, Theorem 10.4.9].

Theorem 6.3.3. *Let the functions* $F, G : \mathbb{R}^n \to \mathbb{R}^n$ *be continuously differentiable and let the sequences* $(x^k)_k$, $(A_k)_k$, $(B_k)_k$*, and* $(\Lambda_k)_k$ *be generated by Algorithm 4. Furthermore, suppose that assumption* (F.1) *is satisfied. Then, it holds:*

(i) *Every accumulation point* x^* *of the sequence* $(x^k)_k$ *is a stationary point of the D-gap function* H^{A_*,B_*}.

(ii) *Suppose that infinitely many Newton steps are acceptable to the filter. In this case, every accumulation point of the sequence* $(x^k)_k$ *is a solution of the generalized variational inequality* $(\mathcal{P}_{\text{vip}})$.

(iii) *Let* x^* *be an accumulation point of the sequence* $(x^k)_k$ *and suppose that* x^* *is a solution of the problem* $(\mathcal{P}_{\text{vip}})$*. Furthermore, assume that the conditions* (F.2)–(F.4) *are satisfied. Then, the following statements are valid:*

- *The whole sequence* $(x^k)_k$ *converges to* x^*.
- *Algorithm 4 eventually turns into a pure semismooth Newton method and the sequence* $(x^k)_k$ *converges locally q-superlinearly to* x^*.
- *If, in addition, the proximity operator* $\operatorname{prox}_{\varphi}^{\Lambda_*}$ *is* α*-order semismooth at* u^* *for some* $\alpha \in (0, 1]$ *and the derivatives* $DF(x)$ *and* $DG(x)$ *are Lipschitz continuous near* x^**, then the order of convergence is* $1 + \alpha$.

Proof. Since part (i) clearly follows from Theorem 6.3.1 and part (ii), we first verify the second part. Therefore, let $x^* \in \mathbb{R}^n$ be an arbitrary accumulation point of the sequence $(x^k)_k$ and let $(x^k)_K$ be a corresponding subsequence that converges to x^*. Similar to our analysis in chapter 4, we will work with the following sets

$$\mathcal{K}_N := \{k : x^k \text{ was generated by the Newton method}\},$$
$$\mathcal{K}_D := \{k : x^k \text{ was generated by the D-gap function-based method}\}.$$

Since we perform infinitely many Newton steps, the sequence $(V^{\Lambda_k}(x^k))_{\mathcal{K}_N}$ converges to zero by the abstract filter result that was presented in Lemma 4.3.1. Moreover, in the case $|K \cap \mathcal{K}_N| = \infty$, this already implies

$$V^{\Lambda_k}(x^k) \to V^{\Lambda_*}(x^*) = 0, \quad K \ni k \to \infty,$$

where we used assumption (F.1) and the continuity of F, G, and $\text{prox}_{\varphi}^{\Lambda_*}$. Hence, by Lemma 6.1.1 (iv), we conclude that x^* is a solution of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$. Next, let us consider the case $|K \cap \mathcal{K}_N| < \infty$. Due to Lemma 6.2.8, Remark 6.2.9, and condition (F.1), there exists a constant $\overline{\lambda}$ such that

$$0 \leq H^{A_*,B_*}(x^k) \leq \overline{\lambda} \cdot \|V^{\Lambda_*}(x^k)\|, \quad \forall\, k \geq k^*.$$

Thus, by taking the limit $\mathcal{K}_N \ni k \to \infty$ and using $H^{A_k,B_k}(x^k) = H^{A_*,B_*}(x^k)$ for all $k \geq k^*$, this establishes

$$H^{A_*,B_*}(x^k) \to 0, \quad \text{as } \mathcal{K}_N \ni k \to \infty.$$

Since the algorithm does not terminate after a finite number of steps, the descent property of the D-gap function-based gradient method,

$$H^{A_*,B_*}(x^{k+1}) < H^{A_*,B_*}(x^k), \quad \forall\, k+1 \in \mathcal{K}_D,\ k \geq k^*,$$

yields $H^{A_*,B_*}(x^*) = 0$. This finishes the proof the first two parts.

Now, in order to prove part (iii), let us additionally assume that the accumulation point x^* is a solution of the generalized variational inequality problem $(\mathcal{P}_{\text{vip}})$. Moreover, for a moment, let us suppose that only finitely many Newton steps are performed. In this case, Algorithm 4 reduces to the D-gap function-based gradient descent method and similar to our preceding discussion, we obtain

$$H^{A_*,B_*}(x^k) \to 0, \quad \text{as } k \to \infty.$$

Consequently, together with part (ii) this implies that *every* accumulation point of the sequence $(x^k)_k$ is a solution of the problem $(\mathcal{P}_{\text{vip}})$. Hence, by assumption (F.4) we conclude that x^* is an isolated accumulation point of $(x^k)_k$. To show convergence of the entire sequence $(x^k)_k$, we again want to apply the result of Moré and Sorensen [160] that was already used in Theorem 4.3.10 and 4.3.12. Thus, let $(x^k)_K$ be a subsequence that converges to the isolated accumulation point x^*. Then, due to condition (F.3), there exist constants $C > 0$ and $k_0 \geq \max\{k_*, k^*\}$ such that $\|M_k^{-1}\| \leq C$ for all $k \geq k_0$. Thus, for all $k \in K$, $k \geq k_0$, it holds

$$\|x^{k+1} - x^k\| \leq \begin{cases} C\|V^{\Lambda_*}(x^k)\| & \text{if } k+1 \in \mathcal{K}_N, \\ \|\nabla H^{A_*,B_*}(x^k)\| & \text{if } k+1 \in \mathcal{K}_D. \end{cases}$$

Next, since x^* is a solution of the problem $(\mathcal{P}_{\text{vip}})$, we have $V^{A_*}(x^k) \to 0$, $V^{B_*}(x^k) \to 0$, and $V^{\Lambda_*}(x^k) \to 0$ as $K \ni k \to \infty$. By Lemma 6.2.6, this implies $\|\nabla H^{A_*,B_*}(x^k)\| \to 0$ for $K \ni k \to \infty$ and altogether, we deduce $(\|x^{k+1} - x^k\|)_K \to 0$. Thus, in this situation, [160, Lemma 4.10] again yields that the whole sequence $(x^k)_k$ converges to x^*.

The rest of the proof is identical to the proof of Theorem 4.3.10 and Theorem 4.3.12. □

Remark 6.3.4. Let us note that, in order to establish the (global) convergence results in Theorem 6.3.3 (i)–(ii), it suffices to assume that the parameter matrices Λ_k, $k \in \mathbb{N}$ stay in a compact set $\mathcal{K} \subset \mathbb{S}^n_{++}$. Moreover, in part (iii), if assumption (F.3) is substituted by the stronger CD-regularity condition that was presented in Remark 4.3.7, then condition (F.4) is again a consequence of the semismoothness of V^{Λ_*} and [182, Proposition 3]. In this case, assumption (F.4) is superfluous and can be omitted.

In contrast to the discussion of Algorithm 2 in chapter 4, the D-gap function-based globalization strategy notably simplifies the convergence analysis of Algorithm 4. On the other hand, without any further assumptions, we can only guarantee global convergence to stationary points of the D-gap function. In particular, if Algorithm 4 is used to solve our initial nonsmooth problem $(\mathcal{P})$, then the generated Newton and D-gap steps only operate on the first order optimality condition,

$$\langle \nabla f(x), y - x \rangle + \varphi(y) - \varphi(x) \geq 0, \quad \forall\, y \in \mathbb{R}^n,$$

and the connection to the underlying optimization problem $(\mathcal{P})$ is no longer taken into account. However, if the Hessian $\nabla^2 f$ is positive definite at some accumulation point x^* and if $V^{\Lambda_*} \equiv F^{\Lambda_*}$ is semismooth at x^*, then Lemma 5.4.2, Remark 6.3.4, and condition (6.2.7) imply that all assumptions of Theorem 6.3.3 (iii) are satisfied. Thus, in this situation, x^* is an isolated stationary point and a globally optimal solution of problem $(\mathcal{P})$ and Algorithm 4 is ensured to converge locally q-superlinearly to x^*. See also Theorem 4.3.10 and 4.3.12 for comparison.

In the spirit of our second order analysis in chapter 5, we will now show that the assumptions (F.3) and (F.4) are fulfilled whenever a certain second-order type condition and the strict complementarity condition hold at a solution of problem $(\mathcal{P}_{\text{vip}})$. The proof of the following theorem essentially relies on the techniques that were presented and used in Theorem 5.4.4 and 5.5.2 and finishes this section.

Theorem 6.3.5. *Let $F, G : \Omega \to \mathbb{R}$ be continuously differentiable functions and let $\Omega \subset \mathbb{R}^n$ be an open set that contains the domain $G^{-1}(\text{dom } \varphi)$. Furthermore, let $\bar{x} \in G^{-1}(\text{dom } \varphi)$ be a solution of the problem $(\mathcal{P}_{\text{vip}})$ and let G be twice continuously differentiable in a neighborhood of $\bar{x}$. Additionally, let us suppose that φ is C^2-fully decomposable at $G(\bar{x})$ and assume that the strict complementarity condition*

$$-F(\bar{x}) \in \text{ri } \partial\varphi(G(\bar{x}))$$

holds at $\bar{x}$. Then, for every parameter matrix $\Lambda \in \mathbb{S}^n_{++}$, the proximity operator $\text{prox}^\Lambda_\varphi$ is Fréchet differentiable at $\bar{u} := G(\bar{x}) - \Lambda^{-1}F(\bar{x})$ and the second order-type conditions

$$2\langle DG(\bar{x})h, DF(\bar{x})h \rangle + \langle DG(\bar{x})h, [\Lambda^{-\frac{1}{2}} \mathcal{Q}^\Lambda_\varphi(\bar{u})^+ \Lambda^{\frac{1}{2}} - I] DG(\bar{x})h \rangle_\Lambda > 0, \quad \forall\, h \in \mathcal{C}_G(\bar{x}) \setminus \{0\},$$

where $\mathcal{Q}^\Lambda_\varphi(\bar{u}) := \Lambda^{\frac{1}{2}} D\text{prox}^\Lambda_\varphi(\bar{u}) \Lambda^{-\frac{1}{2}}$, imply the following properties:

(i) *The point $\bar{x}$ is an isolated solution of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$.*

(ii) *If the proximity operator $\text{prox}^\Lambda_\varphi$ is semismooth at $\bar{u}$, then the mapping V^Λ is strictly differentiable at $\bar{x}$ and its Fréchet derivative $DV^\Lambda(\bar{x})$ is nonsingular.*

Proof. The first part was already shown in Theorem 6.1.6 and in the subsequent discussion. At this point, let us recall that if the mapping φ is C^2-fully decomposable at $G(\bar{x})$ and if the strict complementarity condition is satisfied at $\bar{x}$, then it holds

$$-\xi^*_{\varphi,h}(-F(\bar{x})) = \langle DG(\bar{x})h, [\Lambda^{-\frac{1}{2}}\mathcal{Q}^\Lambda_\varphi(\bar{u})^+\Lambda^{\frac{1}{2}} - I]DG(\bar{x})h\rangle_\Lambda, \quad \forall\, h \in \mathcal{C}_G(\bar{x}).$$

The strict differentiability is again a consequence of Theorem 2.6.7. Next, let us assume that the matrix $DV^\Lambda(\bar{x})$ is not invertible. Then, there exists $h \in \mathbb{R}^n \setminus \{0\}$ such that

$$DG(\bar{x})h = W(DG(\bar{x}) - \Lambda^{-1}DF(\bar{x}))h, \quad W := D\mathrm{prox}^\Lambda_\varphi(\bar{u}).$$

Hence, Lemma 3.3.6 implies $DG(\bar{x})h \in N_{\partial\varphi(G(\bar{x}))}(-F(\bar{x}))$ and it follows $h \in \mathcal{C}_G(\bar{x}) \setminus \{0\}$. Since the operator $\mathcal{Q}^\Lambda_\varphi(\bar{u})$ has the same basic properties as its associated counterparts in Theorem 5.4.4 or 5.5.2, we also have

$$[\mathcal{Q}^\Lambda_\varphi(\bar{u})^+\mathcal{Q}^\Lambda_\varphi(\bar{u})]\Lambda^{\frac{1}{2}}DG(\bar{x})h = \Lambda^{\frac{1}{2}}DG(\bar{x})h,$$

see, e.g., subsection 5.3.4 for details. Thus, we obtain

$$[\mathcal{Q}^\Lambda_\varphi(\bar{u})^+\mathcal{Q}^\Lambda_\varphi(\bar{u})]\Lambda^{-\frac{1}{2}}DF(\bar{x})h = [I - \mathcal{Q}^\Lambda_\varphi(\bar{u})^+]\Lambda^{\frac{1}{2}}DG(\bar{x})h.$$

Furthermore, by using the symmetry of the matrix $\mathcal{Q}^\Lambda_\varphi(\bar{u})$ and the properties of the Moore-Penrose pseudoinverse, we get

$$\begin{aligned}
&\langle DF(\bar{x})h, DV^\Lambda(\bar{x})h\rangle - \langle DF(\bar{x})h, DG(\bar{x})h\rangle \\
&\qquad = -\langle DF(\bar{x})h, WDG(\bar{x})h\rangle + \langle \Lambda^{-\frac{1}{2}}DF(\bar{x})h, \mathcal{Q}^\Lambda_\varphi(\bar{u})\Lambda^{-\frac{1}{2}}DF(\bar{x})h\rangle \\
&\qquad = -\langle DF(\bar{x})h, WDG(\bar{x})h\rangle + \langle [\mathcal{Q}^\Lambda_\varphi(\bar{u}) - I]\Lambda^{\frac{1}{2}}DG(\bar{x})h, [I - \mathcal{Q}^\Lambda_\varphi(\bar{u})^+]\Lambda^{\frac{1}{2}}DG(\bar{x})h\rangle \\
&\qquad = -\langle DF(\bar{x})h, WDG(\bar{x})h\rangle + \langle [\mathcal{Q}^\Lambda_\varphi(\bar{u}) - I]\Lambda^{\frac{1}{2}}DG(\bar{x})h, \Lambda^{\frac{1}{2}}DG(\bar{x})h\rangle - \xi^*_{\varphi,h}(-F(\bar{x})) \\
&\qquad = -\xi^*_{\varphi,h}(-F(\bar{x})).
\end{aligned}$$

Now, due to $DV^\Lambda(\bar{x})h = 0$ and $h \in \mathcal{C}_G(\bar{x}) \setminus \{0\}$, the latter calculations and the second order-type conditions imply

$$\langle DF(\bar{x})h, DG(\bar{x})h\rangle > 0.$$

However, since the curvature term $-\xi^*_{\varphi,h}(-F(\bar{x}))$ is nonnegative for all $h \in \mathcal{C}_G(\bar{x})$, this yields the contradiction

$$0 = \langle DF(\bar{x})h, DV^\Lambda(\bar{x})h\rangle = \langle DF(\bar{x})h, DG(\bar{x})h\rangle - \xi^*_{\varphi,h}(-F(\bar{x})) > 0.$$

Consequently, we deduce $h = 0$. This concludes the proof of Theorem 6.3.5. □

7. Applications and numerical results

In this chapter, we present numerical results and discuss the competitiveness of the semismooth Newton method that was proposed in chapter 4 for different nonsmooth optimization problems and in comparison with several state-of-the-art algorithms.

Let us note that the numerical comparisons for ℓ_1-regularized problems in section 7.1 and 7.2 are essentially based on the work [157] and that several parts have already appeared in a similar form in [157]. However, we also want to emphasize that the results reported in this thesis were obtained by using a more refined and improved version of Algorithm 2 and thus, are not immediately comparable with the results in [157].

All tests in this chapter were performed under MATLAB v8.5 (R2015a) on an iMac 27" with Intel Core i5 3,2 GHz and 16 GB of memory.

7.1. Convex ℓ_1-regularized least squares problems

At first, based on a test framework in [13], we provide an extensive numerical comparison of different ℓ_1-optimization methods, that are particulary designed to solve either *basis pursuit denoising* problems of the form

$$(\mathrm{BP}_\sigma) \qquad \min_{x\in\mathbb{R}^n} \ \|x\|_1 \quad \text{s.t.} \quad \|Ax-b\|_2 \le \sigma$$

or corresponding ℓ_1-regularized quadratic problems of the form

$$(\mathrm{QP}_\mu) \qquad \min_{x\in\mathbb{R}^n} \ \frac{1}{2}\|Ax-b\|_2^2 + \mu\,\|x\|_1 =: \psi(x).$$

From now on and for ℓ_1-least squares problems of the form (QP_μ), we will refer to the specialized version of Algorithm 2 as SNF-L1 (semismooth Newton filter) method.

Remark 7.1.1. Clearly, due to the nonnegativity of the quadratic term $f(x) = \frac{1}{2}\|Ax-b\|_2^2$, the objective function ψ of problem (QP_μ) is coercive, i.e., there exists at least one solution and the set of all possible solutions must be bounded. Moreover, since f is convex, quadratic, and twice continuously differentiable on $\mathbb{R}^n$ and since the ℓ_1-norm is obviously real valued and positively homogeneous, the assumptions (A.1) and (C.1)–(C.3) are satisfied. Thus, Theorem 4.3.3 and Remark 4.3.4 guarantee that the SNF-L1 method converges globally. Additionally, and as sketched in Example 5.3.12, if a certain submatrix of the Hessian $\nabla^2 f(x) = A^\top A$ is positive definite, then the conditions for local fast convergence are fulfilled.

We start with several implementational aspects of the SNF-L1 algorithm.

7.1.1. Algorithmic details and implementation

We now briefly describe algorithmic and numerical details of the SNF-L1 method. We want to point out that the following considerations mainly focus on the class of convex and quadratic problems (QP_μ).

Λ-strategy. In our implementation, the parameter matrix $\Lambda_k \in \mathbb{S}^n_{++}$ is chosen based on the following simple strategy:

$$\Lambda_k := \tau_k^{-1} I, \quad \tau_k \in [\tau_m, \tau_M], \quad 0 < \tau_m < \tau_M.$$

In the first iteration, we use $\tau_0 = 6$ as initial value. Afterwards, the parameter τ_k is adjusted to approximate the inverse Lipschitz constant of the gradient $\nabla f(x) = A^\top(Ax - b)$. More specifically, we set

$$\lambda_k^1 = \frac{\|x^k - x^{k-1}\|}{\|\nabla f(x^k) - \nabla f(x^{k-1})\|}, \quad \lambda_k^2 = \max\{\min\{\lambda_k^1, \tau_M\}, \tau_m\}, \quad k > 1. \tag{7.1.1}$$

Finally, in order to prevent outliers, we calculate a weighted mean of λ_k^2 and of the previous parameters τ_j, $j = 1, ..., k-1$. This mean is then used as the new step size parameter τ_k.

Newton system. In Example 4.2.16 we have already shown that for ℓ_1-problems of the form (QP_μ) the nonsmooth mapping $F^{\Lambda_k} : \mathbb{R}^n \to \mathbb{R}^n$ is given by

$$F^{\Lambda_k}(x^k) = \tau_k \nabla f(x^k) + \mathcal{P}_{[-\mu\tau_k, \mu\tau_k]^n}(x^k - \tau_k \nabla f(x^k)).$$

Moreover, as also mentioned in Example 4.2.16 and setting $u^k := x^k - \tau_k \nabla f(x^k)$, we will work with the following generalized derivates

$$M(x^k) := \tau_k(I - D(x^k)) \cdot A^\top A + D(x^k),$$

where the diagonal matrix $D(x^k)$ is uniquely determined via

$$D(x^k)_{[ii]} = \begin{cases} 0 & \text{if } |u_i^k| > \mu\tau_k, \\ 1 & \text{if } |u_i^k| \le \mu\tau_k, \end{cases} \qquad \forall\, i = 1, ..., n.$$

Now, in each iteration we have to solve the system of equations

$$M(x^k)s^k = -F^{\Lambda_k}(x^k),$$

in order to obtain the next Newton step s^k. Thus, the performance of our algorithm highly depends on efficient strategies for solving those systems. By taking advantage of the structure of the generalized derivative $M(x^k)$ and using a simple block elimination technique we can reduce the computational complexity and end up with the smaller problem

$$\begin{aligned} s_{\mathcal{A}}^k &= -F_{\mathcal{A}}^{\Lambda_k}(x^k), \\ (A^\top A)_{[\mathcal{I}\mathcal{I}]} s_{\mathcal{I}}^k &= -\tau_k^{-1} F_{\mathcal{I}}^{\Lambda_k}(x^k) - (A^\top A)_{[\mathcal{I}\mathcal{A}]} s_{\mathcal{A}}^k, \end{aligned} \tag{7.1.2}$$

Table 7.1.: Summary of parameters and their default values

$C_1^{\text{cont}}, C_2^{\text{cont}}, C_{\max}^{\text{cont}}$	factors for the continuation update formula, $C_1^{\text{cont}} = 0.535$, $C_2^{\text{cont}} = -\log_{10}(0.65)$ and maximum number of iterations $C_{\max}^{\text{cont}} = 10$
β, γ	parameters for the quasi-Armijo condition, $\beta = 0.1$, $\gamma = 0.1$
τ_0, τ_m, τ_M	parameter for the adaptive choice of Λ_k, $\tau_0 = 6$, $\tau_m = 10^{-3}$, $\tau_M = 10^4$
$\gamma_{\mathcal{F}}$	factor for the filter acceptance criterion, $\gamma_{\mathcal{F}} = 7 \cdot 10^{-2}$
CG-tol, CG-maxit	parameters to control the accuracy of the CG method, CG-tol $= 0.1$, CG-maxit $= 10$

where we set $\mathcal{A} = \mathcal{A}(x^k) := \{i : |u_i^k| \le \mu\tau_k\}$ and $\mathcal{I} = \mathcal{I}(x^k) := \{i : |u_i^k| > \mu\tau_k\}$. Instead of solving (7.1.2) directly, we consider a regularized version of the submatrix of the Hessian $(A^\top A)_{[\mathcal{I}\mathcal{I}]} + \rho I$ with $\rho = \rho(x^k) := \|F^{\Lambda_k}(x^k)\|$. This leads to the numerically more robust formulation

$$(A^\top A + \rho I)_{[\mathcal{I}\mathcal{I}]} s_{\mathcal{I}}^k = -\tau_k^{-1} F_{\mathcal{I}}^{\Lambda_k}(x^k) - (A^\top A)_{[\mathcal{I}\mathcal{A}]} s_{\mathcal{A}}^k \tag{7.1.3}$$

and corresponds to a reformulation of the Newton system with the regularized matrix M_ρ, which was already introduced in (4.3.17). The remaining problem (7.1.3) is approximately solved by an early terminated (preconditioned) CG method. Since ℓ_1-minimization algorithms are usually used for large-scale applications and the matrix A typically involves direct or inverse discrete cosine, wavelet, or related transforms, the computational effort of every iteration is dominated by the number of applications of A and $A^\top$ to a vector. For convenience, we will use the terms A- and $A^\top$-call to describe an application of A or $A^\top$. Furthermore, let $\mathcal{C}_A$ denote the complexity of applying A or $A^\top$. Then the complexity of a single, successful Newton iteration of the SNF-L1 algorithm is given by $2\mathcal{C}_A + 2\mathcal{C}_A \cdot \texttt{cg-iter}$ (two calls are used to evaluate the right-hand side of equation (7.1.3)). Furthermore, if the current iterate is not acceptable to the filter, we have to apply A (and $A^\top$) once more to obtain an alternative shrinkage step. This complexity bound motivates us to solve the linear system (7.1.3) only approximately, in order to keep the number of CG iterations as low as possible. Hence, we choose a rather mild stopping criterion for the CG method and set the relative tolerance to 10^{-1} and the maximum number of iterations to 10.

Filter. In our implementation, we choose a filter function $\theta : \mathbb{R}^n \to \mathbb{R}_+^p$ of type (4.2.8) with the following decomposition pattern

$$\mathcal{I}_1 = \{1, ..., \ell\},\ \mathcal{I}_2 = \{\ell + 1, ..., 2\ell\},\ ...,\ \mathcal{I}_p = \{(p-1)\ell + 1, ..., n\}, \quad \ell = \left[\tfrac{n}{p}\right].$$

We use $p = 1000$, but experiments show that the algorithm is quite insensitive to the choice of p. Of course, the required filter storage increases proportional to p.

Continuation. The continuation with respect to μ [98] has become a common and successful tool to further improve the performance of ℓ_1-optimization algorithms. The idea is to solve the problem (QP_μ) for a sequence of different μ values. At first, starting with a usually large parameter $\mu_0 > \mu$, an approximate solution x_0^* of the problem (QP_{μ_0}) is computed. We then decrease the regularization parameter, i.e., we choose μ_1 satisfying $\mu_0 > \mu_1 \ge \mu$

and solve (QP_{μ_1}) with x_0^* as initial point. This procedure is repeated until the current regularization parameter μ_j coincides with our desired parameter μ or a termination criterion for the problem (QP_μ) is satisfied. Practical experience and numerical experiments in [78, 98, 256, 13] showed that this homotopy scheme can enhance the performance of ℓ_1-optimization methods significantly. Encouraged by this general observation we embedded the SNF-L1 method in a continuation framework. Particularly, we choose

$$\mu_0 = \max\left\{C_0^{\mathrm{cont}}\|b\|_\infty, \mu\right\}, \quad C_0^{\mathrm{cont}} = \min\left\{0.25, 2.2 \cdot (\|b\|_\infty/\mu)^{-\frac{1}{3}}\right\}$$

and decrease the current homotopy parameter μ_j according to the following update formula

$$\mu_{j+1} = \max\{\gamma_j\mu_j, \mu\}, \quad (0,1] \ni \gamma_j = 1 - C_1^{\mathrm{cont}}\left(\frac{\mu_j}{\mu_0}\right)^{C_2^{\mathrm{cont}}},$$

$C_1^{\mathrm{cont}}, C_2^{\mathrm{cont}} \in (0,1)$ (see Table 7.1). Here, our specific choice of μ_0 is based on a scale invariance argument: if the magnitude of the input data b is increased by a factor, the continuation scheme is supposed to adapt in a similar way. This motivates the above ansatz $\mu_0 = C_0^{\mathrm{cont}}\|b\|_\infty$, where the additional damping factor C_0^{cont} is introduced to avoid disproportionately high initial values. Furthermore, we observed that a logarithmically decreasing update of the regularization parameter yields better numerical results. Thus, we have chosen an adaptive update formula such that the reduction of μ_j gets smaller if the total number of continuation steps increases. We set $x_j^* := x_j^{k+1}$ and reduce μ_j whenever the total number of iterations within a single continuation phase exceeds the bound $C_{\max}^{\mathrm{cont}} = 10$ or a good Newton step is performed, i.e., when the Newton iterate $x_j^{k+1} = x_j^k + s_j^k$ satisfies the following decrease condition

$$\|F^{\Lambda_{k+1}}(x_j^{k+1})\| \leq 0.5\,\|F^{\Lambda_k}(x_j^k)\|, \quad k \geq 0,$$

where x_j^k denotes the k-th iterate of the j-th subproblem (QP_{μ_j}), $k > 0$ and $x_j^0 := x_{j-1}^*$. An exemplary visualization of the development of the continuation phases with respect to the number of A- and $A^\top$-calls can be found in Figure 7.2.

Initial point and stopping tolerance. We choose $x^0 = 0$ as initial point and terminate SNF-L1 when the current residual falls below a given tolerance ε, i.e.,

$$\|F^{\Lambda_k}(x^k)\| \leq \varepsilon, \tag{7.1.4}$$

where we dropped the additional continuation index for convenience. We want to emphasize that the term F^{Λ_k} has to be understood in its original sense, i.e., here F^{Λ_k} depends on the initial regularization parameter μ. The tolerance ε influences the level of accuracy, we will work with $\varepsilon \in \{1, 10^{-1}, 10^{-2}, 10^{-4}, 10^{-6}\}$. Table 7.1 summarizes the default setting of the parameters of SNF-L1.

7.1.2. State-of-the-art methods

In this section we state main ideas and basic structural aspects of several state of the art methods, which will be used later in our numerical comparison. We will work with ℓ_1-

algorithms that are designed for efficient large-scale optimization and can take advantage of fast implementations of the A or $A^\top$ application.

Fixed Point Continuation method (FPC) [98]. The Fixed Point Continuation method is a first order algorithm for solving the problem (QP$_\mu$) or more general ℓ_1-problems with convex f. It is a direct realization of the fixed point iteration

$$x^{k+1} = \mathrm{prox}^{\Lambda}_{\mu\|\cdot\|_1}(x^k - \tau \nabla f(x^k)), \quad \Lambda = \tau^{-1} I, \quad \tau > 0 \tag{7.1.5}$$

with an additional continuation scheme for the ℓ_1-regularization parameter μ. FPC-BB is an advanced version of FPC that uses Barzilai-Borwein steps to improve performance. The code can be found online at `http://www.caam.rice.edu/~optimization/L1/fpc/`. All parameters were set to default values.

FPC Active Set (FPC-AS) [253]. FPC-AS is an extended two-phase version of the FPC method and is designed to solve (QP$_\mu$). In the first stage a specialized, nonmonotone version of Algorithm 1 with $\Lambda = \tau^{-1}I$ and a Barzilai-Borwein heuristic for the parameter τ is used to determine an active set. Motivated by Greedy algorithms for ℓ_1-optimization, FPC-AS then solves a smooth subproblem on this active set with a L-BFGS method. The algorithm is embedded in a continuation scheme. The code is available at `http://www.caam.rice.edu/~optimization/L1/FPC_AS/`. All parameters were set to default values.

Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [12]. FISTA is an accelerated proximal gradient method that resembles Nesterov's fast gradient schemes for convex problems [169]. It can be seen as an extension of the fixed-point iteration (7.1.5) with an additional extrapolation step. In our numerical experiments we implemented FISTA as described [12, Section 4] and with constant step size $L = 1$.

Gradient Projections for Sparse Reconstruction (GPSR) [78]. GPSR is based on the well-known projected gradient technique for constrained optimization problems. By splitting $x = v_1 - v_2$ into its positive and negative part v_1 and v_2 the ℓ_1-regularized problem (QP$_\mu$) is smoothed and reformulated as a quadratic program with positivity constraints for v_1 and v_2. In our experiments we tested GPSR with continuation and an alternative version with Barzilai-Borwein step sizes (we refer to GPSR-BB). As recommended by the authors and proposed in [13], all parameters were set to default except the number of continuation steps was set to 40, the `ToleranceA` variable was set to 10^{-3}, and the `MiniterA` variable was set to 1. The code is available at `http://www.lx.it.pt/~mtf/GPSR/`.

NESTA [13]. NESTA is built on Nesterov's smoothing technique for convex and possibly nonsmooth functions and applies this technique to the constrained ℓ_1-problem (BP$_\sigma$). In [13] it was shown that the performance of Nesterov's framework can be significantly improved by using a continuation scheme on the smoothing parameter that characterizes the level of smoothing of the ℓ_1-norm. We tested NESTA with different smoothing parameters, $\mu \in \{0.01, 0.02, 10^{-8}\}$ (unfortunately, the meaning of μ here is different from its standard use in this section) and two continuation scenarios, where the number of continuation steps was set to either $T = 4$ or $T = 5$. All other parameters were set to default, except, as proposed in [13], the tolerance variable δ was set to 10^{-7}. The code can be found at `http://www-stat.stanford.edu/~candes/nesta/`.

Primal-Dual Method (PD) [42]. Chambolle and Pock's primal-dual method is able to solve general nonsmooth problems of the form

$$\min_{x\in\mathbb{R}^n} \varphi(x) + \varrho(Kx), \tag{7.1.6}$$

where the functions $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ and $\varrho : \mathbb{R}^m \to (-\infty, +\infty]$ are convex, proper, and lower semicontinuous functions and $K \in \mathbb{R}^{m\times n}$ is a given matrix. In each iteration, the method successively performs the following steps:

$$\begin{aligned} y^{k+1} &= \mathrm{prox}_{\varrho^*}^{\sigma^{-1}I}(y^k + \sigma K\bar{x}^k), \\ x^{k+1} &= \mathrm{prox}_{\varphi}^{\tau^{-1}I}(x^k - \tau K^\top y^{k+1}), \\ \bar{x}^{k+1} &= x^{k+1} + \theta(x^{k+1} - x^k). \end{aligned}$$

If the step sizes σ and τ satisfy $\sigma\tau\|K\|^2 < 1$ and if it holds $\theta = 1$, then the sequence (x^k, y^k) converges to a saddle point of the corresponding primal-dual formulation of problem (7.1.6). For the ℓ_1-problem (QP_μ), we set $\varphi(x) := \mu\|x\|_1$, $\varrho(y) := \frac{1}{2}\|y - b\|_2^2$, and $K := A$. Moreover, since the convex conjugate ϱ^* is strongly convex with convexity parameter $\gamma = 1$, we can use an accelerated version of the primal-dual method that is presented in [42, Section 5] (in this variant the variables σ, τ, and θ are updated adaptively in each iteration). We implemented the accelerated primal-dual method as specified in [42, Algorithm 2] and set $\tau_0 = 1$, $\sigma_0 = 1$, and $\gamma = 0.9$.

Sparse reconstruction by separable approximation (SpaRSA) [256]. The method SpaRSA was developed to solve the general problem ($\mathcal{P}$). For (QP_μ) it is an iterative shrinkage-based algorithm and therefore resembles FPC. SpaRSA also uses Barzilai-Borwein steps and a continuation technique to accelerate its performance. Online code can be obtained at `http://www.lx.it.pt/~mtf/SpaRSA/`. Again, as recommended, we set all parameter to default and adopt the parameter modifications in GPSR-BB.

Spectral projected gradient (SPGL1) [242]. SPGL1 solves the basis pursuit problem (BP_σ) via finding roots of a corresponding one-dimensional nonlinear equation. This procedure involves solving a sequence of so-called LASSO problems

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|_2 \quad \text{s.t.} \quad \|x\|_1 \le \tau$$

for different values of τ. In [242] a spectral projected gradient method is used to efficiently compute approximate solutions of the above least squares problems. The code is available at `http://www.cs.ubc.ca/labs/scl/spgl1/`. All parameters were set to default values.

Alternating Direction Method of Multipliers (YALL1) [258]. The YALL1 package provides a general alternating direction method [92, 89, 91, 71] that can solve a variety of constrained and unconstrained ℓ_1-problems. Similar to the primal-dual method it introduces an auxiliary variable to split the objective function into two separate parts. The resulting reformulated problem is then solved with an Augmented Lagrange method in an alternating fashion. Let us note that the performance of YALL1 strongly depends on how efficiently its corresponding subproblems can be solved. In our numerical comparison, since the measurement matrix A will be chosen as an orthogonal projector with $AA^\top = I$, the different steps of the YALL1-algorithm are given by simple and fast update rules. For more details on the alternating direction method we refer to [91, 71, 258] and to section 7.3.2. Online code can be found at `http://yall1.blogs.rice.edu/`. All parameters were set to default values.

7.1.3. Numerical comparison

To compare SNF-L1 with several other methods we use a slightly modified test framework from the NESTA package [13]. The problem setting is identical to the one proposed in [13] and is specified as follows. At first, we generate a sparse signal $\bar{x} \in \mathbb{R}^n$ of length $n = 512^2 = 262144$ with $k = [n/40] = 5553$ nonzero entries. Here, the k different indices $i \in \{1, ..., n\}$ are randomly chosen and the magnitude of each nonzero component is determined via

$$\bar{x}_i = \eta_1(i) 10^{d\eta_2(i)/20},$$

where $\eta_1(i) \in \{-1, +1\}$ is a symmetric random sign and $\eta_2(i)$ is uniformly distributed in $[0, 1]$. The signal has dynamic range of d dB and we consider $d \in \{20, 40, 60, 80\}$. The matrix $A \in \mathbb{R}^{m \times n}$ takes $m = n/8 = 32768$ random cosine measurements, i.e., $Ax = (\texttt{dct}(x))_J$, where the index set $J \subset \{1, ..., n\}$, $|J| = m$, is initialized randomly and `dct` is the discrete cosine transform. Finally, the input data $b \in \mathbb{R}^m$ is obtained by adding Gaussian noise with standard deviation $\bar{\sigma} = 0.1$ to $A\bar{x}$.

Since NESTA and SPGL1 solve the basis pursuit problem (BP_σ) while all other mentioned algorithms solve the unconstrained problem (QP_μ) we need to compute a corresponding pair (σ, μ) to gain comparable results at first. Therefore, we run SPGL1 to generate an approximate solution of the problem (BP_{σ_0}) with $\sigma_0 = \sqrt{m + 2\sqrt{2m}}\bar{\sigma}$ and to obtain an estimate $\mu(\sigma_0)$ from its dual solution. Afterwards, we use the SNF-L1 algorithm with stopping criterion

$$\|F^I(x^k)\| \leq 10^{-12}$$

to compute a high precision solution x^* of the problem ($\text{QP}_{\mu(\sigma_0)}$) and set $\sigma = \|Ax^* - b\|$. Then the problems (QP_μ) and (BP_σ) should be almost equivalent. Now, the SNF-L1 method is run again with stopping rule

$$(\mathcal{C}_{\text{nat}}) \qquad \|F^I(x^k)\| \leq \varepsilon$$

and with different tolerances ε to create a series of reference solutions. We modified the stopping criterion of each algorithm; the other algorithms now terminate at iteration k when the current iterate x^k_{alg} satisfies the following relative stopping criterion

$$(\mathcal{C}_{\text{rel}}) \qquad \frac{|\psi(x^k_{\text{alg}}) - \psi(x^*)|}{\psi(x^*)} \leq \frac{|\psi(x^*_{\text{snf}}) - \psi(x^*)|}{\psi(x^*)},$$

where x^*_{snf} denotes the solution of the SNF-L1 method. Moreover, we also performed a second independent test series where each algorithm uses the condition ($\mathcal{C}_{\text{nat}}$) as stopping criterion. In this case, termination and accuracy of the different methods is solely controlled by the stopping rule ($\mathcal{C}_{\text{nat}}$) and do no longer depend on the results of SNF-L1.

Table 7.2.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 20 dB using the stopping criteria $(\mathcal{C}_{\text{rel}})$ and $(\mathcal{C}_{\text{nat}})$ (best NESTA configuration was used: $\mu = 10^{-8}$, $T = 4$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-L1	91	91	227	227	260	260	289	289	342	342
FISTA	70	56	203	149	474	349	1879	1480	4333	3549
FPC	372	374	514	432	778	646	1490	1332	2385	2118
FPC-BB	148	150	180	174	269	206	960	802	DNC	1588
FPC-AS	80	74	218	212	311	312	393	409	454	434
NESTA	570	570	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
GPSR	551	561	675	614	902	818	1481	1387	DNC	2026
GPSR-BB	390	418	445	435	522	487	753	706	1055	964
PD	86	104	420	307	1127	839	DNC	7698	DNC	DNC
SpaRSA	485	503	528	521	559	550	628	621	702	688
SPGL1	66	58	152	124	192	181	DNC	DNC	DNC	DNC
YALL1	60	56	144	126	223	202	399	374	620	558

In the following, we consider the dynamic ranges $d \in \{20, 40, 60, 80\}$ and the tolerances $\varepsilon \in \{1, 10^{-1}, 10^{-2}, 10^{-4}, 10^{-6}\}$. Since an application of A or $A^\top$ corresponds to the evaluation of a `dct` or `idct` function, the total number of A- and $A^\top$-calls is an important measure of efficiency. For our two different test series these total numbers will be denoted by $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$, respectively. Thus, along with the corresponding total runtime of each algorithm, our numerical comparison is based on a discussion of the different, achieved $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ values. We report `DNC` (did not converge) if convergence is not reached after $\#A_{\text{rel}} = 20000$ or $\#A_{\text{nat}} = 20000$ calls. The Tables 7.2, 7.3, 7.4, and 7.5 contain the mean values of the numbers $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ over 10 random trials; the three best results from each column are shaded. Accordingly, the Tables A.1, A.2, A.3, and A.4 contain the corresponding mean values of the total runtimes t_{rel} and t_{nat} of the different algorithms. Here, we set $t_{\text{rel}} = \texttt{DNC}$ or $t_{\text{nat}} = \texttt{DNC}$ when the respective algorithm did not converge within 20000 A-calls. Again, the three best results fro each column are shaded. For the sake of clarity, the Tables A.1–A.4 have been moved to the Appendix A.4.1. In Figure 7.1 we illustrate the change of the absolute value of 128 randomly chosen components of the iterate with respect to the number of A- and $A^\top$-calls. Each plot is associated with one of the tested algorithms and shows the development of the iterate for a single run with dynamic range $d = 40$ dB. Furthermore, the maximum number of A- and $A^\top$-calls that is necessary to capture all zero components of the optimal solution x^* is marked with a green line. Similar illustrations for the other dynamic ranges $d \in \{20, 60, 80\}$ can be found in the Appendix A.4.1, see Figures A.1–A.3.

It was observed in [13, 139] that ℓ_1-optimization algorithms react sensitively on changes of the dynamic range and their performances usually deteriorate with increasing dynamic range. Our experiments also confirm this behavior for the modified test framework.

Table 7.3.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 40 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$) (best NESTA configuration was used: $\mu = 10^{-1}$, $T = 5$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-L1	180	180	459	459	496	496	542	542	610	610
FISTA	196	234	555	421	1193	833	4280	2934	7543	6402
FPC	367	378	997	679	1531	1203	2867	2355	3925	3561
FPC-BB	160	169	441	223	975	648	2309	1797	DNC	3192
FPC-AS	252	164	448	324	542	526	616	608	703	676
NESTA	374	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
GPSR	461	472	836	698	1082	942	1692	1469	DNC	DNC
GPSR-BB	428	467	583	493	781	668	1287	1101	1695	1559
PD	165	240	718	582	2158	1769	DNC	DNC	DNC	DNC
SpaRSA	478	519	566	538	621	596	738	703	821	799
SPGL1	100	110	286	228	342	318	DNC	DNC	DNC	DNC
YALL1	131	202	451	392	812	676	1894	1517	2804	2516

SPGL1 is very efficient at low and middle precisions. However, SPGL1 does not converge in the high precision examples. In the low precision case, it takes about 4 times as many A applications on the 80 dB signal as on the 20 dB signal (for lower tolerances this factor diminishes to 3). Figure 7.1 (k) demonstrates that SPGL1 quickly detects zero and nonzero components of the optimal solution x^*.

The Barzilai-Borwein version of GPSR outperforms the regular GPSR version in both number of A- and $A^\top$-calls and CPU time, though it cannot keep up with the results of SNF-L1. In general, GPSR-BB requires twice as many A- and $A^\top$-calls at 80 dB than at 20 dB. The runtimes for 80 dB and 20 dB differ by a factor between 1.6 and 2.6.

SpaRSA needs a comparatively large number of A- and $A^\top$-calls at low dynamic range and low precision tests. For fixed dynamic range it only requires about 1.5–1.9 times as many A- and $A^\top$-calls to compute a very accurate solution as in the low precision case. Besides, SpaRSA shows good performance with large dynamic range and requires less A- and $A^\top$-calls than SNF-L1 in the 80 dB example (see Table 7.5). However, concerning computational time, SpaRSA does not succeed in outperforming SNF-L1. Both GPSR-BB and SpaRSA show a similar development of their iterates. The Figures 7.1 (h) and (j) illustrate that almost all zero components of x^* are captured within the very first iterations. However, there are several outliers that prevent fast convergence to the correct zero pattern. Compared to SNF-L1, a large number of A-calls is needed to detect the small nonzero components of x^*, see also Figures A.1–A.3. In general, we observe that both SpaRSA and GPSR-BB become more competitive in the problems with higher dynamic range.

The NESTA algorithm did not converge in most of our tests. This can be traced back to NESTA's smoothing of the ℓ_1-norm that is used in the algorithm and the resulting relatively

Table 7.4.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 60 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$) (here, NESTA did not converge).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-L1	378	378	666	666	702	702	765	765	860	860
FISTA	641	630	1149	946	1896	1543	4939	4213	9340	8511
FPC	625	521	1700	1177	2328	1937	3820	3497	5376	5098
FPC-BB	236	225	1212	690	1838	1446	3329	3006	DNC	DNC
FPC-AS	243	214	648	598	717	690	791	779	898	883
NESTA	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
GPSR	726	455	1233	1007	1519	1354	2200	2066	DNC	DNC
GPSR-BB	453	462	800	614	1049	912	1634	1526	2253	2158
PD	420	561	1636	1422	4597	4765	DNC	DNC	DNC	DNC
SpaRSA	546	563	649	609	721	686	847	827	943	927
SPGL1	168	168	389	319	459	434	DNC	DNC	DNC	DNC
YALL1	606	1046	2492	2322	3986	3961	7610	7598	11346	11019

Table 7.5.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 80 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$) (here, NESTA did not converge).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-L1	644	644	939	939	979	979	1058	1058	1174	1174
FISTA	2297	2293	3030	2757	4019	3650	7948	7222	12716	12753
FPC	1243	834	2526	1816	3271	2800	5080	4723	6674	6685
FPC-BB	499	256	1773	1064	2518	2047	4331	3975	DNC	DNC
FPC-AS	390	397	784	738	836	824	950	935	1072	1083
NESTA	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
GPSR	2079	980	2674	2366	3017	2818	3846	3700	DNC	DNC
GPSR-BB	596	424	1004	782	1249	1100	1849	1737	2388	2389
PD	1158	1476	4140	3729	11196	12357	DNC	DNC	DNC	DNC
SpaRSA	584	529	747	673	820	789	964	947	1072	1076
SPGL1	250	245	475	378	562	529	DNC	DNC	DNC	DNC
YALL1	5445	8530	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC

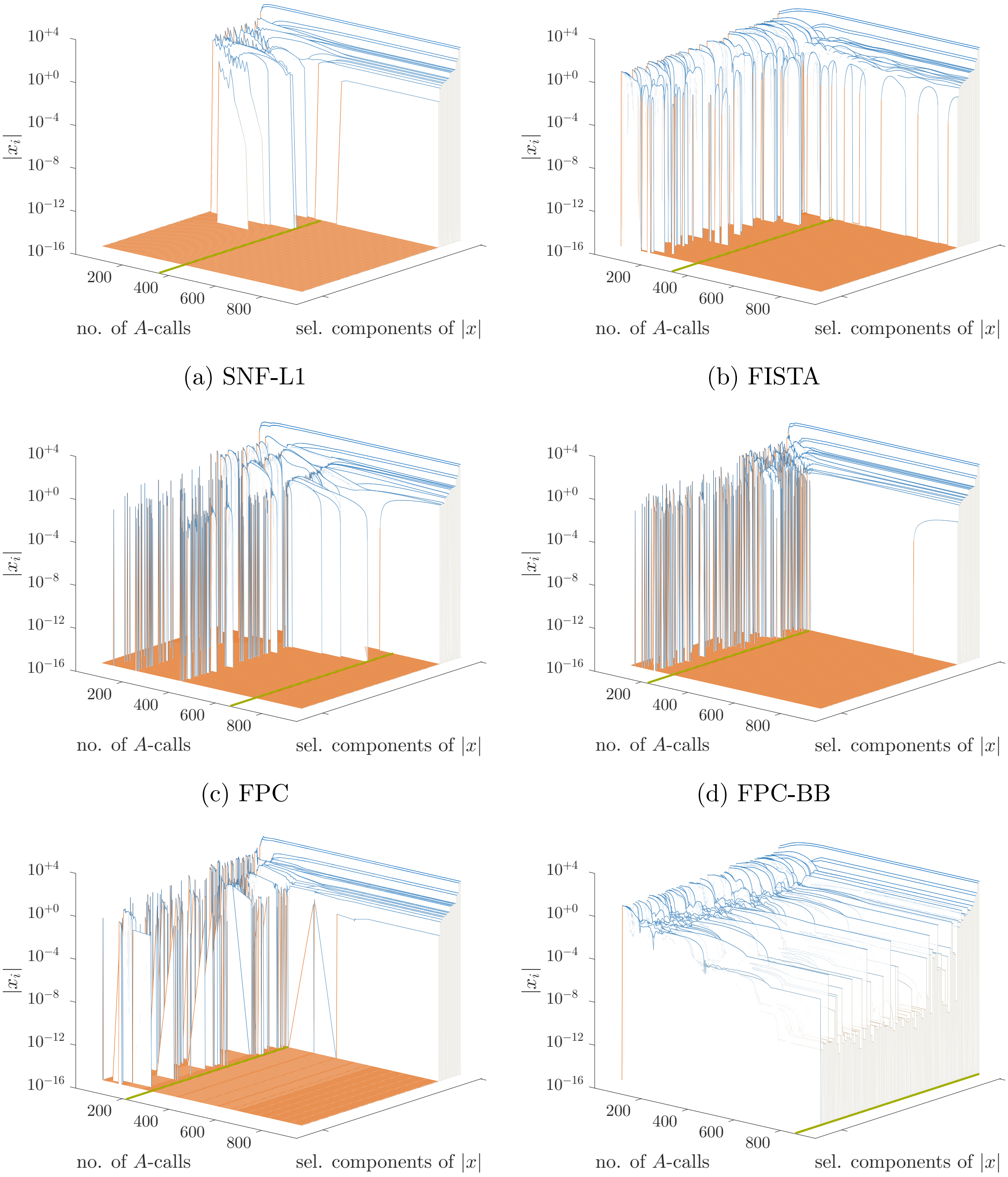
Table 7.6.: Total number of iterations, shrinkage steps $\mathcal{S}$-iter and A- and $A^\top$-calls averaged over 10 independent runs. The minimum and maximum value of $\mathcal{S}$-iter over the 10 runs are shown in smaller font. The method SNF-L1a^4 uses a fixed parameter choice $\tau_k = \tau = 4$.

Method		$\varepsilon : 10^{-2}$			$\varepsilon : 10^{-6}$			$\varepsilon : 10^{-10}$		
		iter	$\mathcal{S}$-iter	$\#A_{\text{nat}}$	iter	$\mathcal{S}$-iter	$\#A_{\text{nat}}$	iter	$\mathcal{S}$-iter	$\#A_{\text{nat}}$
SNF-L1	20 dB	22	$0_{(0/0)}$	260	25	$0_{(0/0)}$	342	31	$0_{(0/0)}$	461
SNF-L1a		22	$0_{(0/0)}$	260	24	$0_{(0/0)}$	346	25	$0_{(0/0)}$	410
SNF-L1a^4		22	$0_{(0/0)}$	272	26	$0_{(0/0)}$	373	27	$0_{(0/0)}$	437
SNF-L1	40 dB	46	$0_{(0/0)}$	496	51	$0_{(0/0)}$	610	58	$0_{(0/0)}$	773
SNF-L1a		46	$0_{(0/0)}$	496	48	$0_{(0/0)}$	595	50	$0_{(0/0)}$	724
SNF-L1a^4		49	$0_{(0/0)}$	506	52	$0_{(0/0)}$	609	53	$0_{(0/0)}$	737
SNF-L1	60 dB	70	$0_{(0/0)}$	702	77	$0_{(0/0)}$	860	86	$0_{(0/0)}$	1060
SNF-L1a		70	$0_{(0/0)}$	702	73	$0_{(0/0)}$	820	75	$0_{(0/0)}$	943
SNF-L1a^4		87	$5_{(4/9)}$	807	89	$5_{(4/9)}$	917	91	$5_{(4/9)}$	1053
SNF-L1	80 dB	105	$1_{(0/2)}$	979	114	$1_{(0/2)}$	1174	125	$1_{(0/2)}$	1410
SNF-L1a		105	$1_{(0/2)}$	979	109	$1_{(0/2)}$	1131	110	$1_{(0/2)}$	1238
SNF-L1a^4		105	$2_{(0/4)}$	999	109	$2_{(0/4)}$	1122	111	$2_{(0/4)}$	1264

low sparsity of NESTA's solutions (see Figures 7.1 and A.1–A.3 (b)). Thus, with increasing accuracy NESTA seems to fail at sufficiently decreasing the ℓ_1-norm of its iterates and satisfying the conditions ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$). Nevertheless, the results of the low precision problems and the results reported in [13, 157] indicate that NESTA is a very robust method regarding changes of the dynamic range.

At low and middle precision the FPC-BB method outperforms its regular FPC version and converges much faster. For fixed dynamic range the performance of both approaches degrades as the stopping tolerance is reduced, requiring about 6–18 times more iterations. Furthermore, FPC-BB did not converge for almost all high precision examples. FPC-AS generally performed very well; it takes about 2.3–4.9 times as many A- and $A^\top$-calls at 80 dB than at 20 dB. For dynamic range $d \in \{20, 40, 60\}$ about 2.8–5.6 times as many A- and $A^\top$-calls are used to compute a very accurate solution as to compute a low precision solution, whereas SNF-L1 only needs 1.8–3.7 times more calls. According to Figure 7.1 all variants of FPC need about 200–400 A-calls until they start to reliably identify zero components of x^*. While FPC-BB and FPC-AS manage to detect all zero components within 200 A-calls, FPC requires a large number of additional A-calls to find the correct sparsity pattern. Besides SPGL1, FPC-AS is the only tested method that detects small nonzero components of x^* as fast as SNF-L1.

FISTA and the PD method are outperformed by most other algorithms. For fixed dynamic range, FISTA needs about 5.5–60 times as many A- and $A^\top$-calls to compute a high precision solution as to calculate a low precision solution. In our experiments, FISTA seems to be mainly attractive for problems with low dynamic range and low accuracy. As SPGL1, the



(a) SNF-L1

(b) FISTA

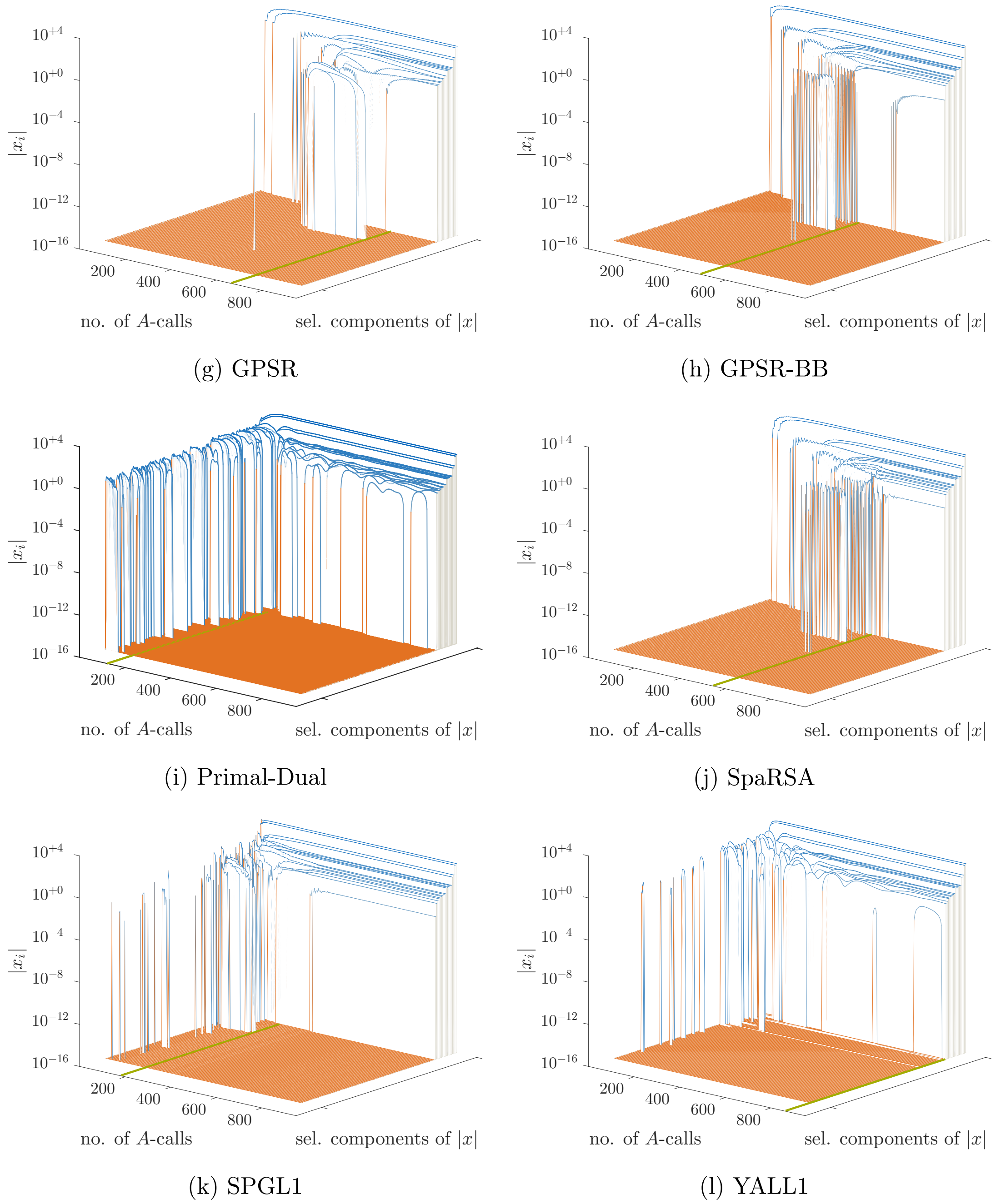
(c) FPC

(d) FPC-BB

(e) FPC-AS

(f) NESTA

Figure 7.1.: Change of the absolute value of 128 randomly chosen components of the iterate of a single run with dynamic range 40 dB with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero components of the high precision solution x^* within the sample (the following NESTA configuration was used: $\mu = 10^{-8}$, $T = 4$).



(g) GPSR

(h) GPSR-BB

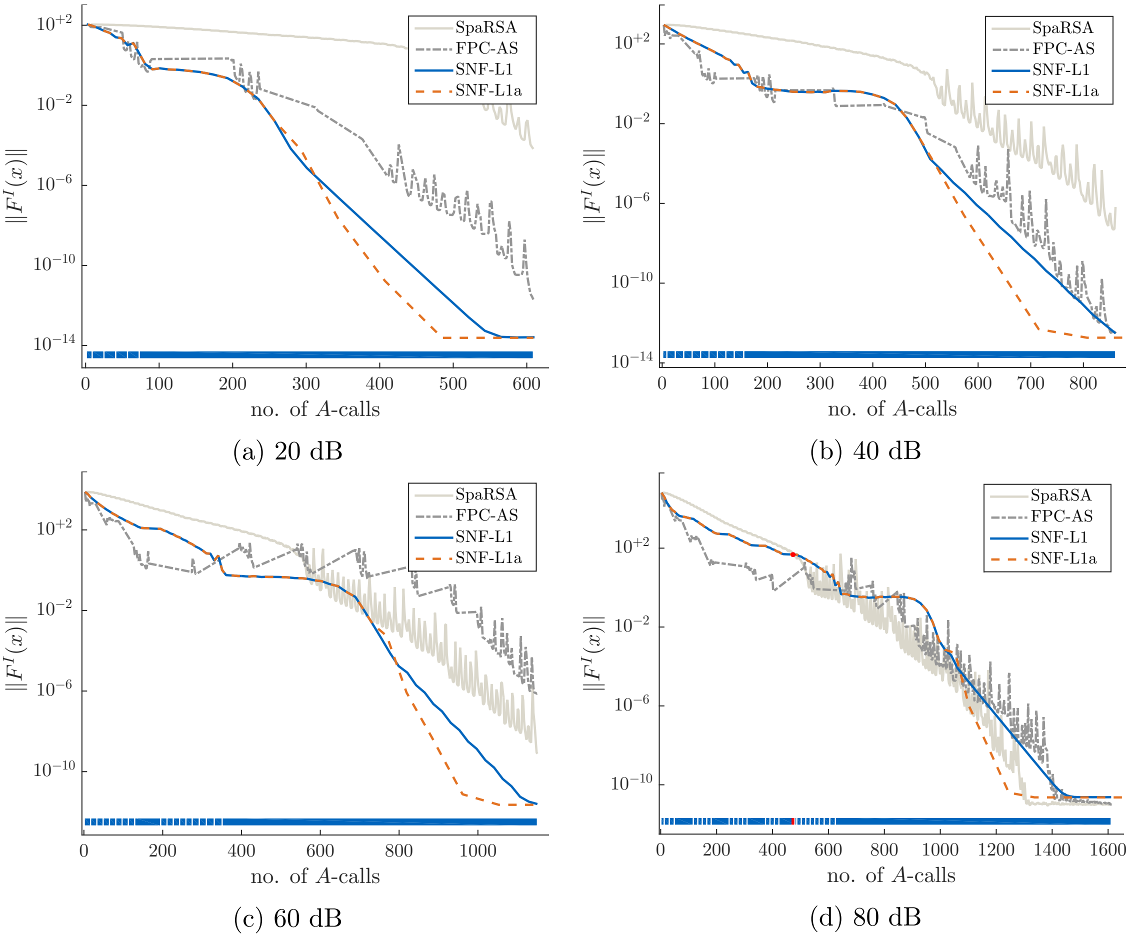
(i) Primal-Dual

(j) SpaRSA

(k) SPGL1

(l) YALL1

Figure 7.1.: Change of the absolute value of 128 randomly chosen components of the iterate of a single run with dynamic range 40 dB with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero components of the high precision solution x^* within the sample.



(a) 20 dB

(b) 40 dB

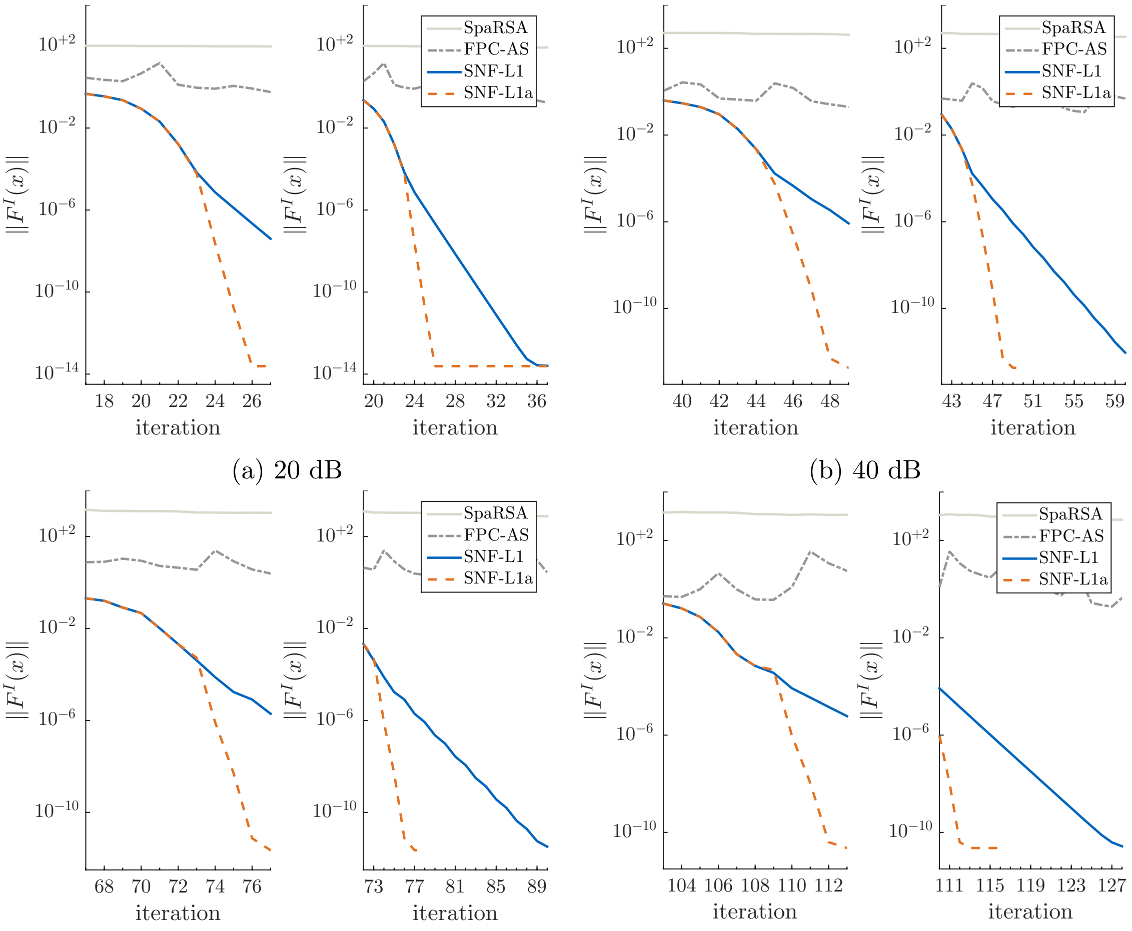
(c) 60 dB

(d) 80 dB

Figure 7.2.: Change of residual with respect to the total number of A- and $A^\top$-calls. The blue bar at the bottom of each plot visualizes the continuation scheme of SNF-L1. Each single box depicts one specific continuation phase with fixed μ_j. Shrinkage steps are marked in red color.

primal-dual method did not converge in the high precision examples. Usually, it requires about 10 times as many A-calls for the middle precision examples as for the low precision problems with $\varepsilon = 1$. Although the Figures 7.1 and A.1–A.2 (i) show that the PD method quickly detects all zero components of x^* for dynamic range $d \in \{20, 40, 60\}$, both FISTA and the PD method fail in successfully locating the small nonzero components of the optimal solution x^*.

The performance of the YALL1 algorithm rapidly decreases with increasing dynamic range. While YALL1 achieves good results in the 20 dB case, it performs rather poor in the examples with higher dynamic range. More specifically, YALL1 takes about 18–21 times as many A applications on the 40 and 60 dB signal than on the 20 dB signal. Furthermore, it did not converge in the 80 dB examples (see Table 7.5). The Figures A.2 and A.3 (l) in the appendix also indicate that YALL1 does not correctly detect all zero components of x^*.



(a) 20 dB

(b) 40 dB

(c) 60 dB

(d) 80 dB

Figure 7.3.: Convergence of the residual with respect to the number of iterations. Each subfigure contains two images that depict the same trial for different intervals of iterations.

The proposed method SNF-L1 compares quite positively to the other solvers. It performs especially well at low dynamic range. The Tables 7.2 and 7.3 demonstrate the efficiency of SNF-L1 and its competitiveness in this regime. Although SNF-L1 generally is a bit more sensitive to increasing dynamic range than, e.g., GPSR-BB or SpaRSA, its performance consistently stays competitive with the other methods. Moreover, our experiments in the Tables 7.2–7.5 show the particular strength of SNF-L1 in efficiently computing high accuracy solutions.

We now investigate the local convergence properties in some more detail. Since SNF-L1 terminates the CG iteration quite early, we also consider SNF-L1a, a version of SNF-L1 where the stopping criteria for the inner CG solves are adaptively adjusted to enforce more accurate solves in the final phase of convergence. To achieve this, SNF-L1a applies a simple adaptive update rule for the CG parameters. If the residual $\|F^{\Lambda_k}(x^k)\|$ falls below a certain tolerance we adjust CG-maxit and CG-tol appropriately and solve the Newton system with higher accuracy. For SpaRSA, FPC-AS, SNF-L1 and SNF-L1a, we investigate the change of the current residual $\|F^I(x^k)\|$ depending on the number of A- and $A^\top$-calls (Figure 7.2) and depending on the total number of iterations (Figure 7.3). We perform a single trial and consider dynamic ranges $d \in \{20, 40, 60, 80\}$. The results are summarized in the Figures 7.2 and 7.3.

The following effects can be observed. Independently of the dynamic range, both SNF-L1 and SNF-L1a provide competitive results and generally need much less A- and $A^\top$-calls than SpaRSA or FPC-AS to compute very accurate solutions. Figure 7.3 shows that the adaptive choice of the CG parameters results in local superlinear convergence of SNF-L1a, as predicted by the theory. Due to the coarse accuracy used in SNF-L1, superlinear convergence cannot be expected here and thus our original SNF-L1 implementation converges at a (good) linear rate. We also see that adaptivity as implemented in SNF-L1a can be a means for further increasing the performance of SNF-L1, since in the considered cases, SNF-L1a generally requires even less A- and $A^\top$-calls than SNF-L1, although the system (7.1.3) is solved with much higher accuracy. Thus, Figures 7.2 and 7.3 clearly confirm the potential and the efficiency of the SNF-L1 method in the high precision regime.

We conclude this section with a short discussion of the filter. Typically, in consistency with our observations regarding SNF-L1's rate of convergence, only few Newton iterates are rejected by the filter and the acceptance condition. To be more precise, our experiments and the results in Table 7.6 indicate that unacceptable Newton iterations mainly occur among trials with high dynamic range $d = 80$ dB and at low precision (see also Figure 7.2). Moreover, as shown in Table 7.6, the additional adaptive scheme for the step size parameter τ_k further stabilizes the SNF-L1 method and reduces the overall number of proximal gradient steps and A- or $A^\top$-calls. Hence, in the convex quadratic example, our numerical results demonstrate that the SNF-L1 method finally turns into a locally fast converging, pure semismooth Newton method.

7.2. Nonconvex ℓ_1-problems with Student's-t penalty

One particular strength of Algorithm 2 is that it is applicable to nonconvex problems. To evaluate the semismooth Newton method for nonconvex problems, we replace the Gaussian noise by errors with a Student's-t distribution, which is heavy-tailed and thus generates more outliers. It is known that least squares functionals $\|Ax - b\|^2$ are tailored to Gaussian noise, while a suitable misfit measure for data contaminated by Student's-t errors is given by [4, 2, 3] $\sum_{i=1}^m \psi([Ax - b]_i)$ where

$$\psi : \mathbb{R} \to \mathbb{R}, \quad \psi(y) := \log\left(1 + \frac{y^2}{\nu}\right)$$

Table 7.7.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 20 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-T	354	354	906	906	1280	1280	1888	1888	2323	2323
FPD	716	574	3283	1869	5452	4174	DNC	17087	DNC	DNC
FPD-BB$_\mu$	554	1996	2372	2082	4337	4327	DNC	DNC	DNC	DNC
FPD-BB$_\nu$	752	631	3411	1648	6371	6106	DNC	DNC	DNC	DNC

Table 7.8.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 40 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-T	1298	1298	2651	2651	3134	3134	3956	3956	4545	4545
FPD	4365	4788	9731	7785	14255	11495	DNC	DNC	DNC	DNC
FPD-BB$_\mu$	2252	3405	4369	3608	9070	7800	DNC	DNC	DNC	DNC
FPD-BB$_\nu$	448	703	1700	852	6528	5249	DNC	DNC	DNC	DNC

is the Student's-t penalty function with the degrees of freedom parameter $\nu > 0$ (see Figure 7.4 (a)). The function ψ is *nonconvex* and we consider the following nonconvex problem:

$$\min_{x\in\mathbb{R}^n} \sum_{i=1}^{m} \psi([Ax-b]_i) + \mu\|x\|_1. \tag{7.2.1}$$

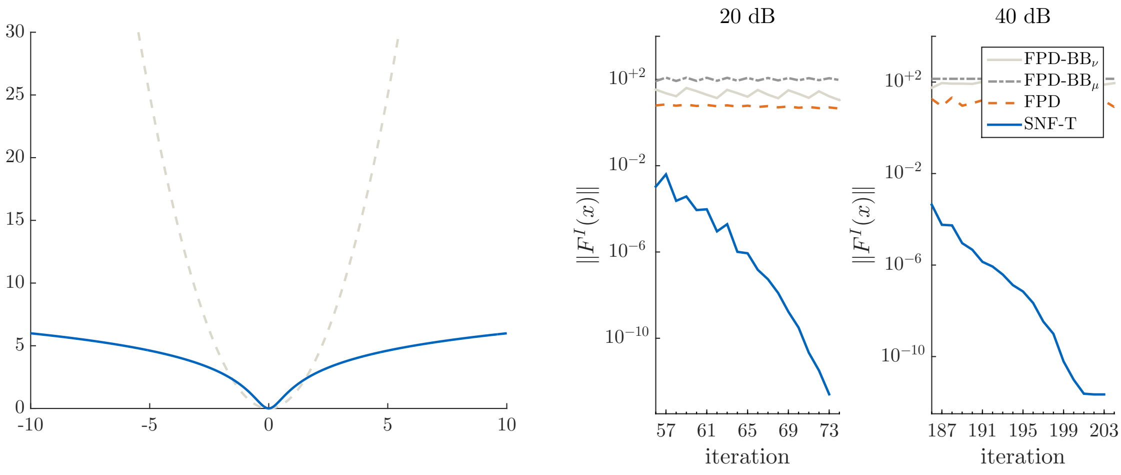
For more information about robust inversion, Student's-t approaches and related applications we refer to [2, 1] and the references therein.

In the following, we discuss the behavior of Algorithm 2 for the nonconvex problem (7.2.1). Based on an extension of the test framework of the convex example we compare the globalized semismooth Newton method with variants of Algorithm 1 and FPC-BB.

7.2.1. Algorithms and implementational details

In this section we list implementational details of the semismooth Newton method for (7.2.1) and describe the setting of the generalized fixed point descent (FPD) methods.

Semismooth Newton Method (SNF-T). We will refer to the Student's-t version of Algorithm 2 as SNF-T. SNF-T inherits its structure and concepts from SNF-L1 (section 7.1.1). As SNF-L1a, it implements an adaptive strategy for the CG parameters. Thus, the system (7.1.3) is solved with higher accuracy when the current residual is small enough. The parameters of the additional filter conditions (4.2.18) and (4.2.19) are set to $\alpha_1 = \alpha_2 = 5$,



(a)

(b)

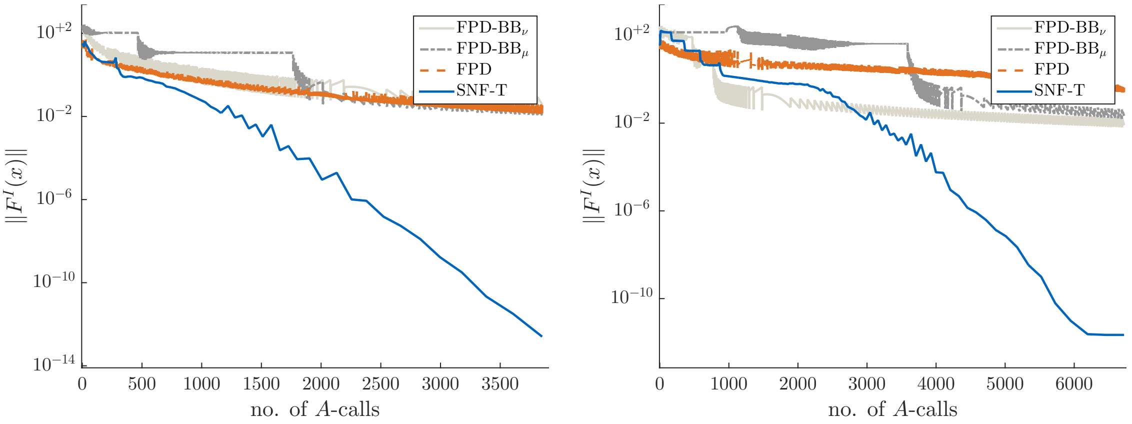
Figure 7.4.: In subfigure (a), a plot of the Gaussian (- -) and the Student's-t penalty function (-) with $\nu = 0.25$ is shown. Subfigure (b) shows the convergence of the residual with respect to the number of iterations. The left figure contains the results for a single run with 20 dB; on the right, dynamic range $d = 40$ dB is considered.

$\alpha_3 = 10^{-1}$ and $\eta = 0.8$. Due to the different scale invariance properties, continuation is not performed with respect to the regularization parameter μ but for the degree of freedom ν, i.e., μ is kept fixed throughout the iteration process. We used the same update formula as in the convex case except the damping factor C_0^{cont} was set to

$$C_0^{\text{cont}} = \min\left\{0.1, 2.2 \cdot (\|b\|_\infty/\nu^2)^{-\frac{1}{3}}\right\}$$

and $C_{\max}^{\text{cont}}$ was set to 20. Finally, $x^0 = A^\top b$ was used as initial point and we worked with fixed parameter matrices of the form $\Lambda = \tau^{-1}I$ and $\tau = 6$. All remaining parameters were not changed.

Fixed point descent methods (FPD). FPD is an implementation of the globally convergent proximal gradient method, Algorithm 1. The parameters for the quasi-Armijo condition were set to $\beta = 0.1$, $\gamma = 0.1$ and we used $\Lambda = \tau^{-1}I$, $\tau = 6$. Furthermore, we tested two variants of FPD that are based on FPC-BB. FPD-BB$_\mu$ uses a continuation strategy for μ and the quasi-Armijo condition is substituted by a Barzilai-Borwein framework with a nonmonotone linesearch technique. The continuation scheme and the choice of the BB steps and parameters were adopted from FPC-BB. As SNF-T, FPD-BB$_\nu$ applies a continuation to the degree of freedom parameter ν. Again the scheme and concept of FPC-BB was used except the initial value ν_0 was set to $\max\{0.1 \cdot \|b\|_\infty, \nu\}$. All methods were initialized with $x^0 = A^\top b$.



(a) 20 dB

(b) 40 dB

Figure 7.5.: Change of the residual with respect to the total number of A- and $A^\top$-calls.

7.2.2. Numerical comparison

Our comparison is based on the test framework of the convex example. More precisely, the reference signal $\bar{x} \in \mathbb{R}^n$, $n = 512^2$, and the matrix $A \in \mathbb{R}^{m\times n}$, $m = n/8$, are generated as specified in section 7.1.3 and the input data $b \in \mathbb{R}^m$ is obtained by adding Student's-t noise with degree of freedom 4, that is rescaled by 0.1, to $A\bar{x}$. Now, as described in 7.1.3, several reference solutions are computed by SNF-T and the stopping criteria of the other methods are changed to ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$), respectively.

All algorithms were run with $\mu = 0.07$ and degree of freedom $\nu = 0.25$. We consider dynamic range $d \in \{20, 40\}$ and tolerances $\varepsilon \in \{1, 10^{-1}, 10^{-2}, 10^{-4}, 10^{-6}\}$. Again we report `DNC`, when convergence is not reached after a total number of $\#A_{\text{rel}} = 20000$ or $\#A_{\text{nat}} = 20000$ A-calls. Table 7.7 and 7.8 contain the mean values of $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ over 10 random trials; the best result from each column is shaded. The corresponding total CPU times can be found in Table A.5 and A.6 in the appendix.

At first, we observe that, due to the additional nonconvexity all algorithms need a relatively high number of A- and $A^\top$-calls to show convergence. As in the convex case, the Barzilai-Borwein methods FPD-BB$_\mu$ and FPD-BB$_\nu$ generally outperform their regular FPD version and converge faster. The continuation strategy of FPD-BB$_\nu$ proves to be very efficient at low precision and at larger dynamic range. However, in the 20 dB case, FPD and FPD-BB$_\mu$ achieve better results and converge faster than FPD-BB$_\nu$ with respect to the CPU time. Compared to SNF-T, FPD and FPD-BB$_\nu$ degrade much faster as the stopping tolerance is reduced, requiring about 8–30 times more iterations. Surprisingly, none of the other algorithms converged at the high precision examples.

The results in Table 7.7 and 7.8 demonstrate that SNF-T possesses similar convergence properties as SNF-L1 and is, again, especially well-suited for recovering high precision solutions. Figure 7.4 (b) and 7.5 strengthen this impression and illustrate that transition to local superlinear convergence is also achieved in the nonconvex setting. As in the convex example only few Newton iterates are rejected by the filter and consequently, due to the rare update of φ_k, we observed that the majority of the accepted Newton iterations satisfies the growth condition (4.2.19). The remaining iterates, which satisfy condition (4.2.18), are usually performed at the beginning or the end of a trial or after a proximal gradient step.

7.3. Group sparse least squares problems

In this paragraph, we evaluate the semismooth Newton method on convex group sparse problems of the following type

$$(\mathrm{GS}_\mu) \qquad \min_{x\in\mathbb{R}^n} \; f(x) + \mu \sum_{i=1}^{s} \|x_{g_i}\|_2, \quad f(x) := \frac{1}{2}\|Ax-b\|_2^2.$$

As usual, we assume that the different groups $g_i \subset \{1,...,n\}$, $i = 1,...,s$, form a disjoint partitioning of the set $\{1,...,n\}$. Our numerical comparison will be again based on the test framework presented in [13] and section 7.1. A detailed description of the construction of the subsampled data vector $b \in \mathbb{R}^m$, the measurement matrix $A \in \mathbb{R}^{m\times n}$, and of the test setting can be found in section 7.3.3.

In contrast to the ℓ_1-norm regularization, the group sparse penalty term

$$(7.3.1) \qquad \varphi(x) = \mu \sum_{i=1}^{s} \|x_{g_i}\|_2$$

allows to model and add information about the sparsity pattern of the solutions of problem (GS_μ). In particular, any solution $\bar{x} \in \mathbb{R}^n$ of the latter least squares problem will possess a certain group sparse structure, i.e., the components of $\bar{x}$ are clustered in different groups that are either zero or nonzero. The minimization problem (GS_μ) is also known as the *group lasso problem* [262] and is a specific example of a problem with so-called *joint sparsity constraints* [70, 235, 85]. Its effectiveness has been proven in various applications such as variable selection [262, 115], machine and multiple kernel learning [7, 115], or gene selection and logistic regression [144].

In the following, we compare Algorithm 2 with several state-of-the-art methods that were already considered in the convex ℓ_1-example. For instance, we will also investigate a variant of the SPGL1 method that solves a constrained version of the problem (GS_μ):

$$(\mathrm{GS}_\sigma) \qquad \min_{x\in\mathbb{R}^n} \sum_{i=1}^{s} \|x_{g_i}\|_2 \quad \text{s.t.} \quad \|Ax-b\|_2 \le \sigma.$$

By reusing the basic constructions of the ℓ_1-comparison, we build a similar, high dimensional test framework for group sparse problems. We first start with a more detailed discussion of the implementation of the semismooth Newton method for problem (GS_μ).

7.3.1. Algorithmic description

We will refer to the group sparse version of Algorithm 2 as SNF-GS. In the following, we describe the basic components of the SNF-GS method.

Λ-strategy. As in the ℓ_1-case, we will work with simple parameter matrices of the form

$$\Lambda_k := \tau_k^{-1} I, \quad \tau_k \in [\tau_m, \tau_M], \quad k > 1,$$

where the variable τ_k is again chosen to estimate the inverse Lipschitz constant of the gradient $\nabla f(x) = A^\top (Ax - b)$ and is updated according to (7.1.1). The initial value is set to $\tau_0 = 500$ and we used $\tau_m = 10^{-3}$, $\tau_M = 10^4$.

Newton system. As described in Example 4.2.18, the nonsmooth residual function F^{Λ_k} can be computed group-wise via

$$F^{\Lambda_k}(x^k)_{g_i} = \tau_k \nabla f(x^k)_{g_i} + \mathcal{P}_{B_{\|\cdot\|_2}(0,\mu\tau_k)}(x^k_{g_i} - \tau_k \nabla f(x^k)_{g_i}), \quad \forall\, i = 1, ..., s.$$

Moreover, setting $u^k := x^k - \tau_k \nabla f(x^k)$, we will work with the following generalized derivatives

$$M(x^k) := \tau_k (I - D(x^k)) \cdot A^\top A + D(x^k),$$

where the block-structured matrix $D(x^k)$ is uniquely determined by

$$D(x^k)_{[g_i g_j]} = 0, \quad D(x^k)_{[g_i g_i]} = \begin{cases} I & \text{if } \|u^k_{g_i}\|_2 \le \mu\tau_k, \\ \frac{\mu\tau_k}{\|u^k_{g_i}\|_2} I - \frac{\mu\tau_k}{\|u^k_{g_i}\|_2^3} u^k_{g_i} (u^k_{g_i})^\top & \text{if } \|u^k_{g_i}\|_2 > \mu\tau_k, \end{cases}$$

for all $1 \le i, j \le s$ and $i \ne j$. As usual, the most expensive part of the algorithm is the computation of the next Newton step s^k which again involves finding a solution of the linear system of equations

$$M(x^k) s^k = -F^{\Lambda_k}(x^k). \tag{7.3.2}$$

Let us emphasize that the matrix $M(x^k)$ is never build explicitly in our implementation. Instead, we utilize an iterative method which only requires matrix-vector multiplications to compute an approximate solution of the latter system. Here, in contrast to the SNF-L1 method, we directly solve the *full* and nonsymmetric system (7.3.2) with a GMRES method. Let us note that by introducing the index sets $\mathcal{A} = \mathcal{A}(x^k) := \{i : \|u^k_{g_i}\|_2 \le \mu\tau_k\}$ and $\mathcal{I} = \mathcal{I}(x^k) := \{i : \|u^k_{g_i}\|_2 > \mu\tau_k\}$ and using a group-wise block elimination, the system (7.3.2) can again be reduced to a symmetric and smaller system of equations that, for instance, can be solved with a conjugate gradient method. However, in our numerical experiments, we observed that the CG method needs a relatively large number of iterations to achieve convergence. Hence, we implemented a GMRES-based strategy to solve the system (7.3.2). Furthermore and similar to the ℓ_1-case, we also consider the regularized and numerically more robust formulation

$$[M(x^k) + \rho I] s^k = -F^{\Lambda_k}(x^k), \quad \rho = \rho(x^k) = \|F^{\Lambda_k}(x^k)\|.$$

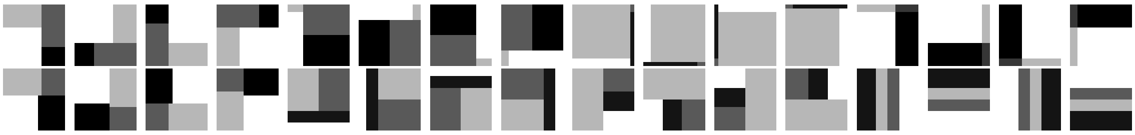


Figure 7.6.: Construction of the group pattern. Each image visualizes one basic group configuration for a 8×8 signal. Larger group patterns are generated by combining the depicted group blocks and randomly assigning the different groups.

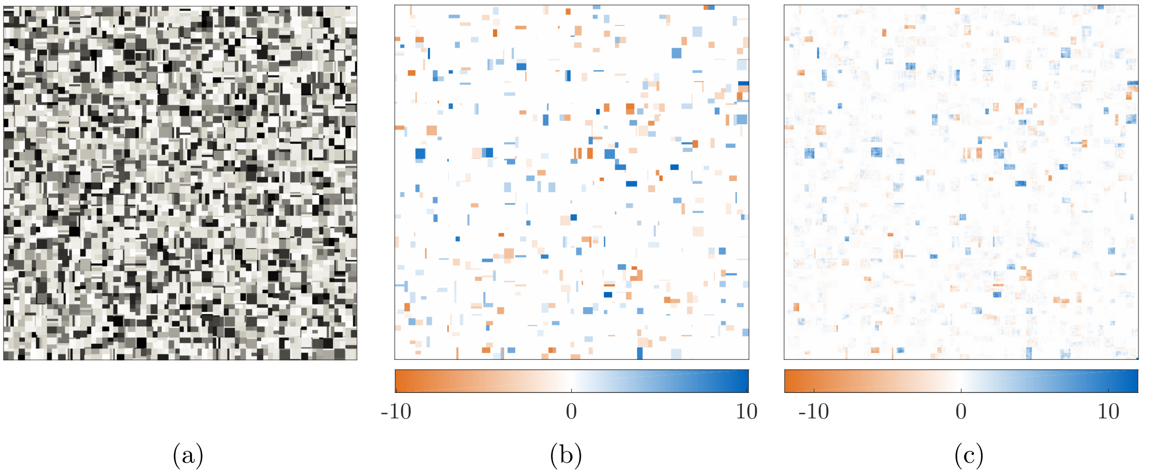


Figure 7.7.: Visualization of the experimental data. In subfigure (a), the group pattern of a 512×512 reference signal x^* with a total of 4096 groups is shown. Subfigure (b) illustrates the corresponding sparsity pattern of x^*. The generated signal has dynamic range 20 dB and 409 nonzero groups. In subfigure (c), a reconstruction of a noisy and undersampled version of the signal x^* is presented (we took 12.5% random discrete cosine measurements and added Gaussian noise with zero mean and variance $\sigma = 0.1$).

As in the ℓ_1-example, we will count the number of applications of the matrices A and $A^\top$ and use this number as a measure of efficiency and to compare the different algorithms. The terms A-call, $A^\top$-call, and $\mathcal{C}_A$ will again denote an application of the matrices A or $A^\top$ and will specify the complexity of an A- or $A^\top$-call, respectively. Since each inner step of the GMRES solver requires two A-calls, the complexity of a single Newton iteration of the SNF-GS method is given by $2\mathcal{C}_A + 2\mathcal{C}_A \cdot \texttt{gmres-iter}$ (two A-calls are used to calculate $F^{\Lambda_k}(x^k)$). To reduce the overall number of A-calls, we implemented an adaptive scheme that controls the maximum number of GMRES iterations and the accuracy of the GMRES method. Specifically, at the beginning, we set the relative tolerance GMRES-tol to 0.2 and the maximum number of iterations GMRES-max to 10. If the current residual $\|F^{\Lambda_k}(x^k)\|$ falls below a certain tolerance, we adjust GMRES-tol and GMRES-max and solve the system (7.3.2) with higher accuracy.

Continuation. Motivated by our previous experiments and results, we implemented a simple continuation framework for the regularization parameter μ. Here, we choose $\mu_0 =$

Table 7.9.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 20 dB and noise level $\bar{\sigma} = 0.1$.

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-GS	109	109	149	149	199	199	289	289	289	289
ADMM	211	68	387	240	757	450	1904	888	1904	1330
FISTA	130	68	294	202	696	568	3353	2160	3353	4716
FPD-BB	131	65	203	159	320	279	630	519	630	741
PD	87	91	118	148	190	259	1688	1252	1688	6795
SpaRSA	389	388	415	409	488	480	736	651	736	853
SPG-GS	84	70	129	114	213	194	DNC	DNC	DNC	DNC

$0.05 \cdot \|A^\top b\|_\infty$ and use the following update formula

$$\mu_{j+1} = \max\{\gamma_j \mu_j, \mu\}, \quad (0,1] \ni \gamma_j = \min\{0.01 \cdot \texttt{iter}, 0.8\},$$

where `iter` denotes the current iteration number. Thus, in contrast to SNF-L1, the damping factor γ_j increases linearly as the total number of iterations increases. Again, the current regularization parameter μ_j will be reduced whenever the number of inner iterations of a single continuation phase is larger than $C_{\max}^{\text{cont}} = 10$.

Filter and globalization. In our numerical tests, we experienced that the implemented continuation scheme and the adaptive step size strategy ensure convergence of semismooth Newton method even without the filter globalization. This is not completely surprising since a similar behavior can also be observed for the ℓ_1-problems in section 7.1; see, e.g., Table 7.6. Nevertheless, we integrated the filter mechanism as a safeguard in our implementation. Here, we again choose a filter function $\theta : \mathbb{R}^n \to \mathbb{R}^s_+$ of the type (4.2.8) with the following specifications:

$$\gamma_{\mathcal{F}} = 10^{-3}, \quad \mathcal{I}_j = g_j, \quad \forall\, j = \{1, ...s\}.$$

Hence, in this case, the filter function θ does also take account of the group-wise structure of the natural residual F^{Λ_k}.

Initial point and stopping criterion. As in the ℓ_1-case, we choose $x^0 = 0$ as default initial point and terminate SNF-GS if the current residual $\|F^{\Lambda_k}(x^k)\|$ is smaller than a given tolerance.

7.3.2. State-of-the-art methods

For our numerical comparison, we will mainly focus on methods that were already considered in the ℓ_1-test framework and that can also handle group sparse problems of the form (GS$_\mu$) or (GS$_\sigma$). In particular, we will reuse FISTA, SpaRSA, SPGL1, and the primal dual method for our experiments. (Here, we refer to the group sparse version of SPGL1 as SPG-GS). For each approach, additional code for the computation of the objective function and of the

Table 7.10.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 20 dB and noise level $\bar{\sigma} = 0.01$.

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-GS	209	209	283	283	402	402	561	561	686	686
ADMM	133	29	366	147	815	347	1495	799	1813	1258
FISTA	302	16	992	270	3722	824	DNC	6342	DNC	DNC
FPD-BB	518	93	1401	195	2465	1063	3955	3020	4647	5000
PD	171	26	271	261	363	360	451	476	485	567
SpaRSA	445	452	613	473	1251	593	2220	1670	2730	3066
SPG-GS	142	137	440	184	1355	340	DNC	DNC	DNC	DNC

group sparse proximity operator is provided. Moreover, we also use the same modifications and parameter settings as specified in section 7.1.2.

Although SpaRSA's continuation scheme has been developed for ℓ_1-problems, it also led to a significant speedup in the group sparse case. Thus, in our tests, SpaRSA is run with continuation and all continuation parameters were set to default except the number of continuation steps was set to 40 and the initial regularization parameter μ_0 was changed to

$$\mu_0 = \max_{i=1,\dots,s} \|[A^\top b]_{g_i}\|_2.$$

In the following, we derive an alternating direction method for the constrained problem (GS_σ) and briefly describe a variant of the FPC-BB algorithm for the group sparse least-squares problem (GS_μ).

Alternating Direction Method of Multipliers (ADMM). In [61], Deng et al. present and provide an extension of the YALL1 package for group sparse problems. Unfortunately, their algorithmic framework and their online code is not directly applicable to the problems considered in this section. Thus, we implemented our own version of ADMM to solve the constrained problem (GS_σ). Next, we give an overview of ADMM and of our proposed version; for more general information about ADMM, see, e.g., [91, 71]. The classical alternating direction method goes back to Glowinski and Marrocco [92] and Gabay and Mercier [89] and is designed for problems of the form

$$\min_{x,y} \ \varphi(x) + \psi(y) \quad \text{s.t.} \quad Ax + By = c,$$

where $\varphi : \mathbb{R}^n \to (-\infty, +\infty]$ and $\psi : \mathbb{R}^m \to (-\infty, +\infty]$ are convex, proper, and lower semicontinuous functions and $A \in \mathbb{R}^{\ell \times n}$, $B \in \mathbb{R}^{\ell \times m}$, $c \in \mathbb{R}^\ell$ are given. The constraint $Ax + By = c$ couples the variables x and y and usually arises from a reformulation of a general convex composite problem. The basic idea of the alternating direction method is to

Table 7.11.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 40 dB and noise level $\bar{\sigma} = 0.1$.

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-GS	257	257	386	386	453	453	556	556	668	668
ADMM	377	192	778	380	1130	555	1416	901	1807	1247
FISTA	793	293	3028	876	5528	2146	9220	8276	15117	DNC
FPD-BB	1016	233	1728	903	2335	1489	2832	2672	DNC	3844
PD	249	253	333	312	403	383	510	648	729	1156
SpaRSA	516	478	982	530	1435	844	1811	1732	2337	2620
SPG-GS	283	147	812	275	1266	648	DNC	DNC	DNC	DNC

perform a successive and alternating minimization of the associated augmented Lagrangian

$$\mathcal{L}(x, y, \lambda) := \varphi(x) + \psi(y) + \langle \lambda, Ax + By - c \rangle + \frac{\beta}{2} \|Ax + By - c\|_2^2,$$

where $\lambda \in \mathbb{R}^\ell$ denotes the Lagrange multiplier and $\beta > 0$ is a penalty parameter. Now, for the group sparse problem (GS_σ), we introduce the auxiliary variables $z = x$ and $t = \sigma$. This leads to the following formulation:

$$\min_{x,z,t} \; \mu \sum_{i=1}^{s} \|x_{g_i}\|_2 \quad \text{s.t.} \quad (Az - b, t) \in \operatorname{epi} \|\cdot\|_2, \quad x - z = 0, \quad t - \sigma = 0.$$

In our implementation, we also use different penalty parameters for the constraints $x = z$ and $t = \sigma$, i.e., setting $y = (z, t) \in \mathbb{R}^{n+1}$ and $\lambda = (\lambda_z, \lambda_t) \in \mathbb{R}^{n+1}$, we consider the Lagrangian

$$\mathcal{L}(x, y, \lambda) := \varphi(x) + \psi(y) + \langle \lambda, y - (x^\top, \sigma)^\top \rangle + \frac{\beta_1}{2} \|x - z\|_2^2 + \frac{\beta_2}{2} (t - \sigma)^2,$$

where $\psi(y) = \psi(z, t) = \iota_{\operatorname{epi} \|\cdot\|_2}(Az - b, t)$ and φ is the group sparse penalty term as defined in (7.3.1). In this case, the update steps of the alternating direction method are given by

$$\begin{aligned} x^{k+1} &= \operatorname*{arg\,min}_{x \in \mathbb{R}^n} \; \mathcal{L}(x, y^k, \lambda^k) \;\;= \operatorname{prox}_\varphi^{\beta_1 I}(z^k + \beta_1^{-1} \lambda_z^k), \\ y^{k+1} &= \operatorname*{arg\,min}_{y \in \mathbb{R}^{n+1}} \; \mathcal{L}(x^{k+1}, y, \lambda^k) = \operatorname{prox}_\psi^B \left(\begin{pmatrix} x^k \\ \sigma \end{pmatrix} - B^{-1} \lambda^k \right), \quad B = \begin{pmatrix} \beta_1 I & 0 \\ 0 & \beta_2 \end{pmatrix}, \\ \lambda^{k+1} &= \lambda^k + \gamma B (y^{k+1} - ((x^{k+1})^\top, \sigma)^\top). \end{aligned}$$

Clearly, the most complex part of each ADMM iteration is the computation of the proximity operator $\operatorname{prox}_\psi^B$. However, if the matrix A is an orthogonal projector and satisfies $AA^\top = I$, then Example 3.2.9 implies that $\operatorname{prox}_\psi^B$ has an explicit representation. In this situation, the total complexity of a single ADMM iteration reduces to two A-calls. In our implementation,

Table 7.12.: Total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ averaged over 10 independent runs with dynamic range 40 dB and noise level $\bar{\sigma} = 0.01$.

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-GS	401	401	476	476	788	788	1024	1024	1224	1224
ADMM	307	179	535	383	835	568	1468	936	1695	1303
FISTA	2282	165	5255	1178	12928	3439	DNC	DNC	DNC	DNC
FPD-BB	4733	308	7885	404	11291	4986	DNC	13372	DNC	DNC
PD	656	224	806	783	946	974	1179	1244	1254	1448
SpaRSA	528	518	2018	535	5583	978	12708	7738	DNC	15942
SPG-GS	337	297	2876	336	7200	665	14869	10166	DNC	DNC

we use the parameters $\gamma = 1.618$, $\beta_1 = 0.3/\texttt{mean}(b)$, $\beta_2 = 3/\texttt{mean}(b)$ and set $x^0 = 0$ and $\lambda^0 = 0$. Furthermore, no continuation was used for the penalty parameters β_1 and β_2.

Fixed point descent method (FPD-BB). As in the nonconvex ℓ_1-test framework, FPD-BB is an implementation of the globally convergent, proximal gradient scheme, Algorithm 1. Similar to FPC-BB, we use Barzilai-Borwein step sizes to build the parameter matrices $\Lambda_k = \tau_k^{-1} I$ and the quasi-Armijo condition is substituted by a nonmonotone line search technique. Again, we apply a simple continuation strategy to adapt the parameter μ and to accelerate convergence. More specifically, we set $\mu_0 = 0.05 \cdot \|A^\top b\|_\infty$ and after a fixed number of steps the new regularization parameter μ_{j+1} is calculated via $\mu_{j+1} = \max\{0.7 \cdot \mu_j, \mu\}$.

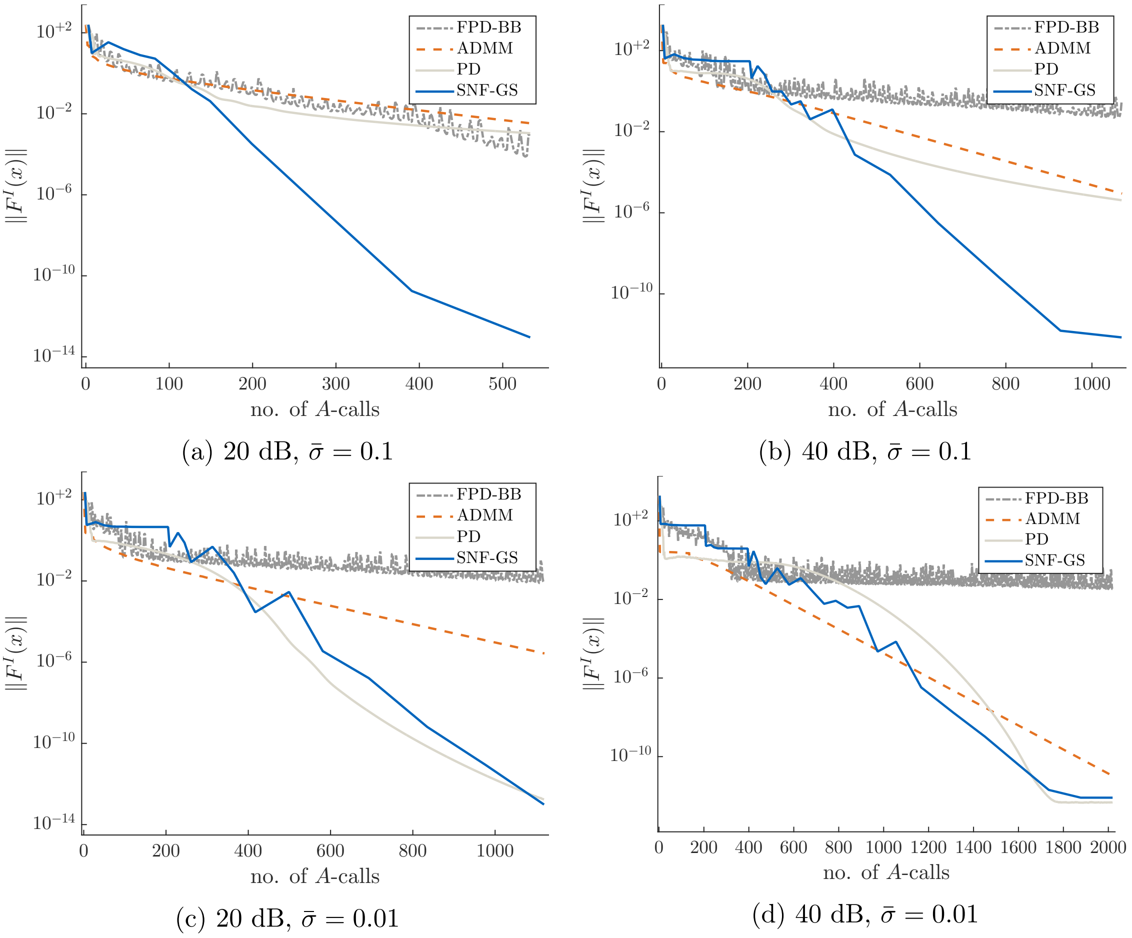
7.3.3. Numerical comparison

Again, our test framework is based on the NESTA package [13] and on the experiments in section 7.1. At first, we generate a random group pattern by combining a suitable number of basic and fixed, lower dimensional group configurations and by randomly assigning the different groups. A visualization of the construction of the group pattern and a specific example can be found in Figure 7.6 and 7.7. Next, we generate a group sparse reference signal $\bar{x} \in \mathbb{R}^n$ of length $n = 512^2 = 262144$ with a total of $s = \lceil n/64 \rceil = 4096$ groups and with $k = 409$ nonzero groups. Similar to the ℓ_1-test framework, the k different groups $i \in \{1, ..., s\}$ are randomly chosen and the nonzero components of $\bar{x}$ are determined via

$$\bar{x}_{g_i} = \eta_1(i) 10^{d\eta_2(i)/20} \cdot \mathbb{1}_{g_i},$$

where $\eta_1(i) \in \{-1, +1\}$ is a random sign and $\eta_2(i)$ is uniformly distributed in $[0, 1]$. The matrix $A \in \mathbb{R}^{m \times n}$, $m = n/8$, is based on random cosine measurements and is constructed as specified in section 7.1.3. Finally, the input data $b \in \mathbb{R}^m$ is obtained by adding Gaussian noise with standard deviation $\bar{\sigma} \in \{0.1, 0.01\}$ to $A\bar{x}$.

Since ADMM and SPG-GS solve the constrained group sparse problem (GS_σ) while all other algorithms solve the unconstrained problem (GS_μ), we again need to compute a corresponding pair of parameters σ and μ first. Here, we simply reuse the procedure described in



(a) 20 dB, $\bar{\sigma} = 0.1$

(b) 40 dB, $\bar{\sigma} = 0.1$

(c) 20 dB, $\bar{\sigma} = 0.01$

(d) 40 dB, $\bar{\sigma} = 0.01$

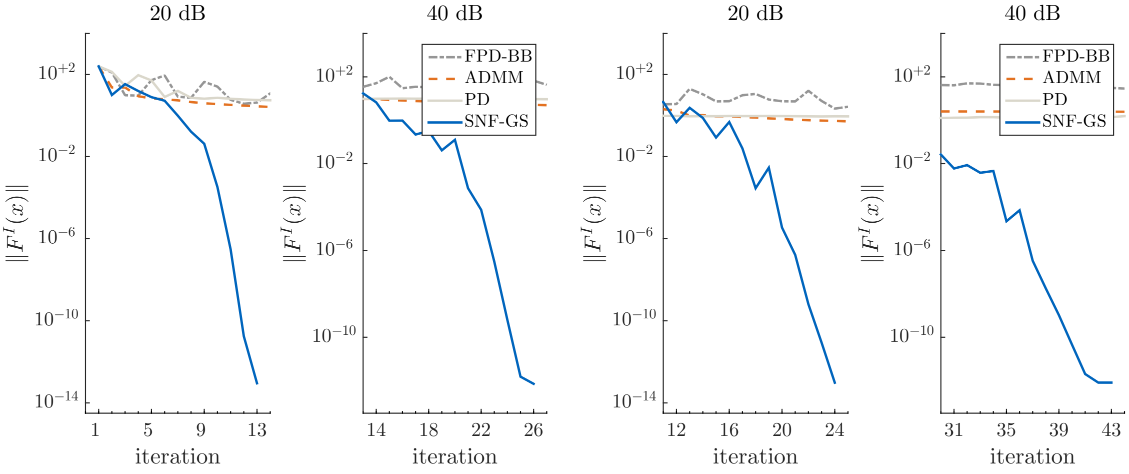
Figure 7.8.: Change of the residual with respect to the total number of A- and $A^\top$-calls.

section 7.1.3 to obtain comparable problem settings. In particular, as an intermediate step, SNF-GS is used to compute a high precision solution x^* of the group sparse problem (GS_μ) that satisfies

$$\|F^I(x^*)\| \leq 10^{-12}.$$

Afterwards and identical to the convex ℓ_1-test framework, the SNF-GS method is run with different stopping tolerances to generate a series of reference solutions. Moreover, the stopping criteria of the other methods are again changed to ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$).

In the following test, we consider the dynamic ranges $d \in \{20, 40\}$, the noise level $\bar{\sigma} = \{0.1, 0.01\}$, and the tolerances $\varepsilon \in \{1, 10^{-1}, 10^{-2}, 10^{-4}, 10^{-6}\}$. Since an application of A or $A^\top$ again corresponds to the evaluation of a `dct` or `idct` function, our numerical comparison will focus on the total number of A- and $A^\top$-calls $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$, respectively. We report `DNC` (did not converge) if convergence is not reached after a total number of $\#A_{\text{rel}} = 20000$ or $\#A_{\text{nat}} = 20000$ calls. The Tables 7.9, 7.10, 7.11, and 7.12 contain the mean values of the numbers $\#A_{\text{rel}}$ and $\#A_{\text{nat}}$ over 10 random trials; the two best results from each column are shaded. The mean values of the corresponding, total CPU times t_{rel} and t_{nat} are summarized



(a) $\bar{\sigma} = 0.1$

(b) $\bar{\sigma} = 0.01$

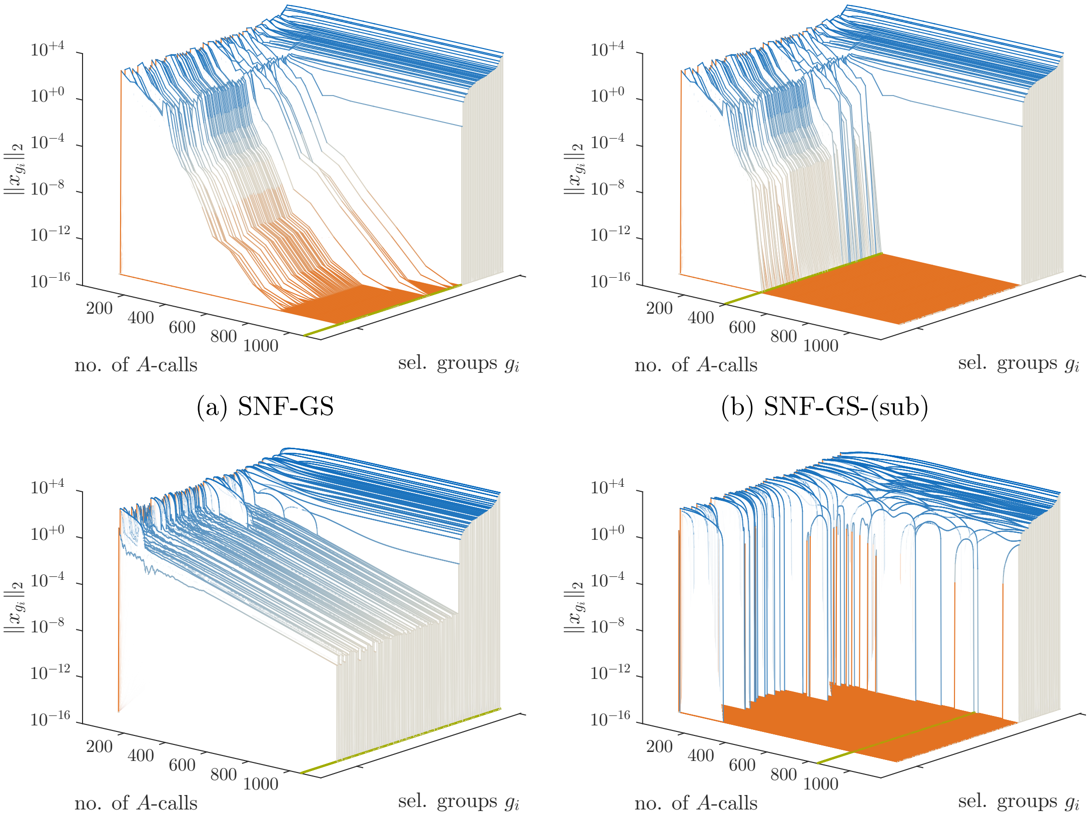
Figure 7.9.: Convergence of the residual with respect to the number of iterations. The left figure contains the results for a single run with $\bar{\sigma} = 0.1$; on the right, the noise level $\bar{\sigma} = 0.01$ is considered.

in the Tables A.7–A.10 and can be found in section A.4.3 in the appendix. In Figure 7.10 we visualize the change of the ℓ_2-norm of 128 randomly chosen group-components of the iterate with respect to the number of A- and $A^\top$-calls. Each plot shows the development of the iterate for a single run with dynamic range $d = 40$ dB. Additionally, the maximum number of A- and $A^\top$-calls that is required to detect all zero groups of the optimal solution x^* is marked with a green line. A similar visualization for dynamic range $d = 20$ dB is provided in the appendix, see Figure A.4.

SPG-GS has a similar behavior as its ℓ_1-version SPGL1 and performs especially well on the low and middle precision problems. However, for $\bar{\sigma} = 0.01$ and using the relative stopping criterion $(\mathcal{C}_{\text{rel}})$, it requires up to 10 times as many A-calls as SNF-GS to show convergence. Similar to the ℓ_1-experiments, SPG-GS does not converge in the high precision examples. Figure 7.10 (h) also again demonstrates that SPG-GS quickly detects the correct zero and nonzero groups of the solution x^*.

The proximal gradient schemes FISTA and FPD-BB are outperformed by the other algorithms in most cases. More specifically, the performance of both methods quickly degrades as the stopping tolerance ε is reduced. In general, FISTA needs 10–100 times as many A-calls to compute a very accurate solution as to compute a low precision solution and thus, it often does not converge in the harder examples with smaller noise level $\bar{\sigma} = 0.01$. While FPD-BB achieves good results in the experiment with dynamic range 20 dB and $\bar{\sigma} = 0.1$ (see Table 7.9), its overall performance deteriorates in the other test examples. Both methods require a relatively large number of steps and A-calls to detect the correct group sparsity pattern. Surprisingly and in contrast to our previous observations, the performance of both methods also significantly depends on the chosen stopping criterion.

For group sparse problems, the overall performance of SpaRSA is not as stable as in the



(a) SNF-GS

(b) SNF-GS-(sub)

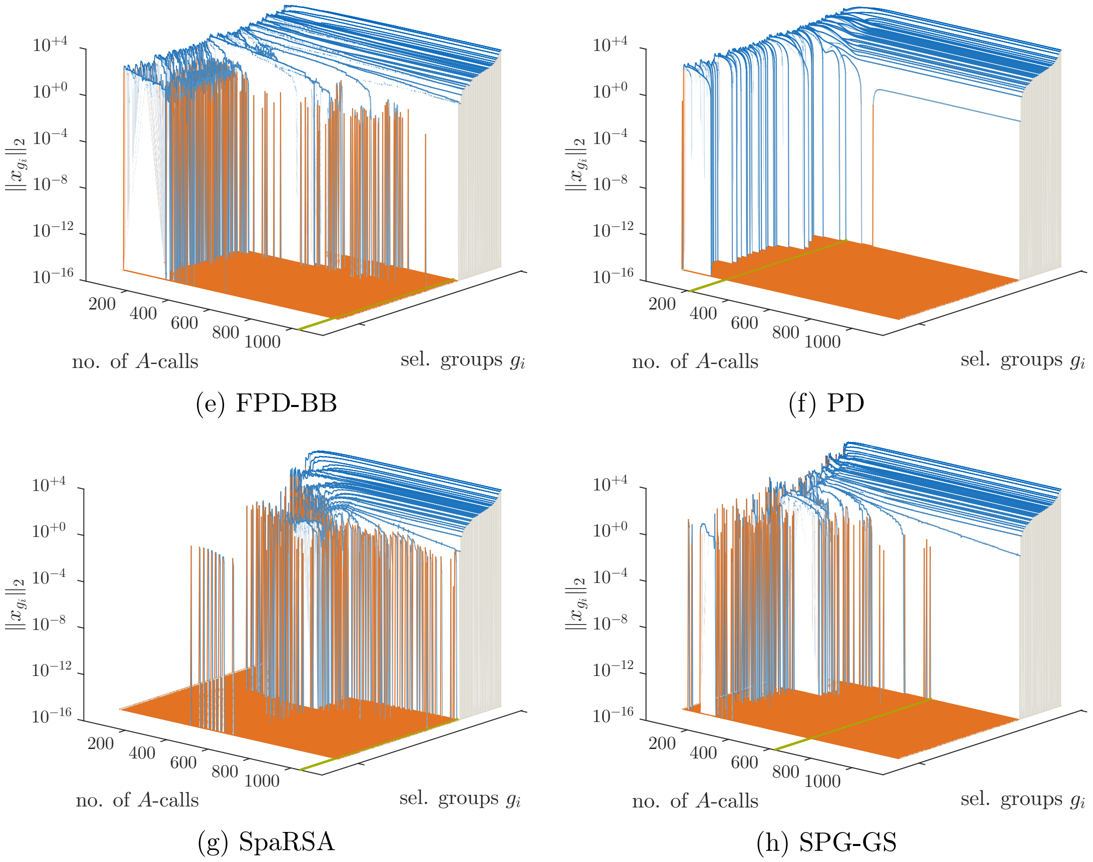
(c) ADMM

(d) FISTA

Figure 7.10.: Change of the ℓ_2-norm of 128 randomly chosen groups of the iterate of a single run with dynamic range 40 dB and $\bar{\sigma} = 0.1$ with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero groups of the high precision solution x^* within the sample.

ℓ_1-case. For fixed dynamic range SpaRSA requires about 3.7–6 times as many A-calls at $\bar{\sigma} = 0.01$ than at the more noisy case $\bar{\sigma} = 0.1$. As FPD-BB, it needs a relatively large number of A applications to capture the active and inactive groups of x^*, see Figure 7.10 (g) and A.4 (g).

In contrast to the ℓ_1-experiments in section 7.1, the primal-dual method achieves very good results. In particular, it is very efficient for the test problems with lower noise level $\bar{\sigma} = 0.01$, see also Figure 7.8. Here, the PD method takes about 3 times as many A-calls on the 40 dB signal than on the 20 dB examples (see Table 7.10 and 7.12). Similar to the other tested methods, its performance generally depends on the chosen stopping criterion (we refer to Table 7.9 for details). The Figures 7.10 (f) and A.4 (f) demonstrate that the PD method quickly and efficiently captures the correct sparsity pattern of the solution x^*. However, it usually requires more iterations and A applications to locate groups with small nonzero components. ADMM achieves very stable results throughout the test framework and performs especially well on problems with lower noise level. For fixed dynamic range



(e) FPD-BB

(f) PD

(g) SpaRSA

(h) SPG-GS

Figure 7.10.: Change of the ℓ_2-norm of 128 randomly chosen groups of the iterate of a single run with dynamic range 40 dB and $\bar{\sigma} = 0.1$ with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero groups of the high precision solution x^* within the sample.

and fixed noise level $\bar{\sigma}$, it requires about 5.5–13 times as many A-calls to compute a high precision solution than to compute a low precision solution, whereas SNF-GS only needs 2.6–3.2 times as many A-calls. This rather slow convergence can be seen in Figure 7.8 and is also reflected in the slow reduction of the zero groups (see also Figure A.4 (c)). Nevertheless, our results show that ADMM is very robust regarding changes of the dynamic range or of the noise level $\bar{\sigma}$.

The results in the Tables 7.9–7.12 show that SNF-GS compares very positively to the other algorithms and performs particularly well at the higher noise level $\bar{\sigma} = 0.1$. Again, our tests confirm that the semismooth Newton method is very efficient in computing high precision solutions. Moreover, the Figures 7.8 and 7.9 clearly demonstrate that transition to local, q-superlinear convergence can also be observed in the group sparse setting. In constrast to the ℓ_1-problems, SNF-GS only approximately detects the correct sparsity pattern and requires a comparably high number of A-calls to reduce the zero group components. This effect is mainly caused by our regularization strategy that is applied to the full system (7.3.2);

see section 7.3.1. Hence, we also tested SNF-GS-(sub), a simple variant of SNF-GS, that performs a subspace correction step on the active set $\mathcal{A} = \mathcal{A}(x^k)$. As illustrated in Figure 7.10 and A.4 this adjustment clearly improves the overall quality of the iterates.

Thus, in summary, our experiments indicate the competitiveness of the globalized semismooth Newton method and demonstrate its particular strength in obtaining high accuracy solutions for different, convex, and nonconvex problems.

A. Additional material

A.1. The proximal gradient method: Global convergence

In the following, we prove global convergence of the proximal gradient method presented in section 4.2.2. The proof is essentially based on a result of Tseng and Yun [236] for a nonsmooth coordinate descent method. Let us also note that the proof uses similar arguments as the standard convergence analysis of the classical gradient descent method; see, e.g., [17, Proposition 1.2.1] or [239, Satz 7.7].

Proof A.1.1 (Proof of Theorem 4.2.10). Suppose that Algorithm 1 does not terminate within a finite number of steps and let the sequences $(x^k)_k$, $(\Lambda_k)_k$, and $(\sigma_k)_k$ be generated by Algorithm 1. As in the proof of Lemma 4.2.8, we have

$$\limsup_{\sigma \downarrow 0} \frac{\psi(x^k + \sigma d^k) - \psi(x^k)}{\sigma} - \gamma \Delta^k \leq -(1-\gamma)\|d^k\|^2_{\Lambda_k} < 0.$$

Hence, it holds $(\sigma_k)_k \subset (0,1]$ and $\psi(x^{k+1}) < \psi(x^k)$ for all $k \in \mathbb{N}$.

Now, let $\bar{x} \in \mathbb{R}^n$ be an arbitrary accumulation point of the sequence $(x^k)_k$ and let $(x^k)_K$ be a corresponding subsequence that converges to $\bar{x}$. Since the sequence $(\psi(x^k))_k$ is monotonically decreasing, it converges to some limit $\xi \in \mathbb{R} \cup \{-\infty\}$. On the other hand, the lower semicontinuity of ψ implies

$$\liminf_{K \ni k \to \infty} \psi(x^k) \geq \psi(\bar{x}).$$

Consequently, it holds $\xi \geq \psi(\bar{x}) \in \mathbb{R}$. Next, by using the Armijo step size rule, we have

$$\psi(x^0) - \psi(\bar{x}) \geq \sum_{k=0}^{\infty} \psi(x^k) - \psi(x^{k+1}) \geq \sum_{k=0}^{\infty} \sigma_k \gamma \lambda_m \|d^k\|^2.$$

This immediately yields $\sigma_k \|d^k\|^2 \to 0$, as $k \to \infty$. Now, suppose that $F^\Lambda(\bar{x}) \neq 0$ for some $\Lambda \in \mathbb{S}^n_{++}$. Due to the continuity of F^Λ and using the convergence $x^k \to \bar{x}$, $K \ni k \to \infty$, there exists $\ell \in K$ such that

$$\|F^\Lambda(x^k)\| \geq \frac{1}{2}\|F^\Lambda(\bar{x})\| > 0, \quad \forall\, k \in K,\ k \geq \ell.$$

Next, since the parameter matrices remain in a bounded set, we can utilize Remark 4.1.4, i.e., there exists $\underline{\lambda} = \underline{\lambda}(\lambda_m, \lambda_M, \Lambda) > 0$ such that

$$\text{(A.1.1)} \qquad \|d^k\| = \|F^{\Lambda_k}(x^k)\| \geq \underline{\lambda} \cdot \|F^\Lambda(x^k)\| > 0, \quad \forall\, k \in K,\ k \geq \ell.$$

In particular, this implies

$$\sigma_k \to 0, \quad K \ni k \to \infty.$$

Thus, there exists $\tilde{\ell} \in K$, $\tilde{\ell} \geq \ell$ such that $\sigma_k \leq \beta$ for all $k \in K$, $k \geq \tilde{\ell}$ and the Armijo condition in step **S2** of Algorithm 1 is not satisfied for $\beta^{-1}\sigma_k > \sigma_k$, i.e., it holds

$$\psi(x^k + \beta^{-1}\sigma_k d^k) - \psi(x^k) > \gamma\beta^{-1}\sigma_k\Delta^k.$$

We obtain

$$\begin{aligned}\gamma\Delta^k < \frac{\psi(x^k + \beta^{-1}\sigma_k d^k) - \psi(x^k)}{\beta^{-1}\sigma_k} \\ &\leq \frac{f(x^k + \beta^{-1}\sigma_k d^k) - f(x^k)}{\beta^{-1}\sigma_k} + \varphi(\mathrm{prox}_{\varphi}^{\Lambda_k}(u(x^k))) - \varphi(x^k) \\ &= \frac{f(x^k + \beta^{-1}\sigma_k d^k) - f(x^k)}{\beta^{-1}\sigma_k} - \nabla f(x^k)^\top d^k + \Delta^k\end{aligned}$$

Now, as in Lemma 4.2.8, we have $x^k + \beta^{-1}\sigma_k d^k \in \mathrm{dom}\ \varphi \subset \Omega$ for all $k \in K$, $k \geq \tilde{\ell}$. Finally, a first order Taylor expansion yields

$$(1-\gamma)\lambda_m\|d^k\| \leq -(1-\gamma)\frac{\Delta^k}{\|d^k\|} \leq \frac{f(x^k + \beta^{-1}\sigma_k d^k) - f(x^k)}{\beta^{-1}\sigma_k\|d^k\|} - \frac{\nabla f(x^k)^\top d^k}{\|d^k\|} = \frac{o(\sigma_k\|d^k\|)}{\sigma_k\|d^k\|}.$$

Taking the limit $K \ni k \to \infty$, this clearly leads to a contradiction to inequality (A.1.1) and we can conclude the proof of Theorem 4.2.10.

A.2. Second order conditions

In this section, we present a detailed proof of the second order necessary and sufficient conditions, which were introduced and discussed in section 5.2. The proof summarizes and recreates several significant results of Bonnans and Shapiro, [27, Theorem 3.45, 3.83, and Proposition 3.105], and rigorously "transfers" their argumentation to the convex composite setting. For the sake of completeness, let us also refer to Bonnans et al. [24, Theorem 3.1, 3.2, and 4.1; Section 5]. Let us note that this section is intended to complement our second order results on a theoretical level and to illustrate the application and interaction of the concepts of second order regularity, second order tangent sets, and higher order, parabolic epidifferentiability.

In the proof, we will utilize the following property of second order tangent sets, see [27, Proposition 3.34]. Let $S \subset \mathbb{R}^n$ be a convex set and let $x \in S$, $h \in T_S(x)$ be arbitrary. Then, it follows

$$T_S^{i,2}(x,h) + T_{T_S(x)}(h) \subset T_S^{i,2}(x,h) \subset T_{T_S(x)}(h)$$

and

$$T_S^2(x,h) + T_{T_S(x)}(h) \subset T_S^2(x,h) \subset T_{T_S(x)}(h). \tag{A.2.1}$$

Consequently, if $0 \in T_S^{i,2}(x,h)$, then we have $T_S^{i,2}(x,h) = T_S^2(x,h) = T_{T_S(x)}(h)$. (For instance, this is satisfied, when S is a polyhedral set).

Now, as usual, let $\phi : \mathbb{R}^n \to (-\infty, +\infty]$ be a convex, proper, and lower semicontinuous function and let $x \in \text{dom } \phi$, $h \in \mathbb{R}^n$ be given. Moreover, let us suppose that the epiderivative $\phi^{\downarrow}(x;h)$ is finite. Then, due to Example 2.1.5 and Lemma 2.5.3, it holds

$$\begin{aligned} T_{T_{\text{epi } \phi}(x,\phi(x))}(h, \phi^{\downarrow}(x;h)) &= \text{cl}\{T_{\text{epi } \phi}(x,\phi(x)) + \text{sp}[(h, \phi^{\downarrow}(x;h))]\} \\ &= \text{cl}\{\text{epi } \phi^{\downarrow}(x;\cdot) + \text{sp}[(h, \phi^{\downarrow}(x;h))]\}. \end{aligned}$$

Thus, combing (A.2.1) and Lemma 5.2.4, it follows

$$\text{cl}\{\mathcal{T} + \text{epi } \varphi^{\downarrow}(x;\cdot)\} \subset \text{epi } \varphi_-^{\downarrow\downarrow}(x;h,\cdot) \tag{A.2.2}$$

for any subset $\mathcal{T} \subset \text{epi } \varphi_-^{\downarrow\downarrow}(x;h,\cdot)$. (Here, we used the fact that the outer second order tangent set epi $\varphi_-^{\downarrow\downarrow}(x;h,\cdot)$ is a closed set). We now turn to the proof of Theorem 5.2.8.

Proof A.2.1 (Proof of Theorem 5.2.8). Let us start with the verification of the second order necessary conditions.

Proof of part (i). Let $\bar{x}$ be a local solution of problem $(\mathcal{P}_c)$. Then, it holds $\psi_c(x) \geq \psi_c(\bar{x})$ for all x in a certain neighborhood of $\bar{x}$. Consequently, for every $w \in \mathbb{R}^n$ and $h \in \mathcal{C}(\bar{x})$, it follows

$$0 \leq \liminf_{t\downarrow 0,\, \tilde{w}\to w} \frac{\psi_c(\bar{x} + th + \frac{1}{2}t^2\tilde{w}) - \psi_c(\bar{x}) - t(\psi_c)_-^{\downarrow}(\bar{x};h)}{\frac{1}{2}t^2} = (\psi_c)_-^{\downarrow\downarrow}(\bar{x};h,w).$$

By taking the infimum of the latter inequality over all $w \in \mathbb{R}^n$, we obtain

$$\inf_w \, (\psi_c)_-^{\downarrow\downarrow}(\bar{x};h,w) \geq 0 \tag{A.2.3}$$

for all $h \in \mathcal{C}(\bar{x})$. Since Robinson's constraint qualification is satisfied at $\bar{x}$ and the term $\phi^{\downarrow}(F(\bar{x}); DF(\bar{x})h)$ is finite for all $h \in \mathcal{C}(\bar{x})$ we can apply the chain rule in Lemma 5.2.5 to compute the second order directional epiderivative $(\psi_c)_-^{\downarrow\downarrow}(\bar{x};h,w)$; it holds:

$$(\psi_c)_-^{\downarrow\downarrow}(\bar{x};h,w) = \nabla f(\bar{x})^\top w + h^\top \nabla^2 f(\bar{x})h + \phi_-^{\downarrow\downarrow}(F(\bar{x}); DF(\bar{x})h, DF(\bar{x})w + D^2F(\bar{x})[h,h]).$$

Thus, the infimum expression on the left side of the second order condition (A.2.3) is equivalent to the following problem

$$\inf_{w,t} \; \nabla f(\bar{x})^\top w + h^\top \nabla^2 f(\bar{x})h + t \quad \text{s.t.} \quad \phi_-^{\downarrow\downarrow}(F(\bar{x}); DF(\bar{x})h, DF(\bar{x})w + \bar{w}) \leq t,$$

where $\bar{w} := D^2F(\bar{x})[h,h]$. Furthermore, setting $\xi_{\phi,h}(\cdot) := \phi_-^{\downarrow\downarrow}(F(\bar{x}); DF(\bar{x})h, \cdot)$, we obtain

$$\inf_{w,t} \; \nabla f(\bar{x})^\top w + h^\top \nabla^2 f(\bar{x})h + t \quad \text{s.t.} \quad (DF(\bar{x})w + \bar{w}, t) \in \text{epi } \xi_{\phi,h}.$$

In order to dualize the latter problem, we need to replace the possibly nonconvex epigraph of the lower second order directional epiderivative by an appropriately chosen convex expression. Of course, if we minimize over a slightly smaller set, the optimal value of the resulting problem will also be bounded below by zero. Hence, let us consider an arbitrary convex function $\zeta(\cdot) \geq \xi_{\phi,h}(\cdot) = \phi_{-}^{\downarrow\downarrow}(F(\bar{x}), DF(\bar{x})h, \cdot)$. Then, it follows epi $\zeta \subset$ epi $\xi_{\phi,h}$ and, due to (A.2.2), we obtain

$$\Xi := \mathrm{cl}\{\mathrm{epi}\ \zeta + \mathrm{epi}\ \phi^{\downarrow}(F(\bar{x}); \cdot)\} \subset \mathrm{epi}\ \xi_{\phi,h}.$$

Moreover, since Ξ is the closure of the sum of two convex sets, this set is also closed and convex. Thus, we can apply the Fenchel-Rockafellar duality framework to the convex problem

$$\text{(A.2.4)} \qquad \inf_{w,t}\ \nabla f(\bar{x})^\top w + h^\top \nabla^2 f(\bar{x}) h + t + \iota_\Xi(DF(\bar{x})w + \bar{w}, t).$$

Specifically, defining $\Pi(y,\gamma) := \iota_\Xi(y + \bar{w}, \gamma)$ and $\varrho(y,\gamma) := \nabla f(\bar{x})^\top y + h^\top \nabla^2 f(\bar{x}) h + \gamma$, the dual problem of (A.2.4) is formally given by

$$\max_{v,\tau}\ -\varrho^*(DF(\bar{x})^\top v, \tau) - \Pi^*(-v, -\tau),$$

(see, e.g., [11, Chapter 15 and Definition 15.19]). The convex conjugates ϱ^* and Π^* can be computed as follows

- $\begin{aligned}\varrho^*(DF(\bar{x})^\top v, \tau) &= -h^\top \nabla^2 f(\bar{x}) h + \sup_{y,\gamma}\ \langle y, DF(\bar{x})^\top v - \nabla f(\bar{x})\rangle + \gamma(\tau - 1) \\ &= -h^\top \nabla^2 f(\bar{x}) h + \iota_{\{y\colon \nabla f(\bar{x}) + DF(\bar{x})^\top y = 0\} \times \{1\}}(-v, \tau),\end{aligned}$
- $\Pi^*(-v, -\tau) = \sup_{y,\gamma}\ -\langle y, v\rangle - \gamma\tau - \iota_\Xi(y + \bar{w}, \gamma) = \langle \bar{w}, v\rangle + \sigma_\Xi(-v, -\tau).$

Thus, using $\lambda \equiv -v$, the dual problem can be rewritten as the following constrained program

$$\max_{\lambda}\ h^\top \nabla^2 f(\bar{x}) h + \langle \lambda, D^2 F(\bar{x})[h,h]\rangle - \sigma_\Xi(\lambda, -1) \quad \text{s.t.} \quad \nabla f(\bar{x}) + DF(\bar{x})^\top \lambda = 0.$$

Next, let us fix an arbitrary element $(\bar{y}, \bar{\gamma}) \in$ epi ζ. Then, by Lemmas 2.1.8 and 2.5.3, it follows

$$\mathrm{dom}\ \sigma_\Xi \subset \mathrm{dom}\ \sigma_{(\bar{y},\bar{\gamma}) + \mathrm{epi}\ \phi^{\downarrow}(F(\bar{x});\cdot)} \subset N_{\mathrm{epi}\ \phi}(F(\bar{x}), \phi(F(\bar{x}))).$$

Moreover, due to Lemma 2.5.14, we have

$$\partial\phi(F(\bar{x})) \times \{-1\} = N_{\mathrm{epi}\ \phi}(F(\bar{x}), \phi(F(\bar{x}))) \cap \mathbb{R}^m \times \{-1\},$$

and, consequently, it holds $\sigma_\Xi(\lambda, -1) = +\infty$ if $\lambda \notin \partial\phi(F(\bar{x}))$. Hence, the dual problem takes the following final form

$$\text{(A.2.5)} \qquad \max_{\lambda \in \mathcal{M}(\bar{x})}\ h^\top \nabla^2 f(\bar{x}) h + \langle \lambda, D^2 F(\bar{x})[h,h]\rangle - \sigma_\Xi(\lambda, -1).$$

In addition, utilizing [11, Theorem 15.23 and Proposition 15.24], there is no duality gap between the primal problem (A.2.4) and the dual problem (A.2.5) when the following regularity condition is satisfied:

$$0 \in \text{int}\left\{\begin{pmatrix} DF(\bar{x})\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - \text{dom }\Pi\right\} = \text{int}\left\{\begin{pmatrix} \bar{w} + DF(\bar{x})\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - \Xi\right\}. \tag{A.2.6}$$

In section 5.1, we worked with an equivalent representation of Robinson's constraint qualification, see page 92 or [27, Proposition 2.97, Corollary 2.98] for details. Using this alternative formulation, Robinson's constraint qualification implies

$$\begin{pmatrix} \mathbb{R}^m \\ \mathbb{R} \end{pmatrix} = \begin{pmatrix} DF(\bar{x})\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - \text{epi } \phi^{\downarrow}(F(\bar{x}), \cdot) \subset \begin{pmatrix} \bar{y} \\ \bar{\gamma} \end{pmatrix} + \begin{pmatrix} DF(\bar{x})\mathbb{R}^n \\ \mathbb{R} \end{pmatrix} - \Xi.$$

Clearly, this easily establishes (A.2.6) and shows that problem (A.2.4) and (A.2.5) coincide and have the same optimal value, which is bounded below by zero. Now, due to epi $\zeta \subset \Xi$ and Lemma 2.1.8, it immediately follows

$$\sigma_{\Xi}(\lambda, -1) \geq \sigma_{\text{epi } \zeta}(\lambda, -1) = \sup_{(y,\gamma)\in\text{epi } \zeta} \langle y, \lambda\rangle - \gamma = \zeta^*(\lambda).$$

and, by (A.2.5) and (A.2.3), we have

$$\max_{\lambda\in\mathcal{M}(\bar{x})} \; h^\top \nabla^2 f(\bar{x})h + \langle \lambda, D^2F(\bar{x})[h,h]\rangle - \zeta^*(\lambda) \;\geq\; \inf_w \; (\psi_c)_-^{\downarrow\downarrow}(\bar{x}; h, w) \;\geq\; 0,$$

for all $h \in \mathcal{C}(\bar{x})$. This concludes the proof of the first part of Theorem 5.2.8.

Proof of part (ii). Let us suppose that the second order growth condition does not hold at the stationary point $\bar{x}$. Then, there exist sequences $(x^k)_k \subset F^{-1}(\text{dom } \varphi)$ and $(\alpha_k)_k \subset \mathbb{R}$ such that

$$x^k \to \bar{x}, \quad \alpha_k \downarrow 0, \quad k \to \infty$$

and

$$\psi_c(x^k) \leq \psi_c(\bar{x}) + \alpha_k \|x^k - \bar{x}\|^2. \tag{A.2.7}$$

Let us define $t_k := \|x^k - \bar{x}\|$ and $h^k := (x^k - \bar{x})/t_k$. Then, by passing to a subsequence if necessary, we can assume that $(h^k)_k$ converges to some $h \in \mathbb{R}^n$ with $\|h\| = 1$. This readily establishes

$$(\psi_c)_-^{\downarrow}(\bar{x}; h) = \liminf_{k\to\infty} \frac{\psi_c(\bar{x} + t_k h^k) - \psi_c(\bar{x})}{t_k} \leq \liminf_{k\to\infty} \; \alpha_k \|x^k - \bar{x}\| = 0.$$

Thus, since $\bar{x}$ satisfies the first order necessary conditions, it follows $h \in \mathcal{C}(\bar{x}) \setminus \{0\}$. Furthermore, it can be easily seen, that the vector x^k can be rewritten in the following form

$$x^k = \bar{x} + t_k h + \frac{1}{2}t_k^2[2t_k^{-1}(h^k - h)].$$

Setting $w^k := 2t_k^{-1}(h^k - h)$, we have $t_k w^k \to 0$ and using inequality (A.2.7), it holds

$$\psi_c(\bar{x} + t_k h + \frac{1}{2}t_k^2 w^k) \leq \psi_c(\bar{x}) + t_k(\psi_c)_-^{\downarrow}(\bar{x}; h) + \frac{1}{2}t_k^2 \cdot 2\alpha_k.$$

Hence, since $h \in \mathcal{C}(\bar{x})$ and Robinson's constraint qualification is satisfied at $\bar{x}$, the outer second order regularity of ϕ and Lemma 5.2.7 imply that there exist sequences $(\tilde{w}^k)_k$ and $(\tilde{\alpha}_k)_k$ such that $\tilde{w}^k - w^k \to 0$, $\tilde{\alpha}_k - \alpha_k \to 0$, and

$$2\tilde{\alpha}_k \geq (\psi_c)^{\downarrow\downarrow}_{-}(\bar{x}; h; \tilde{w}^k).$$

Moreover, the second order sufficient condition (5.2.4) implies that there exist $\varepsilon > 0$ and $\bar{\lambda} \in \mathcal{M}(\bar{x})$ such that

$$h^\top \nabla^2 f(\bar{x}) h + \langle \bar{\lambda}, D^2F(\bar{x})[h,h] \rangle - \xi^*_{\phi,h}(\bar{\lambda}) \geq \varepsilon. \tag{A.2.8}$$

Now, applying the chain rule for lower second order directional epiderivatives, $\bar{\lambda} \in \mathcal{M}(\bar{x})$, Fenchel's inequality, and (A.2.8), we obtain

$$\begin{aligned} 0 \geq h^\top \nabla^2 f(\bar{x}) h + \langle \bar{\lambda}, D^2F(\bar{x})[h,h] \rangle - \langle \bar{\lambda}, DF(\bar{x})\tilde{w}^k + D^2F(\bar{x})[h,h] \rangle \\ + \phi^{\downarrow\downarrow}_{-}(F(\bar{x}); DF(\bar{x})h, DF(\bar{x})\tilde{w}^k + D^2F(\bar{x})[h,h]) - 2\tilde{\alpha}_k \\ \geq h^\top \nabla^2 f(\bar{x}) h + \langle \bar{\lambda}, D^2F(\bar{x})[h,h] \rangle - \xi^*_{\phi,h}(\bar{\lambda}) - 2\tilde{\alpha}_k \geq \varepsilon - 2\tilde{\alpha}_k. \end{aligned}$$

Since $(\alpha_k)_k$ converges to zero, taking the limit $k \to \infty$ yields the desired contradiction. This finishes the proof of the second order sufficient conditions (5.2.4).

Finally, let us suppose that the function $\xi_{\phi,h}(\cdot) = \phi^{\downarrow\downarrow}_{-}(F(\bar{x}); DF(\bar{x})h, \cdot)$ is convex and that the quadratic growth condition (5.2.5) is fulfilled in a certain neighborhood of $\bar{x}$. Let $\tilde{w} \in \mathbb{R}^n$ and $h \in \mathcal{C}(\bar{x}) \setminus \{0\}$ be arbitrary. Then, for all $t > 0$ sufficiently small it follows

$$\frac{\psi_c(\bar{x} + th + \frac{1}{2}t^2\tilde{w}) - \psi_c(\bar{x}) - t(\psi_c)^{\downarrow}_{-}(\bar{x}; h)}{\frac{1}{2}t^2} \geq 2\alpha \|h + \tfrac{1}{2}t\tilde{w}\|^2$$

Thus, we have

$$(\psi_c)^{\downarrow\downarrow}_{-}(\bar{x}; h, w) \geq 2\alpha\|h\|^2 > 0$$

for all $w \in \mathbb{R}^n$. Since $\xi_{\phi,h}$ is assumed to be convex, we can apply the duality arguments of the proof of part (i) and set $\zeta \equiv \xi_{\phi,h}$. This shows

$$\max_{\lambda \in \mathcal{M}(\bar{x})} \; h^\top \nabla^2 f(\bar{x}) h + \langle \lambda, D^2F(\bar{x})[h,h] \rangle - \xi^*_{\phi,h}(\lambda) \;\geq\; \inf_w \; (\psi_c)^{\downarrow\downarrow}_{-}(\bar{x}; h, w) \;>\; 0$$

and concludes the proof of Theorem 5.2.8.

A.3. Variational inequalities: Existence of solutions

In this paragraph, we verify existence of solutions of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$ under a coercivity and monotonicity assumption. The proof of this result follows the guidelines in [75, Exercise 2.9.11] and is presented for the sake of completeness.

Proof A.3.1 (Proof of Lemma 6.1.4). For $k \in \mathbb{N}$, we consider the following multifunction

$$\Phi_k : \bar{B}_k(0) \rightrightarrows \bar{B}_k(0), \quad \Phi_k(x) := \operatorname*{arg\,min}_{y \in \bar{B}_k(0)} \langle F(x), y \rangle + \varphi(y) =: \theta(y).$$

Since the function $\theta : \mathbb{R}^n \to (-\infty, +\infty]$ is convex, proper, and lower semicontinuous, Weierstrass' Theorem implies that the set $\Phi_k(x)$ is nonempty for all $x \in \bar{B}_k(0)$ and all $k \in \mathbb{N}$ sufficiently large. Moreover, by using the convexity of φ, we immediately see that $\Phi_k(x)$ is also a convex set for all $x \in \bar{B}_k(0)$ and all $k \in \mathbb{N}$. Now, let us consider arbitrary points $\bar{x}, \bar{y} \in \bar{B}_k(0)$ and sequences $(x^n)_n$, $(y^n)_n$ such that

$$x^n \to \bar{x}, \quad y^n \to \bar{y}, \quad y^n \in \Phi_k(x^n), \quad \forall\, n \in \mathbb{N}.$$

Then, from the definition of Φ_k, it follows

$$\langle F(x^n), y^n \rangle + \varphi(y^n) \le \langle F(x^n), y \rangle + \varphi(y), \quad \forall\, y \in \bar{B}_k(0).$$

Hence, taking the limes inferior $n \to \infty$, we readily obtain $\bar{y} \in \Phi_k(\bar{x})$. This shows that the mapping Φ_k is *closed* and thus, by *Kakutani's fixed-point theorem* [75, Theorem 2.1.19], the multifunction Φ_k possesses a fixed point $x^k \in \bar{B}_k(0)$,

$$x^k \in \Phi_k(x^k),$$

for all $k \ge K$ and some $K \in \mathbb{N}$; we refer to [75, Definition 2.1.16 and Theorem 2.1.19] for more details. Since every accumulation point of the sequence $(x^k)_{k \ge K}$ obviously corresponds to a solution of the generalized variational inequality $(\mathcal{P}_{\text{vip}})$, we will now show that the additional conditions in part (ii) and (iii) of Lemma 6.1.4 guarantee boundedness of the sequence $(x^k)_k$. (Clearly, if the sets Ω or $\operatorname{dom} \varphi$ are bounded, then we can conclude the proof here).

Thus, let us suppose that the assumptions in Lemma 6.1.4 (ii) are satisfied and that the sequence $(x^k)_{k \ge K}$ is not bounded. Then, there exist $\vartheta > 0$ and $\xi \ge 0$ such that

$$\liminf_{k \to \infty} \frac{\langle F(x^k), x^k - x^* \rangle}{\|x^k\|^\xi} \ge \vartheta.$$

Consequently, the coercivity of φ implies

$$\liminf_{k \to \infty} \left\{ \frac{\langle F(x^k), x^k - x^* \rangle}{\|x^k\|^\xi} + \frac{\varphi(x^k) - \varphi(x^*)}{\|x^k\|^\xi} \right\} \ge \vartheta.$$

and hence, for all $k \ge K$ sufficiently large, we obtain

$$\langle F(x^k), x^k - x^* \rangle + \varphi(x^k) - \varphi(x^*) \ge \vartheta \cdot \max\{1, \|x^k\|^\xi\} \ge \vartheta.$$

However, this clearly is a contradiction to $x^k \in \Phi_k(x^k)$ and shows that the sequence $(x^k)_{k \ge K}$ has to be bounded.

Next, to prove the third part of Lemma 6.1.4, let us suppose that F is ξ-monotone for some $\xi > 1$ and that φ is subdifferentiable at x^* with $\lambda^* \in \partial\varphi(x^*)$. Then, for all $k \ge K$ with

$x^* \in \bar{B}_k(0)$, we obtain

$$\langle F(x^*), x^* - x^k \rangle \geq \langle F(x^k) - F(x^*), x^k - x^* \rangle + \varphi(x^k) - \varphi(x^*) \geq \|x^k - x^*\|^\xi + \langle \lambda^*, x^k - x^* \rangle.$$

This immediately implies

$$\|x^k - x^*\|^{\xi-1} \leq \|F(x^*) + \lambda^*\|$$

and again establishes the desired boundedness of the sequence $(x^k)_{k \geq K}$. The asserted uniqueness of the solution follows from Lemma 6.1.4 (i).

A.4. Numerical results: Further figures and tables

On the following pages, we provide additional numerical results for the ℓ_1-regularized and group sparse test problems. The tables report the averaged, total CPU times and variants of the Figures 7.1 and 7.10 for different dynamic ranges are shown.

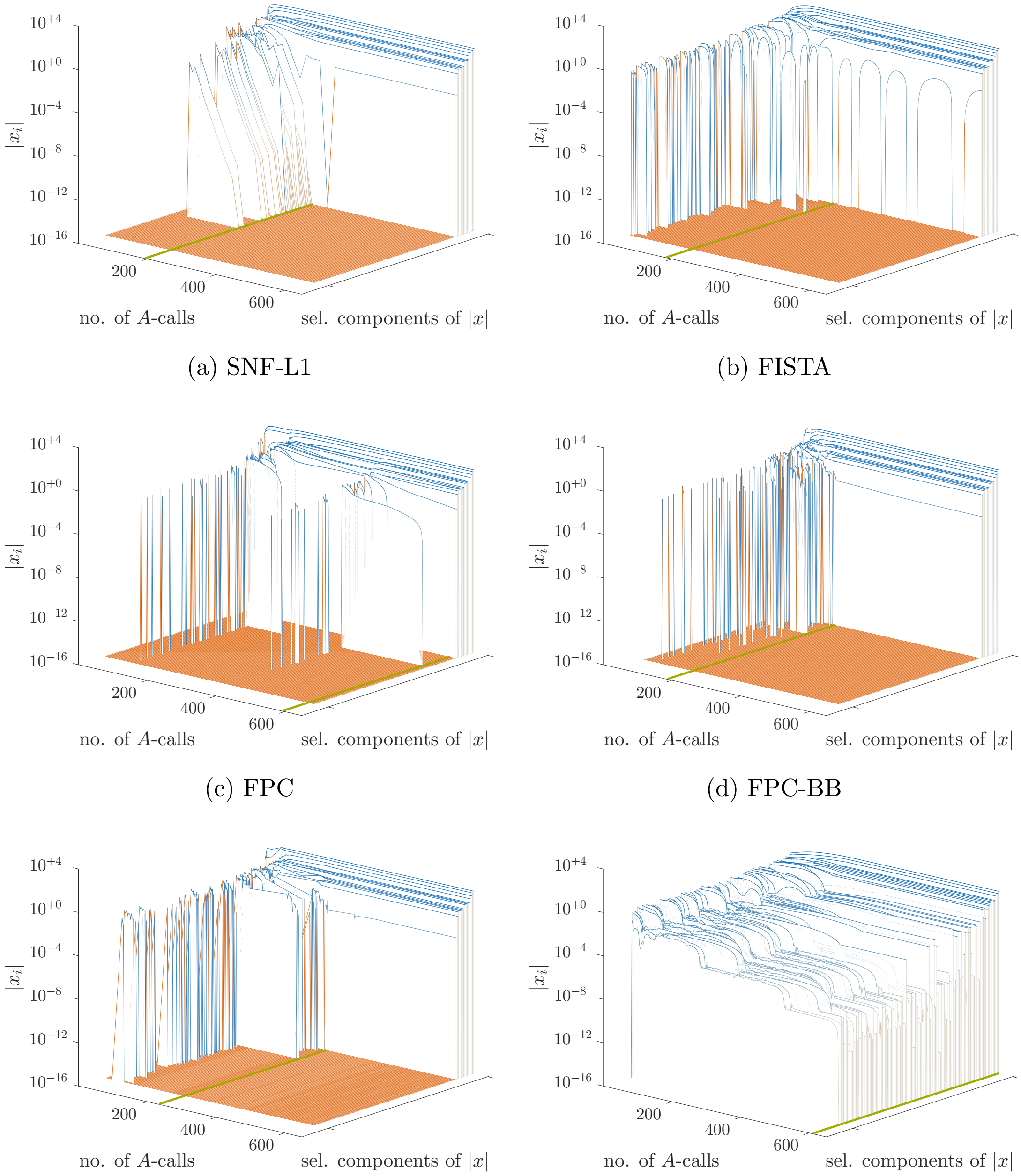
A.4.1. Convex ℓ_1-regularized least squares

Table A.1.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 20 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$) (best NESTA configuration was used: $\mu = 10^{-8}$, $T = 4$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}
SNF-L1	2.5	2.5	5.8	5.8	6.5	6.6	7.2	7.3	8.4	8.4
FISTA	2.2	2.3	6.2	6.0	14.6	14.0	57.8	59.4	133.2	143.2
FPC	12.0	14.7	16.6	17.3	25.2	25.5	48.4	52.8	76.9	84.1
FPC-BB	4.8	5.8	5.8	6.9	8.9	8.3	32.5	34.2	DNC	68.3
FPC-AS	2.5	2.7	5.8	6.6	8.1	9.2	9.8	11.6	11.8	12.7
NESTA	13.5	13.5	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
GPSR	16.2	19.0	20.0	20.9	27.2	28.3	45.7	49.1	DNC	72.1
GPSR-BB	12.7	17.1	14.5	17.7	17.3	20.3	25.6	30.4	36.9	42.1
PD	2.6	4.2	12.7	12.6	33.9	33.4	DNC	312.4	DNC	DNC
SpaRSA	13.5	17.6	14.7	18.6	15.7	19.5	17.8	22.5	20.3	25.2
SPGL1	2.4	2.4	5.4	5.2	7.0	7.6	DNC	DNC	DNC	DNC
YALL1	1.6	1.6	3.8	3.7	6.0	5.9	10.7	11.0	16.8	16.4

Table A.2.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 40 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$) (best NESTA configuration was used: $\mu = 10^{-1}$, $T = 5$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}
SNF-L1	5.1	5.0	11.9	11.9	12.7	12.7	13.8	13.8	15.4	15.3
FISTA	6.1	9.5	17.1	17.0	36.8	32.9	132.4	115.3	231.8	255.0
FPC	11.9	14.9	32.6	26.9	49.7	47.2	92.8	92.7	127.7	139.1
FPC-BB	5.1	6.6	14.9	9.1	33.2	27.4	78.9	77.8	DNC	145.8
FPC-AS	7.0	5.5	11.8	10.4	13.9	15.1	15.7	17.2	18.7	20.2
NESTA	9.1	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
GPSR	13.6	15.9	25.4	24.1	33.3	33.0	52.5	52.1	DNC	DNC
GPSR-BB	14.0	19.1	19.4	20.3	26.8	28.4	45.1	48.6	60.1	69.8
PD	5.0	9.7	21.6	23.3	65.6	69.6	DNC	DNC	DNC	DNC
SpaRSA	13.3	18.3	16.0	18.9	17.7	21.4	21.3	25.5	23.9	29.5
SPGL1	3.9	5.0	10.6	9.8	12.8	13.8	DNC	DNC	DNC	DNC
YALL1	3.5	5.9	12.1	11.4	21.7	19.6	50.9	43.9	75.3	72.7



(a) SNF-L1

(b) FISTA

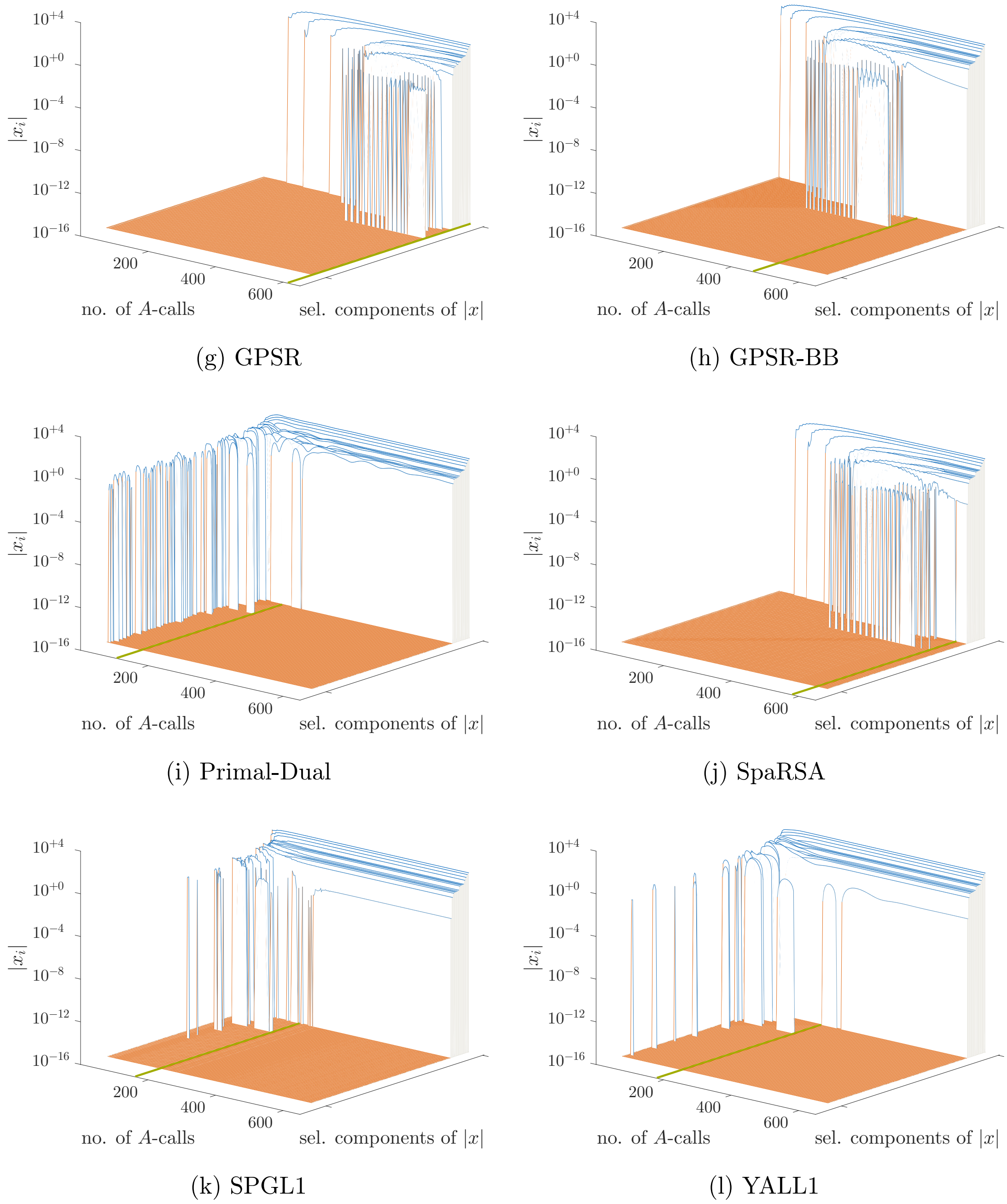
(c) FPC

(d) FPC-BB

(e) FPC-AS

(f) NESTA

Figure A.1.: Change of the absolute value of 128 randomly chosen components of the iterate of a single run with dynamic range 20 dB with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero components of the high precision solution x^* within the sample (the following NESTA configuration was used: $\mu = 10^{-8}$, $T = 4$).



(g) GPSR

(h) GPSR-BB

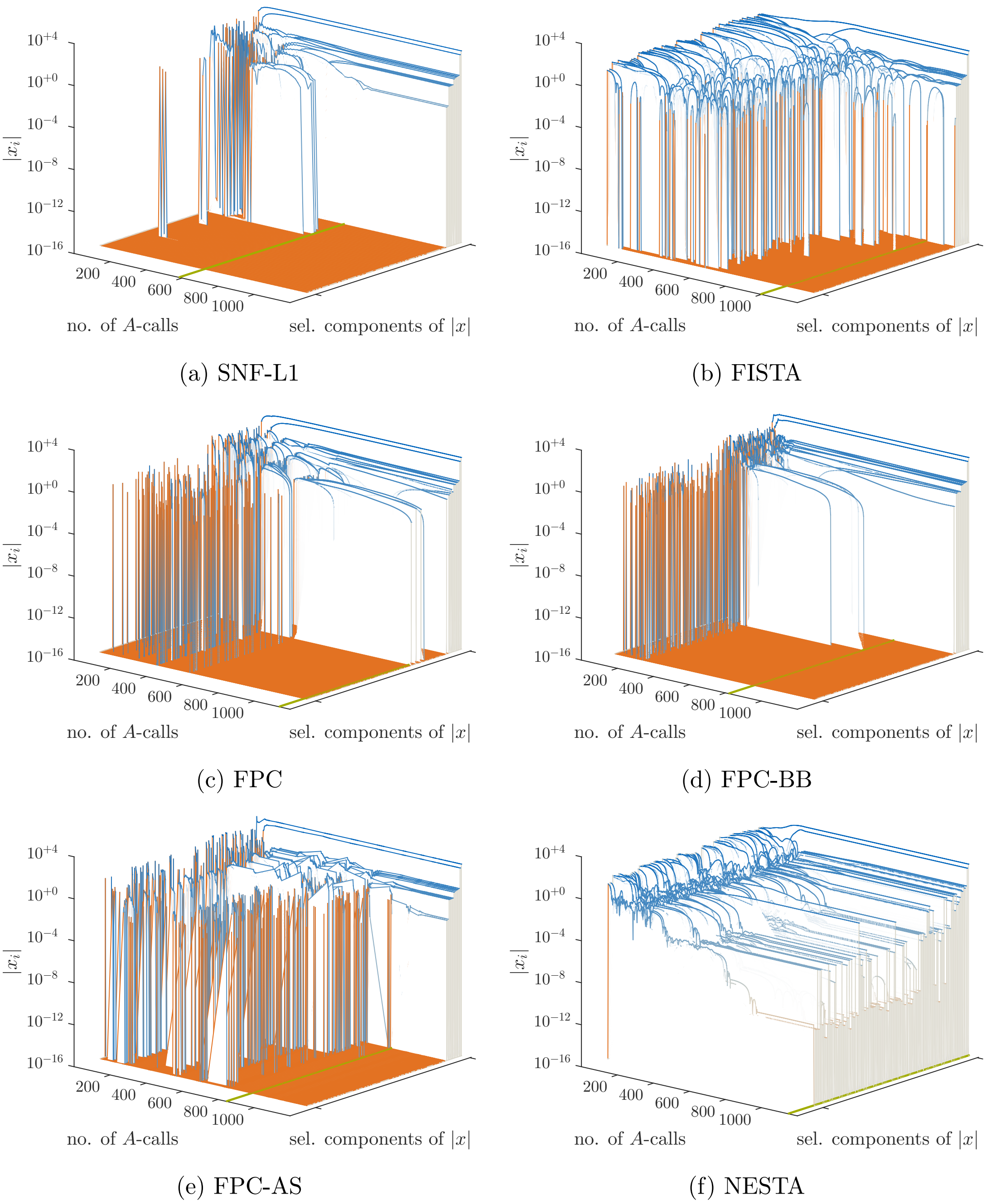
(i) Primal-Dual

(j) SpaRSA

(k) SPGL1

(l) YALL1

Figure A.1.: Change of the absolute value of 128 randomly chosen components of the iterate of a single run with dynamic range 20 dB with respect to the number of A- and $A^{\top}$-calls. Green line: maximum number of A-calls that is needed to detect all zero components of the high precision solution x^{*} within the sample.



(a) SNF-L1

(b) FISTA

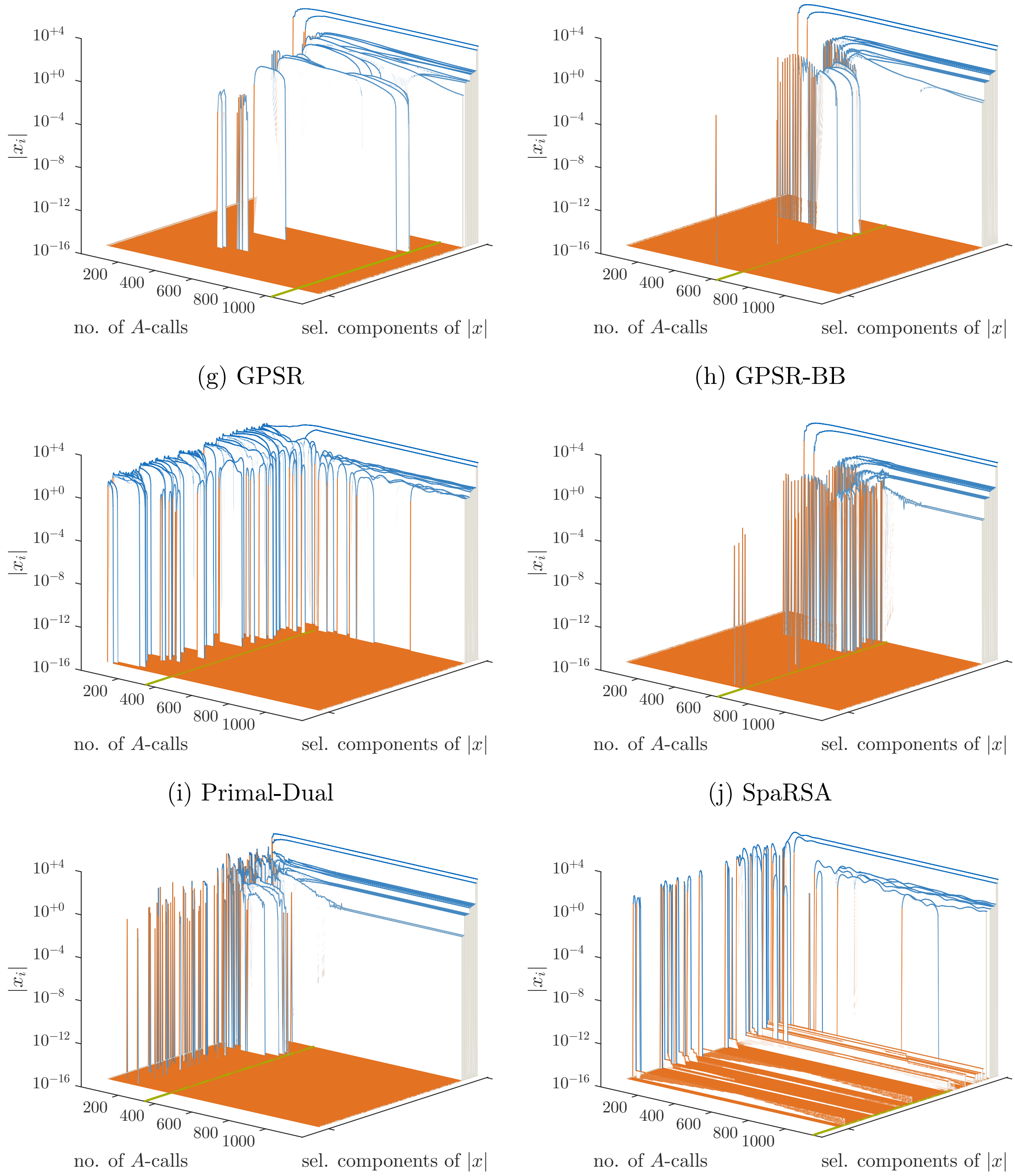
(c) FPC

(d) FPC-BB

(e) FPC-AS

(f) NESTA

Figure A.2.: Change of the absolute value of 128 randomly chosen components of the iterate of a single run with dynamic range 60 dB with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero components of the high precision solution x^* within the sample (the following NESTA configuration was used: $\mu = 10^{-8}$, $T = 4$).



(g) GPSR

(h) GPSR-BB

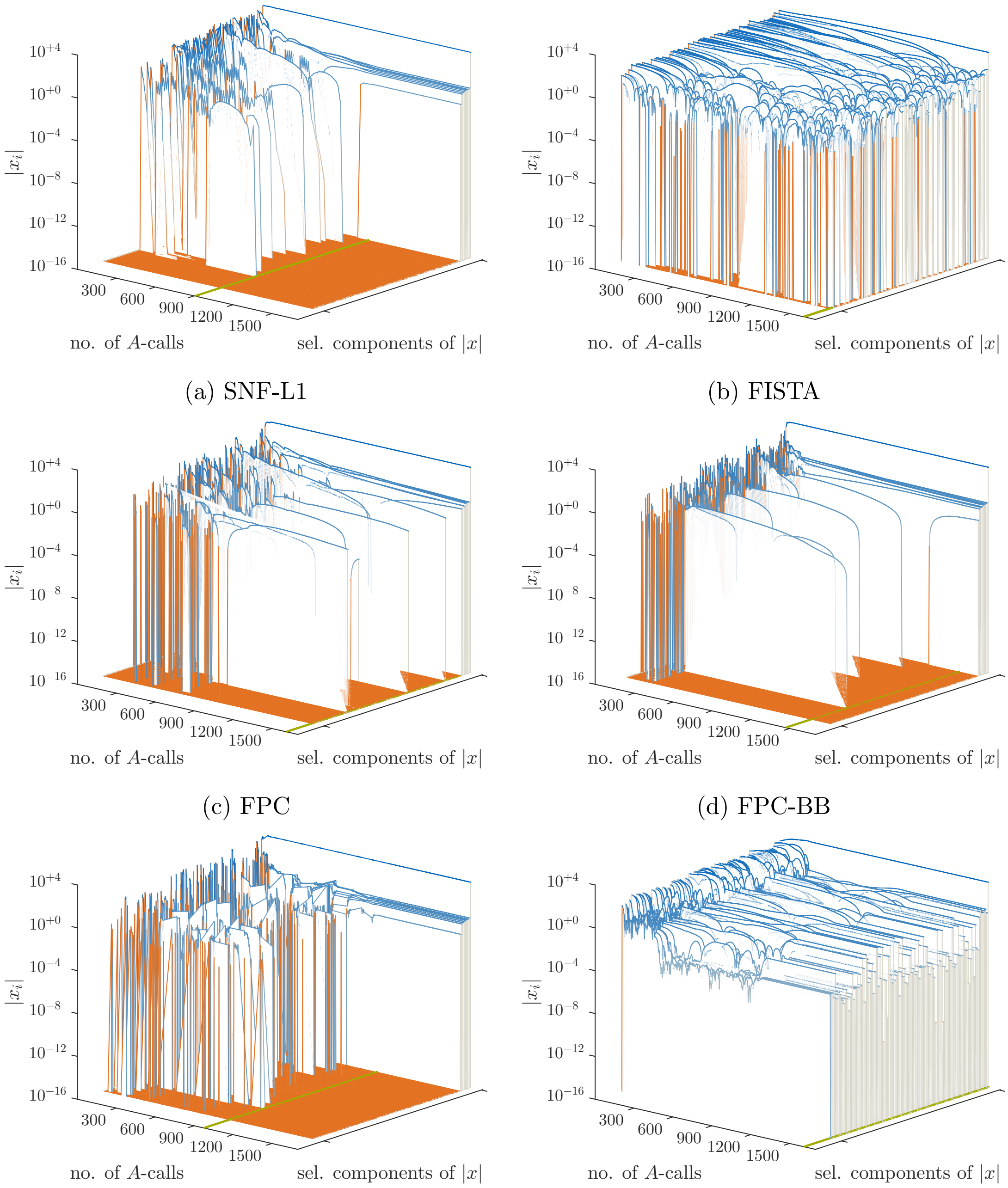
(i) Primal-Dual

(j) SpaRSA

(k) SPGL1

(l) YALL1

Figure A.2.: Change of the absolute value of 128 randomly chosen components of the iterate of a single run with dynamic range 60 dB with respect to the number of A- and $A^{\top}$-calls. Green line: maximum number of A-calls that is needed to detect all zero components of the high precision solution x^* within the sample.



(a) SNF-L1

(b) FISTA

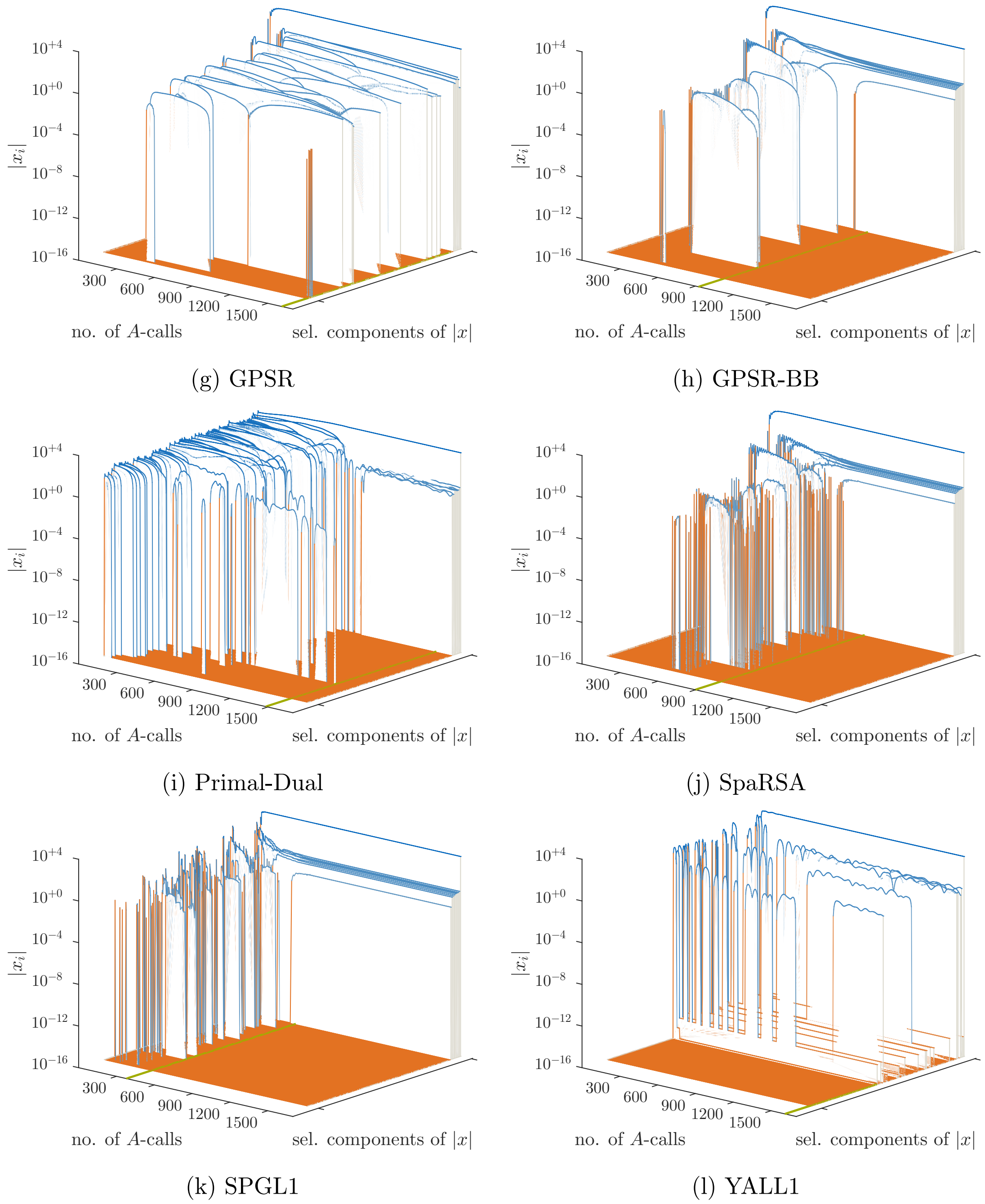
(c) FPC

(d) FPC-BB

(e) FPC-AS

(f) NESTA

Figure A.3.: Change of the absolute value of 128 randomly chosen components of the iterate of a single run with dynamic range 80 dB with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero components of the high precision solution x^* within the sample (the following NESTA configuration was used: $\mu = 10^{-8}$, $T = 4$).



(g) GPSR

(h) GPSR-BB

(i) Primal-Dual

(j) SpaRSA

(k) SPGL1

(l) YALL1

Figure A.3.: Change of the absolute value of 128 randomly chosen components of the iterate of a single run with dynamic range 80 dB with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero components of the high precision solution x^* within the sample.

Table A.3.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 60 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$) (here, NESTA did not converge).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}
SNF-L1	10.6	10.6	17.7	17.6	18.5	18.4	19.9	19.9	22.1	22.1
FISTA	20.1	25.3	35.6	37.9	58.6	61.8	151.8	167.6	289.4	338.1
FPC	20.3	20.6	55.2	46.3	75.4	76.1	124.0	137.3	175.8	200.3
FPC-BB	7.9	9.0	41.3	29.1	62.6	61.8	113.6	128.9	DNC	DNC
FPC-AS	7.0	7.5	17.1	18.3	18.7	20.5	21.2	23.5	24.9	28.3
NESTA	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
GPSR	22.1	15.4	38.3	35.2	47.3	47.8	68.8	73.4	DNC	DNC
GPSR-BB	14.7	18.7	27.5	25.7	36.7	39.5	58.0	68.0	80.2	97.1
PD	12.5	22.3	48.6	56.4	137.3	189.7	DNC	DNC	DNC	DNC
SpaRSA	15.4	20.0	18.6	21.5	20.6	24.7	24.5	30.3	27.1	34.6
SPGL1	6.7	7.9	14.6	14.0	17.4	19.0	DNC	DNC	DNC	DNC
YALL1	16.3	30.7	66.8	68.1	106.6	116.1	204.2	222.6	300.4	323.2

Table A.4.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 80 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$) (here, NESTA did not converge).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}
SNF-L1	18.1	18.4	25.3	25.5	26.2	26.4	28.0	28.4	30.7	31.0
FISTA	71.2	92.7	94.2	110.9	125.1	146.5	244.6	288.3	390.1	508.4
FPC	40.4	33.1	82.0	71.7	106.5	110.5	164.8	186.4	215.4	263.5
FPC-BB	17.0	10.5	60.5	45.6	86.2	88.3	147.3	171.9	DNC	DNC
FPC-AS	10.8	12.7	20.9	22.9	22.4	25.5	26.3	30.2	30.8	37.0
NESTA	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC
GPSR	65.0	34.6	83.6	84.9	94.5	101.3	120.5	133.3	DNC	DNC
GPSR-BB	20.1	17.4	34.8	34.0	43.7	48.8	65.3	78.5	85.0	108.8
PD	35.0	60.2	123.9	148.4	332.0	490.5	DNC	DNC	DNC	DNC
SpaRSA	16.5	18.8	21.5	24.4	23.6	29.0	27.5	35.3	31.6	40.3
SPGL1	10.3	12.0	18.1	17.5	21.7	23.9	DNC	DNC	DNC	DNC
YALL1	146.4	251.6	DNC	DNC	DNC	DNC	DNC	DNC	DNC	DNC

A.4.2. Nonconvex ℓ_1-optimization problems

Table A.5.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 20 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-T	9.2	9.0	21.7	21.5	30.0	29.9	43.5	43.3	53.5	52.3
FPD	23.0	22.9	103.6	74.6	172.8	166.0	DNC	728.1	DNC	DNC
FPD-BB$_\mu$	18.8	86.2	80.5	90.3	146.8	188.5	DNC	DNC	DNC	DNC
FPD-BB$_v$	26.2	27.7	116.8	71.7	216.1	265.5	DNC	DNC	DNC	DNC

Table A.6.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 40 dB using the stopping criteria ($\mathcal{C}_{\text{rel}}$) and ($\mathcal{C}_{\text{nat}}$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$	$\#A_{\text{rel}}$	$\#A_{\text{nat}}$
SNF-T	34.3	33.8	65.8	65.3	76.4	76.0	94.7	93.9	107.3	106.7
FPD	140.0	191.3	309.8	311.8	454.3	461.2	DNC	DNC	DNC	DNC
FPD-BB$_\mu$	77.4	147.9	149.6	157.8	310.8	338.9	DNC	DNC	DNC	DNC
FPD-BB$_v$	15.2	30.2	58.0	36.9	223.5	228.0	DNC	DNC	DNC	DNC

A.4.3. Group sparse problems

Table A.7.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 20 dB and $\bar{\sigma} = 0.1$ using the stopping criteria $(\mathcal{C}_{\mathrm{rel}})$ and $(\mathcal{C}_{\mathrm{nat}})$.

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}
SNF-GS	4.3	4.3	5.9	5.8	8.6	8.6	14.8	14.9	14.8	14.8
ADMM	7.4	2.9	13.4	10.0	26.5	19.0	66.0	37.3	66.1	55.2
FISTA	4.2	2.9	9.5	8.6	22.6	24.2	107.7	91.9	107.9	200.0
FPD-BB	4.9	3.2	7.6	7.7	12.1	13.5	23.6	25.0	23.5	34.8
PD	2.8	3.9	3.8	6.3	6.0	11.1	52.9	53.4	53.4	283.1
SpaRSA	11.0	14.1	11.9	15.0	14.3	18.2	22.5	25.6	22.5	33.8
SPG-GS	3.4	3.4	5.2	5.7	8.7	9.8	DNC	DNC	DNC	DNC

Table A.8.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 20 dB and $\bar{\sigma} = 0.01$ using the stopping criteria $(\mathcal{C}_{\mathrm{rel}})$ and $(\mathcal{C}_{\mathrm{nat}})$.

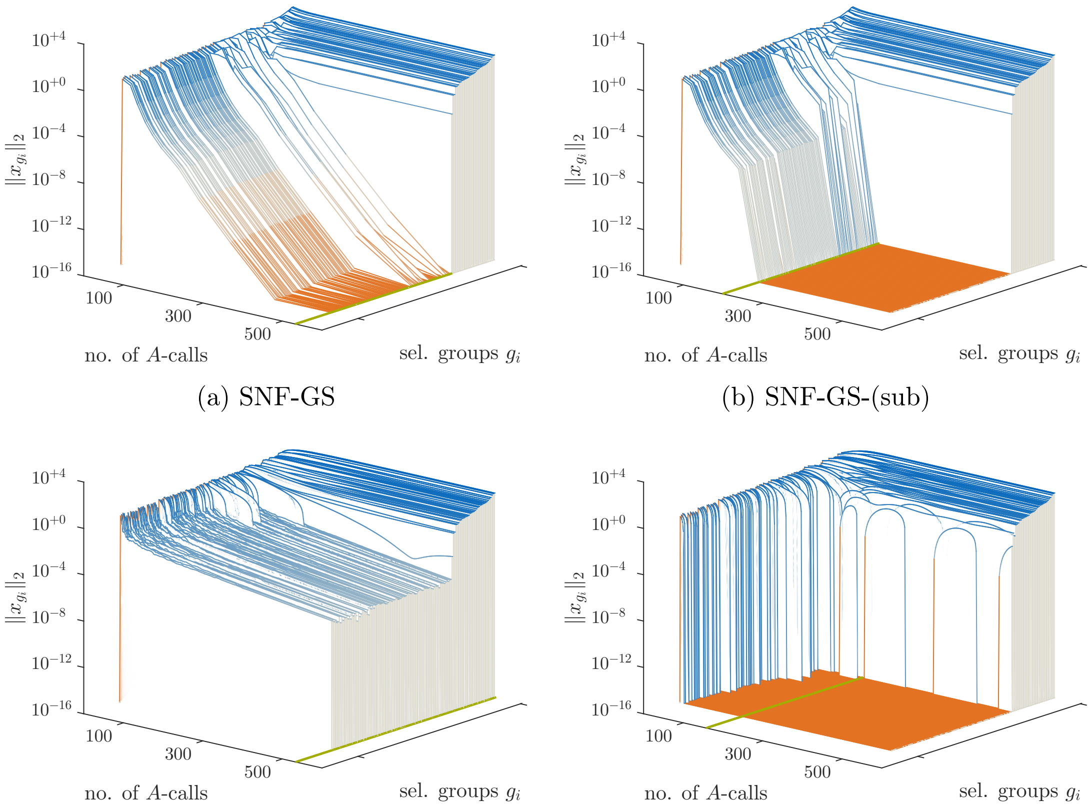
Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}
SNF-GS	8.4	8.4	11.4	11.4	17.9	17.9	28.2	28.2	38.7	38.4
ADMM	4.6	1.2	12.7	6.2	28.1	14.6	52.0	33.7	62.3	53.0
FISTA	9.7	0.7	32.1	11.6	119.9	35.0	DNC	269.9	DNC	DNC
FPD-BB	19.5	4.5	52.8	9.6	92.9	52.1	148.9	148.1	175.9	245.5
PD	5.4	1.1	8.5	11.2	11.5	15.5	14.2	20.4	15.3	24.2
SpaRSA	12.7	16.8	18.4	17.8	39.1	23.1	71.2	69.6	88.5	129.9
SPG-GS	5.5	6.5	16.8	8.6	52.0	16.2	DNC	DNC	DNC	DNC

Table A.9.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 40 dB and $\bar{\sigma} = 0.1$ using the stopping criteria ($\mathcal{C}_{\mathrm{rel}}$) and ($\mathcal{C}_{\mathrm{nat}}$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}
SNF-GS	10.3	10.3	16.1	16.0	19.8	19.8	26.6	26.6	36.1	36.2
ADMM	13.2	7.7	27.1	15.5	39.8	22.3	49.4	36.5	62.9	51.7
FISTA	25.8	12.0	98.4	35.6	180.0	87.2	298.3	336.7	489.6	DNC
FPD-BB	38.6	10.9	65.6	41.9	88.8	68.6	107.0	123.7	DNC	181.6
PD	8.0	10.4	10.6	12.8	13.0	15.6	16.2	26.6	23.2	48.1
SpaRSA	15.0	17.3	30.8	19.4	45.2	32.2	57.9	69.9	74.7	107.7
SPG-GS	11.1	6.7	31.7	12.7	49.6	28.8	DNC	DNC	DNC	DNC

Table A.10.: Total CPU time (in sec.) t_{rel} and t_{nat} averaged over 10 independent runs with dynamic range 40 dB and $\bar{\sigma} = 0.01$ using the stopping criteria ($\mathcal{C}_{\mathrm{rel}}$) and ($\mathcal{C}_{\mathrm{nat}}$).

Method	$\varepsilon : 10^0$		$\varepsilon : 10^{-1}$		$\varepsilon : 10^{-2}$		$\varepsilon : 10^{-4}$		$\varepsilon : 10^{-6}$	
	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}	t_{rel}	t_{nat}
SNF-GS	16.1	16.0	19.1	19.1	35.9	35.8	49.8	50.2	66.0	66.0
ADMM	10.6	7.4	18.5	16.2	29.1	23.7	50.7	39.2	59.3	55.7
FISTA	74.0	7.0	169.9	49.7	418.1	145.0	DNC	DNC	DNC	DNC
FPD-BB	180.0	14.8	300.1	19.6	423.0	238.8	DNC	640.9	DNC	DNC
PD	21.1	9.5	26.0	33.3	30.2	41.2	37.1	53.0	40.4	62.1
SpaRSA	15.4	19.4	64.8	20.1	184.6	39.4	418.1	332.3	DNC	692.2
SPG-GS	12.8	13.4	106.2	15.2	267.8	30.4	582.6	466.5	DNC	DNC



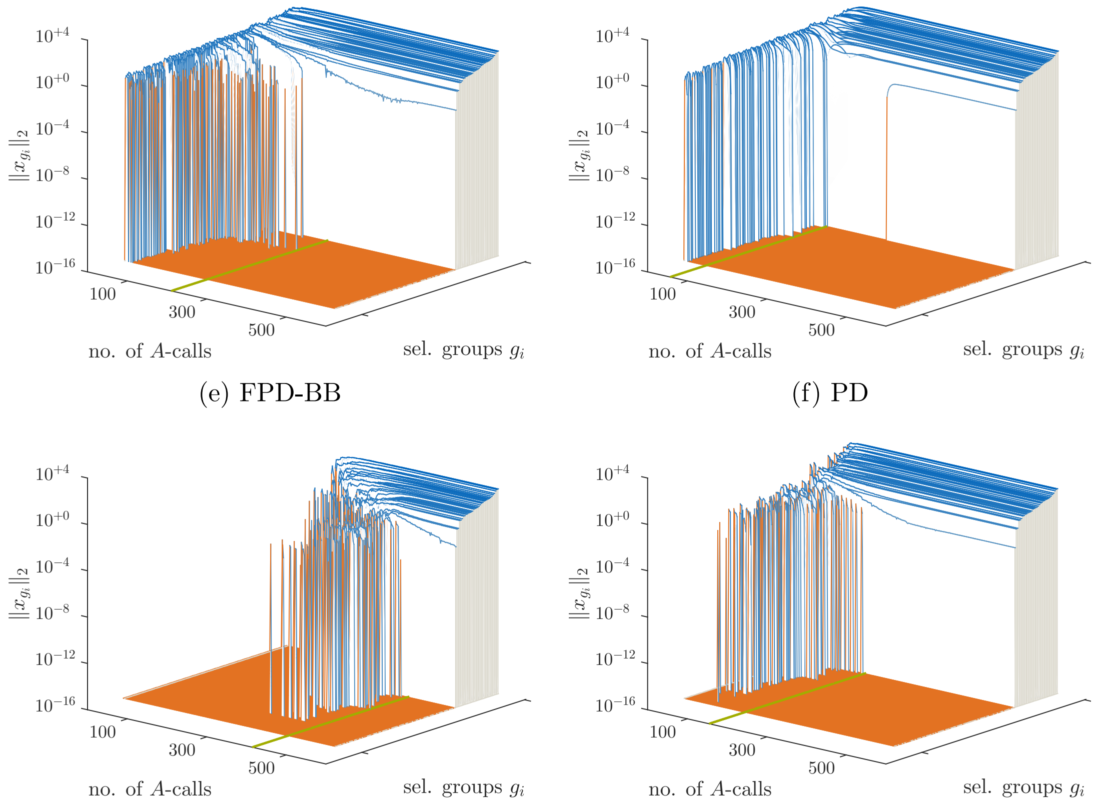
(a) SNF-GS

(b) SNF-GS-(sub)

(c) ADMM

(d) FISTA

Figure A.4.: Change of the ℓ_2-norm of 128 randomly chosen groups of the iterate of a single run with dynamic range 20 dB and $\bar{\sigma} = 0.1$ with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero groups of the high precision solution x^* within the sample.



(e) FPD-BB

(f) PD

(g) SpaRSA

(h) SPG-GS

Figure A.4.: Change of the ℓ_2-norm of 128 randomly chosen groups of the iterate of a single run with dynamic range 20 dB and $\bar{\sigma} = 0.1$ with respect to the number of A- and $A^\top$-calls. Green line: maximum number of A-calls that is needed to detect all zero groups of the high precision solution x^* within the sample.

Bibliography

[1] A. Aravkin, J. V. Burke, and G. Pillonetto, *Robust and trend-following Student's t Kalman smoothers*, SIAM J. Control Optim., 52 (2014), pp. 2891–2916.

[2] A. Aravkin, M. P. Friedlander, F. J. Herrmann, and T. van Leeuwen, *Robust inversion, dimensionality reduction, and randomized sampling*, Math. Program., 134 (2012), pp. 101–125.

[3] A. Aravkin, M. P. Friedlander, and T. van Leeuwen, *Robust inversion via semistochastic dimensionality reduction*, in IEEE Conf. on Acoustics, Speech and Signal Process. (ICASSP), IEEE, 2012, pp. 5245–5248.

[4] A. Aravkin, T. van Leeuwen, and F. J. Herrmann, *Robust full-waveform inversion using the student's t-distribution*, SEG Tech. Program Expanded Abstr., 30 (2011), pp. 2669–2673.

[5] G. Auchmuty, *Variational principles for variational inequalities*, Numer. Funct. Anal. Optim., 10 (1989), pp. 863–874.

[6] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, *Optimization with sparsity-inducing penalties*, Foundations and Trends® in Machine Learning, 4 (2011), pp. 1–106.

[7] F. R. Bach, *Consistency of the group lasso and multiple kernel learning*, J. Mach. Learn. Res., 9 (2008), pp. 1179–1225.

[8] C. Baiocchi and A. Capelo, *Variational and quasivariational inequalities*, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1984.

[9] W. Bajwa, J. Haupt, G. Raz, and R. Nowak, *Compressed channel sensing*, in Conf. on Info. Sciences and Systems (CISS), Princeton, New Jersey, March 2008, pp. 5–10.

[10] J. Barzilai and J. M. Borwein, *Two-point step size gradient methods*, IMA J. Numer. Anal., 8 (1988), pp. 141–148.

[11] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, New York, 2011.

[12] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[13] S. Becker, J. Bobin, and E. J. Candès, *NESTA: a fast and accurate first-order method for sparse recovery*, SIAM J. Imaging Sci., 4 (2011), pp. 1–39.

[14] S. Becker and M. J. Fadili, *A quasi-Newton proximal splitting method.* Preprint, available at arXiv, 1206.1156v2.

[15] A. Ben-Tal and J. Zowe, *Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems*, Math. Program., 24 (1982), pp. 70–91.

[16] ——, *Directional derivatives in nonsmooth optimization*, J. Optim. Theory Appl., 47 (1985), pp. 483–490.

[17] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, 1999. Second edition.

[18] R. Bhatia, *Matrix analysis*, vol. 169 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1997.

[19] J. M. Bioucas-Dias and M. A. T. Figueiredo, *A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration*, IEEE Trans. Image Process., 16 (2007), pp. 2992–3004.

[20] J. Bolte, A. Daniilidis, and A. Lewis, *The łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17 (2006), pp. 1205–1223 (electronic).

[21] ——, *Tame functions are semismooth*, Math. Program., 117 (2009), pp. 5–19.

[22] J. F. Bonnans and R. Cominetti, *Perturbed optimization in Banach spaces. I. A general theory based on a weak directional constraint qualification*, SIAM J. Control Optim., 34 (1996), pp. 1151–1171.

[23] J. F. Bonnans, R. Cominetti, and A. Shapiro, *Sensitivity analysis of optimization problems under second order regular constraints*, Math. Oper. Res., 23 (1998), pp. 806–831.

[24] ——, *Second order optimality conditions based on parabolic second order tangent sets*, SIAM J. Optim., 9 (1999), pp. 466–492.

[25] J. F. Bonnans and H. Ramírez C., *Perturbation analysis of second-order cone programming problems*, Math. Program., 104 (2005), pp. 205–227.

[26] J. F. Bonnans and A. Shapiro, *Nondegeneracy and quantitative stability of parameterized optimization problems with multiple solutions*, SIAM J. Optim., 8 (1998), pp. 940–946.

[27] ——, *Perturbation analysis of optimization problems*, Springer Series in Operations Research, Springer-Verlag, New York, 2000.

[28] J. V. Burke, *Descent methods for composite nondifferentiable optimization problems*, Math. Program., 33 (1985), pp. 260–279.

[29] ——, *Second order necessary and sufficient conditions for convex composite NDO*, Math. Program., 38 (1987), pp. 287–302.

[30] J. V. Burke and M. C. Ferris, *A Gauss-Newton method for convex composite optimization*, Math. Program., 71 (1995), pp. 179–194.

[31] J. V. Burke and R. A. Poliquin, *Optimality conditions for non-finite valued convex composite functions*, Math. Program., 57 (1992), pp. 103–120.

[32] R. H. Byrd, G. M. Chin, J. Nocedal, and F. Oztoprak, *A family of second-order methods for convex ℓ_1-regularized optimization*, Math. Program., (2015), pp. 1–33.

[33] J.-F. Cai, E. J. Candès, and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM J. Optim., 20 (2010), pp. 1956–1982.

[34] E. J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. Math., 9 (2009), pp. 717–772.

[35] E. J. Candès and J. Romberg, *Quantitative robust uncertainty principles and optimally sparse decompositions*, Found. Comput. Math., 6 (2006), pp. 227–254.

[36] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[37] E. J. Candès and T. Tao, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.

[38] ——, *The power of convex relaxation: near-optimal matrix completion*, IEEE Trans. Inform. Theory, 56 (2010), pp. 2053–2080.

[39] E. Casas, R. Herzog, and G. Wachsmuth, *Analysis of spatio-temporally sparse optimal control problems of semilinear parabolic equations.* Preprint, available at `https://www.tu-chemnitz.de/mathematik/part_dgl/publications.de.php`, March 2015.

[40] ——, *Optimality conditions and error analysis of semilinear elliptic control problems with L^1 cost functional*, SIAM J. Optim., 22 (2012), pp. 795–820.

[41] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

[42] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[43] Z. X. Chan and D. Sun, *Constraint nondegeneracy, strong regularity, and nonsingularity in semidefinite programming*, SIAM J. Optim., 19 (2008), pp. 370–396.

[44] C. Chen, Y.-J. Liu, D. Sun, and K.-C. Toh, *A semismooth Newton-CG based dual PPA for matrix spectral norm approximation problems*, Mathematical Program., (2014), pp. 1–36.

[45] J. Chen and X. Huo, *Theoretical results on sparse representations of multiple-measurement vectors*, IEEE Trans. Signal Process., 54 (2006), pp. 4634–4643.

[46] X. Chen, H. Qi, and P. Tseng, *Analysis of nonsmooth symmetric-matrix-valued functions with applications to semidefinite complementarity problems*, SIAM J. Optim., 13 (2003), pp. 960–985 (electronic).

[47] Y. Chen, R. Ranftl, and T. Pock, *A bi-level view of inpainting-based image compression.* Preprint, available at arXiv, 1401.4112v2.

[48] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, *Epigraphical projection and proximal tools for solving constrained convex optimization problems*, Signal, Image and Video Processing, (2014), pp. 1–13.

[49] F. H. Clarke, *On the inverse function theorem*, Pacific J. Math., 64 (1976), pp. 97–102.

[50] ——, *Optimization and nonsmooth analysis*, vol. 5 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 1990.

[51] C. Clason and T. Valkonen, *Stability of saddle points via explicit coderivatives of pointwise subdifferentials.* Preprint, submitted, available at `https://www.uni-due.de/~adf040p/preprints/SaddlepointStability.pdf`, September 2015.

[52] P. L. Combettes and J.-C. Pesquet, *Proximal thresholding algorithm for minimization over orthonormal bases*, SIAM J. Optim., 18 (2007), pp. 1351–1376.

[53] ——, *Proximal splitting methods in signal processing*, in Fixed-point Algorithms for Inverse Problems in Science and engineering, vol. 49 of Springer Optim. Appl., Springer, New York, 2011, pp. 185–212.

[54] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200 (electronic).

[55] R. Cominetti, *Metric regularity, tangent sets, and second-order optimality conditions*, Appl. Math. Optim., 21 (1990), pp. 265–287.

[56] ——, *On pseudo-differentiability*, Trans. Amer. Math. Soc., 324 (1991), pp. 843–865.

[57] M. Coste, *An introduction to o-minimal geometry.* Inst. Rech. Math., Univ. de Rennes, available at `https://perso.univ-rennes1.fr/michel.coste/polyens/OMIN.pdf`, November 1999.

[58] ——, *An introduction to semialgebraic geometry.* Inst. Rech. Math., Univ. de Rennes, `https://perso.univ-rennes1.fr/michel.coste/polyens/SAG.pdf`, October 2002.

[59] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, *Sparse solutions to linear inverse problems with multiple measurement vectors*, IEEE Trans. Signal Process., 53 (2005), pp. 2477–2488.

[60] A. Daniilidis, C. Sagastizábal, and M. Solodov, *Identifying structure of nonsmooth convex functions by the bundle technique*, SIAM J. Optim., 20 (2009), pp. 820–840.

[61] W. Deng, W. Yin, and Y. Zhang, *Group sparse optimization by alternating direction method.* Technical Report, TR11-06, Department of Computational and Applied Mathematics, Rice University, 2011.

[62] J. Dieudonné, *Foundations of modern analysis*, Academic Press, New York-London, 1969.

[63] C. Ding, *An introduction to a class of matrix optimization problems*, PhD Dissertation, National University of Singapore, 2012.

[64] C. Ding, D. Sun, and K.-C. Toh, *An introduction to a class of matrix cone programming*, Math. Program., 144 (2014), pp. 141–179.

[65] Y. Dong, M. Hintermüller, and M. Neri, *An efficient primal-dual method for* L^1TV *image restoration*, SIAM J. Imaging Sci., 2 (2009), pp. 1168–1189.

[66] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[67] A. L. Dontchev and R. T. Rockafellar, *Implicit functions and solution mappings*, Springer Series in Operations Research and Financial Engineering, Springer, New York, second ed., 2014.

[68] A. Dreves, A. von Heusinger, C. Kanzow, and M. Fukushima, *A globalized Newton method for the computation of normalized Nash equilibria*, J. Global Optim., 56 (2013), pp. 327–340.

[69] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, *Single-pixel imaging via compressive sampling*, IEEE Signal Process. Mag., 25 (2008), pp. 83–91.

[70] M. F. Duarte, S. Sarvotham, M. B. Wakin, D. Baron, and R. G. Baraniuk, *Joint sparsity models for distributed compressed sensing*, in Online Proceedings of the Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS), Rennes, France, 2005.

[71] J. Eckstein and D. P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Math. Program., 55 (1992), pp. 293–318.

[72] F. Facchinei, A. Fischer, and C. Kanzow, *Regularity properties of a semismooth reformulation of variational inequalities*, SIAM J. Optim., 8 (1998), pp. 850–869 (electronic).

[73] F. Facchinei and C. Kanzow, *Generalized Nash equilibrium problems*, Ann. Oper. Res., 175 (2010), pp. 177–211.

[74] F. Facchinei, C. Kanzow, and S. Sagratella, *Solving quasi-variational inequalities via their KKT conditions*, Math. Program., 144 (2014), pp. 369–412.

[75] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems. Vol. I*, Springer Series in Operations Research, Springer-Verlag, New York, 2003.

[76] ——, *Finite-dimensional variational inequalities and complementarity problems. Vol. II*, Springer Series in Operations Research, Springer-Verlag, New York, 2003.

[77] M. C. Ferris and J.-S. Pang, *Engineering and economic applications of complementarity problems*, SIAM Rev., 39 (1997), pp. 669–713.

[78] M. Figueiredo, R. D. Nowak, and S. J. Wright, *Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems*, IEEE J. Sel. Topics Signal Process., 1 (2007), pp. 586–598.

[79] A. Fischer, *Solution of monotone complementarity problems with locally Lipschitzian functions*, Math. Program., 76 (1997), pp. 513–532.

[80] S. Fitzpatrick and R. R. Phelps, *Differentiability of the metric projection in Hilbert space*, Trans. Amer. Math. Soc., 270 (1982), pp. 483–501.

[81] R. Fletcher, *Practical methods of optimization*, A Wiley-Interscience Publication, John Wiley & Sons, New York, second ed., 2001.

[82] R. Fletcher, N. I. M. Gould, S. Leyffer, P. L. Toint, and A. Wächter, *Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming*, SIAM J. Optim., 13 (2002), pp. 635–659 (2003).

[83] R. Fletcher and S. Leyffer, *Nonlinear programming without a penalty function*, Math. Program., 91 (2002), pp. 239–269.

[84] R. Fletcher, S. Leyffer, and P. L. Toint, *On the global convergence of a filter-SQP algorithm*, SIAM J. Optim., 13 (2002), pp. 44–59 (electronic).

[85] M. Fornasier and H. Rauhut, *Recovery algorithms for vector-valued data with joint sparsity constraints*, SIAM J. Numer. Anal., 46 (2008), pp. 577–613.

[86] M. Fukushima, *Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems*, Math. Program., 53 (1992), pp. 99–110.

[87] M. Fukushima, Z.-Q. Luo, and P. Tseng, *Smoothing functions for second-order-cone complementarity problems*, SIAM J. Optim., 12 (2001/02), pp. 436–460 (electronic).

[88] M. Fukushima and H. Mine, *A generalized proximal point algorithm for certain nonconvex minimization problems*, Internat. J. Systems Sci., 12 (1981), pp. 989–1000.

[89] D. Gabay and B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximation*, Computers & Mathematics with Applications, 2 (1976), pp. 17–40.

[90] C. Geiger and C. Kanzow, *Theorie und Numerik restringiererter Optimierungsaufgaben*, Springer-Lehrbuch Masterclass, Springer Berlin Heidelberg, 2002.

[91] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*, vol. 9 of SIAM Studies in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1989.

[92] R. Glowinski and A. Marrocco, *Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires*, Rev. Française Automat. Informat. Recherche Opérationnelle, 9 (1975), pp. 41–76.

[93] T. Goldstein, B. O'Donoghue, S. Setzer, and R. Baraniuk, *Fast alternating direction optimization methods*, SIAM J. Imaging Sci., 7 (2014), pp. 1588–1623.

[94] G. H. Golub and C. F. Van Loan, *Matrix computations*, Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD, fourth ed., 2013.

[95] N. I. M. Gould, S. Leyffer, and P. L. Toint, *A multidimensional filter algorithm for nonlinear equations and nonlinear least-squares*, SIAM J. Optim., 15 (2004), pp. 17–38.

[96] N. I. M. Gould, C. Sainvitu, and P. L. Toint, *A filter-trust-region method for unconstrained optimization*, SIAM J. Optim., 16 (2005), pp. 341–357.

[97] R. Griesse and D. A. Lorenz, *A semismooth Newton method for Tikhonov functionals with sparsity constraints*, Inverse Problems, 24 (2008), pp. 035007, 19.

[98] E. T. Hale, W. Yin, and Y. Zhang, *Fixed-point continuation for l_1-minimization: methodology and convergence*, SIAM J. Optim., 19 (2008), pp. 1107–1130.

[99] E. Hans and T. Raasch, *Global convergence of damped semismooth Newton methods for ℓ_1 Tikhonov regularization*, Inverse Problems, 31 (2015), pp. 025005, 31.

[100] W. L. Hare, *Nonsmooth optimization with smooth substructure*, PhD Dissertation, Simon Fraser University, 2004.

[101] ——, *Functions and sets of smooth substructure: relationships and examples*, Comput. Optim. Appl., 33 (2006), pp. 249–270.

[102] ——, *Numerical analysis of $\mathcal{VU}$-decomposition, $\mathcal{U}$-gradient, and $\mathcal{U}$-Hessian approximations*, SIAM J. Optim., 24 (2014), pp. 1890–1913.

[103] W. L. Hare and R. A. Poliquin, *The quadratic sub-Lagrangian of a prox-regular function*, in Proceedings of the Third World Congress of Nonlinear Analysts, Part 2 (Catania, 2000), vol. 47, 2001, pp. 1117–1128.

[104] P. T. Harker and J.-S. Pang, *Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications*, Math. Program., 48 (1990), pp. 161–220.

[105] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM J. Imaging Sci., 5 (2012), pp. 119–149.

[106] R. Herzog, G. Stadler, and G. Wachsmuth, *Directional sparsity in optimal control of partial differential equations*, SIAM J. Control Optim., 50 (2012), pp. 943–963.

[107] N. J. Higham, *Computing a nearest symmetric positive semidefinite matrix*, Linear Algebra Appl., 103 (1988), pp. 103–118.

[108] J.-B. Hiriart-Urruty, *The approximate first-order and second-order directional derivatives for a convex function*, in Mathematical Theories of Optimization (Genova, 1981), vol. 979 of Lecture Notes in Math., Springer, Berlin-New York, 1983, pp. 144–177.

[109] J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of convex analysis*, Grundlehren Text Editions, Springer-Verlag, Berlin, 2001.

[110] J.-B. Hiriart-Urruty, J.-J. Strodiot, and V. H. Nguyen, *Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data*, Appl. Math. Optim., 11 (1984), pp. 43–56.

[111] L. Hoeltgen, S. Setzer, and J. Weickert, *An optimal control approach to find sparse data for Laplace interpolation*, in Energy Minimization Methods in Computer Vision and Pattern Recognition, vol. 8081 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 151–164.

[112] M. Huang, L.-P. Pang, and Z.-Q. Xia, *The space decomposition theory for a class of eigenvalue optimizations*, Comput. Optim. Appl., 58 (2014), pp. 423–454.

[113] T. Ichiishi, *Game theory for economic analysis*, Economic Theory, Econometrics, and Mathematical Economics, Academic Press, Inc., New York, 1983.

[114] A. D. Ioffe, *An invitation to tame optimization*, SIAM J. Optim., 19 (2008), pp. 1894–1917.

[115] R. Jenatton, J.-Y. Audibert, and F. Bach, *Structured variable selection with sparsity-inducing norms*, J. Mach. Learn. Res., 12 (2011), pp. 2777–2824.

[116] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, *Proximal methods for hierarchical sparse coding*, J. Mach. Learn. Res., 12 (2011), pp. 2297–2334.

[117] K. Jiang, D. Sun, and K.-C. Toh, *An inexact accelerated proximal gradient method for large scale linearly constrained convex SDP*, SIAM J. Optim., 22 (2012), pp. 1042–1064.

[118] ——, *Solving nuclear norm regularized and semidefinite matrix least squares problems with linear equality constraints*, in Discrete Geometry and Optimization, vol. 69 of Fields Inst. Commun., Springer, New York, 2013, pp. 133–162.

[119] ——, *A partial proximal point algorithm for nuclear norm regularized matrix least squares problems*, Math. Program. Comput., 6 (2014), pp. 281–325.

[120] C. Kanzow and M. Fukushima, *Solving box constrained variational inequalities by using the natural residual with D-gap function globalization*, Oper. Res. Lett., 23 (1998), pp. 45–51.

[121] ——, *Theoretical and numerical investigation of the D-gap function for box constrained variational inequalities*, Math. Program., 83 (1998), pp. 55–87.

[122] M. Karow, *Geometry of spectral value sets*, PhD Dissertation, Universität Bremen, 2003.

[123] K. Koh, S.-J. Kim, and S. Boyd, *An interior-point method for large-scale l_1-regularized logistic regression*, J. Mach. Learn. Res., 8 (2007), pp. 1519–1555.

[124] M. Kojima and S. Shindo, *Extension of Newton and quasi-Newton methods to systems of PC^1 equations*, J. Oper. Res. Soc. Japan, 29 (1986), pp. 352–375.

[125] Y. Koren, R. Bell, and C. Volinsky, *Matrix factorization techniques for recommender systems*, Computer, 42 (2009), pp. 30–37.

[126] A. S. Kravchuk and P. J. Neittaanmäki, *Variational and quasi-variational inequalities in mechanics*, vol. 147 of Solid Mechanics and its Applications, Springer, Dordrecht, 2007.

[127] J. B. Kruskal, *Two convex counterexamples: A discontinuous envelope function and a nondifferentiable nearest-point mapping*, Proc. Amer. Math. Soc., 23 (1969), pp. 697–703.

[128] B. Kummer, *Lipschitzian inverse functions, directional derivatives, and applications in $C^{1,1}$ optimization*, J. Optim. Theory Appl., 70 (1991), pp. 561–581.

[129] J. D. Lee, Y. Sun, and M. A. Saunders, *Proximal Newton-type methods for minimizing composite functions*, SIAM J. Optim., 24 (2014), pp. 1420–1443.

[130] C. Lemaréchal, F. Oustry, and C. Sagastizábal, *The $\mathcal{U}$-Lagrangian of a convex function*, Trans. Amer. Math. Soc., 352 (2000), pp. 711–729.

[131] C. Lemaréchal and C. Sagastizábal, *More than first-order developments of convex functions: primal-dual relations*, J. Convex Anal., 3 (1996), pp. 255–268.

[132] ——, *Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries*, SIAM J. Optim., 7 (1997), pp. 367–385.

[133] A. S. Lewis, *Active sets, nonsmoothness, and sensitivity*, SIAM J. Optim., 13 (2002), pp. 702–725 (electronic) (2003).

[134] A. S. Lewis and S. J. Wright, *A proximal method for composite minimization.* Preprint, available at arXiv, 0812.0423v2.

[135] C. Li and X. Wang, *On convergence of the Gauss-Newton method for convex composite optimization*, Math. Program., 91 (2002), pp. 349–356.

[136] C. Li, W. Yin, H. Jiang, and Y. Zhang, *An efficient augmented Lagrangian method with applications to total variation minimization*, Comput. Optim. Appl., 56 (2013), pp. 507–530.

[137] T. T. Y. Lin and F. J. Herrmann, *Compressed wavefield extrapolation*, Geophysics, 72 (2007), pp. 77–93.

[138] Y.-J. Liu, D. Sun, and K.-C. Toh, *An implementable proximal point algorithmic framework for nuclear norm minimization*, Math. Program., 133 (2012), pp. 399–436.

[139] D. A. Lorenz, *Constructing test instances for basis pursuit denoising*, IEEE Trans. Signal Process., 61 (2013), pp. 1210–1214.

[140] Y. Lu, L.-P. Pang, F.-F. Guo, and Z.-Q. Xia, *A superlinear space decomposition algorithm for constrained nonsmooth convex program*, J. Comput. Appl. Math., 234 (2010), pp. 224–232.

[141] M. Lustig, D. L. Donoho, and J. M. Pauly, *Sparse MRI: The application of compressed sensing for rapid MR imaging*, Mag. Resonance Med., 58 (2007), pp. 1182–1195.

[142] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, *Compressed sensing MRI*, IEEE Signal Process. Mag., 25 (2007), pp. 72–82.

[143] D. Malioutov, M. Cetin, and A. Willsky, *A sparse signal reconstruction perspective for source localization with sensor arrays*, IEEE Trans. Signal Process., 53 (2005), pp. 3010–3022.

[144] L. Meier, S. van de Geer, and P. Bühlmann, *The group Lasso for logistic regression*, J. R. Stat. Soc. Ser. B Stat. Methodol., 70 (2008), pp. 53–71.

[145] F. Meng, D. Sun, and G. Zhao, *Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization*, Math. Program., 104 (2005), pp. 561–581.

[146] F. Meng, G. Zhao, M. Goh, and R. De Souza, *Lagrangian-dual functions and Moreau-Yosida regularization*, SIAM J. Optim., 19 (2008), pp. 39–61.

[147] C. A. Micchelli, L. Shen, and Y. Xu, *Proximity algorithms for image models: denoising*, Inverse Problems, 27 (2011), pp. 045009, 30.

[148] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optimization, 15 (1977), pp. 959–972.

[149] R. Mifflin, L. Qi, and D. Sun, *Properties of the Moreau-Yosida regularization of a piecewise C^2 convex function*, Math. Program., 84 (1999), pp. 269–281.

[150] R. Mifflin and C. Sagastizábal, *On $\mathcal{VU}$-theory for functions with primal-dual gradient structure*, SIAM J. Optim., 11 (2000), pp. 547–571 (electronic).

[151] ——, *Proximal points are on the fast track*, J. Convex Anal., 9 (2002), pp. 563–579.

[152] ——, *Primal-dual gradient structured functions: second-order results; links to epi-derivatives and partly smooth functions*, SIAM J. Optim., 13 (2003), pp. 1174–1194 (electronic).

[153] ——, *On the relation between $\mathcal{U}$-Hessians and second-order epi-derivatives*, European J. Oper. Res., 157 (2004), pp. 28–38.

[154] ——, *A $\mathcal{VU}$-algorithm for convex minimization*, Math. Program., 104 (2005), pp. 583–608.

[155] ——, *Relating $\mathcal{U}$-Lagrangians to second-order epi-derivatives and proximal-tracks*, J. Convex Anal., 12 (2005), pp. 81–93.

[156] A. Milzarek, *Ein semiglattes Newton-Verfahren mit mehrdimensionaler Filter-Globalisierung zur Lösung von ℓ_1-Minimierungsproblemen*, Bachelor's Thesis, Technische Universität München, 2010.

[157] A. Milzarek and M. Ulbrich, *A semismooth Newton method with multidimensional filter globalization for l_1-optimization*, SIAM J. Optim., 24 (2014), pp. 298–333.

[158] B. S. Mordukhovich, *Variational analysis and generalized differentiation. I*, vol. 330 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 2006.

[159] B. S. Mordukhovich, J. V. Outrata, and M. E. Sarabi, *Full stability of locally optimal solutions in second-order cone programs*, SIAM J. Optim., 24 (2014), pp. 1581–1613.

[160] J. J. Moré and D. C. Sorensen, *Computing a trust region step*, SIAM J. Sci. Statist. Comput., 4 (1983), pp. 553–572.

[161] J.-J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, C. R. Acad. Sci. Paris, 255 (1962), pp. 2897–2899.

[162] ——, *Propriétés des applications "prox"*, C. R. Acad. Sci. Paris, 256 (1963), pp. 1069–1071.

[163] ——, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France, 93 (1965), pp. 273–299.

[164] U. Mosco, *Implicit variational problems and quasi variational inequalities*, in Nonlinear Operators and the Calculus of Variations (Summer School, Univ. Libre Bruxelles, Brussels, 1975), Springer, Berlin, 1976, pp. 83–156. Lecture Notes in Math., Vol. 543.

[165] N. Movahedian, *Nonsmooth calculus of semismooth functions and maps*, J. Optim. Theory Appl., 160 (2014), pp. 415–438.

[166] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig, *Fast ℓ_1-SPIRiT compressed sensing parallel imaging MRI: scalable parallel implementation and clinically feasible runtime*, IEEE Trans. Med. Imag., 31 (2012), pp. 1250–1262.

[167] Z. Naniewicz and P. D. Panagiotopoulos, *Mathematical theory of hemivariational inequalities and applications*, vol. 188 of Monographs and Textbooks in Pure and Applied Mathematics, Marcel Dekker, Inc., New York, 1995.

[168] Y. Nesterov, *Smooth minimization of non-smooth functions*, Math. Program., 103 (2005), pp. 127–152.

[169] ——, *Gradient methods for minimizing composite functions*, Math. Program., 140 (2013), pp. 125–161.

[170] A. Y. Ng, *Feature selection, L_1 vs. L_2 regularization, and rotational invariance*, in Int. Conf. on Mach. Learn. (ICML), Banff, Canada, 2004.

[171] M. K. Ng, R. H. Chan, and W.-C. Tang, *A fast algorithm for deblurring models with Neumann boundary conditions*, SIAM J. Sci. Comput., 21 (1999), pp. 851–866 (electronic).

[172] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*, Cambridge University Press, New York, NY, USA, 2007.

[173] M. A. Noor, *Mixed variational inequalities*, Appl. Math. Lett., 3 (1990), pp. 73–75.

[174] M. A. Noor, *General nonlinear mixed variational-like inequalities*, Optimization, 37 (1996), pp. 357–367.

[175] M. A. Noor, *Generalized mixed variational inequalities and resolvent equations*, Positivity, 1 (1997), pp. 145–154.

[176] ——, *Some developments in general variational inequalities*, Appl. Math. Comput., 152 (2004), pp. 199–277.

[177] M. A. Noor, K. I. Noor, and E. Al-Said, *Existence results for extended general nonconvex quasi-variational inequalities*, in Nonlinear Analysis, vol. 68 of Springer Optim. Appl., Springer, New York, 2012, pp. 503–512.

[178] P. Ochs, Y. Chen, T. Brox, and T. Pock, *iPiano: inertial proximal algorithm for nonconvex optimization*, SIAM J. Imaging Sci., 7 (2014), pp. 1388–1419.

[179] F. Oustry, *A second-order bundle method to minimize the maximum eigenvalue function*, Math. Program., 89 (2000), pp. 1–33.

[180] J. Outrata and H. Ramírez C., *On the Aubin property of critical points to perturbed second-order cone programs*, SIAM J. Optim., 21 (2011), pp. 798–823.

[181] P. D. Panagiotopoulos, *Hemivariational inequalities*, Springer-Verlag, Berlin, 1993.

[182] J.-S. Pang and L. Qi, *Nonsmooth equations: motivation and algorithms*, SIAM J. Optim., 3 (1993), pp. 443–465.

[183] J.-S. Pang, D. Sun, and J. Sun, *Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems*, Math. Oper. Res., 28 (2003), pp. 39–63.

[184] N. Parikh and S. Boyd, *Proximal algorithms*, Foundations and Trends® in Optimization, 1 (2014), pp. 127–239.

[185] M. Patriksson, *Merit functions and descent algorithms for a class of variational inequality problems*, Optimization, 41 (1997), pp. 37–55.

[186] M. Patriksson, *Nonlinear programming and variational inequality problems*, vol. 23 of Appl. Optim., Kluwer Acad. Publ., Dordrecht, 1999.

[187] P. Patrinos, L. Stella, and A. Bemporad, *Forward-backward truncated Newton methods for convex composite optimization.* Preprint, available at arXiv, 1402.6655.

[188] J.-M. Peng, *Equivalence of variational inequality problems to unconstrained minimization*, Math. Program., 78 (1997), pp. 347–355.

[189] J.-M. Peng and M. Fukushima, *A hybrid Newton method for solving the variational inequality problem via the D-gap function*, Math. Program., 86 (1999), pp. 367–386.

[190] J.-M. Peng, C. Kanzow, and M. Fukushima, *A hybrid Josephy-Newton method for solving box constrained variational inequality problems via the D-gap function*, Optim. Methods Softw., 10 (1999), pp. 687–710.

[191] K. Pieper, *Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems*, PhD Dissertation, Technische Universität München, 2015.

[192] T. Pock and A. Chambolle, *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, in IEEE Int. Conf. on Comp. Vision, (ICCV), Nov 2011, pp. 1762–1769.

[193] R. A. Poliquin and R. T. Rockafellar, *Amenable functions in optimization*, in Nonsmooth Optimization: Methods and Applications (Erice, 1991), Gordon and Breach, Montreux, 1992, pp. 338–353.

[194] ——, *A calculus of epi-derivatives applicable to optimization*, Canad. J. Math., 45 (1993), pp. 879–896.

[195] ——, *Generalized Hessian properties of regularized nonsmooth functions*, SIAM J. Optim., 6 (1996), pp. 1121–1137.

[196] ——, *Prox-regular functions in variational analysis*, Trans. Amer. Math. Soc., 348 (1996), pp. 1805–1838.

[197] L. Qi, *Convergence analysis of some algorithms for solving nonsmooth equations*, Math. Oper. Res., 18 (1993), pp. 227–244.

[198] L. Qi and D. Sun, *A survey of some nonsmooth equations and smoothing Newton methods*, in Progress in Optimization, vol. 30 of Appl. Optim., Kluwer Acad. Publ., Dordrecht, 1999, pp. 121–146.

[199] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.

[200] Z. Qin, K. Scheinberg, and D. Goldfarb, *Efficient block-coordinate descent algorithms for the group Lasso*, Math. Program. Comput., 5 (2013), pp. 143–169.

[201] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501.

[202] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.

[203] ——, *Generalized equations and their solutions. II. Applications to nonlinear programming*, Math. Programming Stud., (1982), pp. 200–221. Optimality and Stability in Math. Program.

[204] ——, *Local structure of feasible sets in nonlinear programming. II. Nondegeneracy*, Math. Programming Stud., (1984), pp. 217–230. Math. Program. at Oberwolfach, II (Oberwolfach, 1983).

[205] ——, *Local structure of feasible sets in nonlinear programming. III. Stability and sensitivity*, Math. Programming Stud., (1987), pp. 45–66. Nonlinear Analysis and Optimization (Louvain-la-Neuve, 1983).

[206] R. T. Rockafellar, *First- and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc., 307 (1988), pp. 75–108.

[207] ——, *Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Math. Oper. Res., 14 (1989), pp. 462–484.

[208] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, vol. 317 of Grundlehren der Mathematischen Wissenschaften, Springer-Verlag, Berlin, 1998.

[209] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.

[210] S. Salzo and S. Villa, *Convergence analysis of a proximal Gauss-Newton method*, Comput. Optim. Appl., 53 (2012), pp. 557–589.

[211] S. Scholtes, *Introduction to piecewise differentiable equations*, Springer Briefs in Optimization, Springer, New York, 2012.

[212] A. Shapiro, *Differentiability properties of metric projections onto convex sets.* Preprint, available at `http://www.optimization-online.org/DB_HTML/2013/11/4119.html`, November 2013.

[213] ——, *On differentiability of metric projections in $\mathbf{R}^n$. I. Boundary case*, Proc. Amer. Math. Soc., 99 (1987), pp. 123–128.

[214] ——, *Directionally nondifferentiable metric projection*, J. Optim. Theory Appl., 81 (1994), pp. 203–204.

[215] ——, *First and second order analysis of nonlinear semidefinite programs*, Math. Program., 77 (1997), pp. 301–320. Semidefinite Programming.

[216] ——, *On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints*, SIAM J. Optim., 7 (1997), pp. 508–518.

[217] ——, *On a class of nonsmooth composite functions*, Math. Oper. Res., 28 (2003), pp. 677–692.

[218] ——, *Sensitivity analysis of generalized equations*, J. Math. Sci. (N. Y.), 115 (2003), pp. 2554–2565. Optimization and Related Topics, 1.

[219] ——, *Sensitivity analysis of parameterized variational inequalities*, Math. Oper. Res., 30 (2005), pp. 109–126.

[220] A. Shapiro and M. K. H. Fan, *On eigenvalue optimization*, SIAM J. Optim., 5 (1995), pp. 552–569.

[221] J. Shi, W. Yin, S. Osher, and P. Sajda, *A fast hybrid algorithm for large-scale ℓ_1-regularized logistic regression*, J. Mach. Learn. Res., 11 (2010), pp. 713–741.

[222] Z. Shi and R. Liu, *Large scale optimization with proximal stochastic newton-type gradient descent*, in Machine Learning and Knowledge Discovery in Databases, vol. 9284 of Lecture Notes in Computer Science, Springer International Publishing, 2015, pp. 691–704.

[223] M. V. Solodov, *Merit functions and error bounds for generalized variational inequalities*, J. Math. Anal. Appl., 287 (2003), pp. 405–414.

[224] M. V. Solodov and P. Tseng, *Some methods based on the D-gap function for solving monotone variational inequalities*, Comput. Optim. Appl., 17 (2000), pp. 255–277.

[225] G. Stadler, *Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices*, Comput. Optim. Appl., 44 (2009), pp. 159–181.

[226] D. Sun, *The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications*, Math. Oper. Res., 31 (2006), pp. 761–776.

[227] D. Sun, M. Fukushima, and L. Qi, *A computable generalized Hessian of the D-gap function and Newton-type methods for variational inequality problems*, in Complementarity and Variational Problems (Baltimore, MD, 1995), SIAM, Philadelphia, PA, 1997, pp. 452–473.

[228] D. Sun and J. Han, *On a conjecture in Moreau-Yosida approximation of a nonsmooth convex function*, Chinese Sci. Bull., 42 (1997), pp. 1423–1426.

[229] D. Sun and J. Sun, *Semismooth matrix-valued functions*, Math. Oper. Res., 27 (2002), pp. 150–169.

[230] L. Sun, J. Liu, J. Chen, and J. Ye, *Efficient recovery of jointly sparse vectors*, in Advances in Neural Information Processing Systems, (NIPS), 23, 2009.

[231] K. Taji, M. Fukushima, and T. Ibaraki, *A globally convergent Newton method for solving strong monotone variational inequalities*, Math. Program., 58 (1993), pp. 369–383.

[232] Q. Tran-Dinh, A. Kyrillidis, and V. Cevher, *An inexact proximal path-following algorithm for constrained convex minimization*, SIAM J. Optim., 24 (2014), pp. 1718–1745.

[233] J. S. Treiman, *The linear nonconvex generalized gradient and Lagrange multipliers*, SIAM J. Optim., 5 (1995), pp. 670–680.

[234] ——, *The linear nonconvex generalized gradient*, in World Congress of Nonlinear Analysts '92, Vol. I–IV (Tampa, FL, 1992), de Gruyter, Berlin, 1996, pp. 2325–2336.

[235] J. A. Tropp, *Algorithms for simultaneous sparse approximation: Part II: Convex relaxation*, Signal Process., 86 (2006), pp. 589–602.

[236] P. Tseng and S. Yun, *A coordinate gradient descent method for nonsmooth separable minimization*, Math. Program., 117 (2009), pp. 387–423.

[237] M. Ulbrich, *Nonmonotone trust-region methods for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problems*, SIAM J. Optim., 11 (2001), pp. 889–917.

[238] ——, *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*, vol. 11 of MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.

[239] M. Ulbrich and S. Ulbrich, *Nichtlineare Optimierung*, Birkhäuser, Basel, 2012.

[240] M. Ulbrich, S. Ulbrich, and L. N. Vicente, *A globally convergent primal-dual interior-point filter method for nonlinear programming*, Math. Program., 100 (2004), pp. 379–410.

[241] T. Valkonen, *A primal-dual hybrid gradient method for nonlinear operators with applications to MRI*, Inverse Problems, 30 (2014), pp. 055012, 45.

[242] E. van den Berg and M. P. Friedlander, *Probing the Pareto frontier for basis pursuit solutions*, SIAM J. Sci. Comput., 31 (2008/09), pp. 890–912.

[243] ——, *Theoretical and empirical results for recovery from multiple measurements*, IEEE Trans. Info. Theory, 56 (2010), pp. 2516–2527.

[244] L. van den Dries, *Tame topology and o-minimal structures*, vol. 248 of London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, 1998.

[245] A. von Heusinger, *Numerical methods for the solution of the generalized Nash equilibrium problem*, PhD Dissertation, Universität Würzburg, 2009.

[246] A. von Heusinger and C. Kanzow, *Optimization reformulations of the generalized Nash equilibrium problem using Nikaido-Isoda-type functions*, Comput. Optim. Appl., 43 (2009), pp. 353–377.

[247] A. von Heusinger, C. Kanzow, and M. Fukushima, *Newton's method for computing a normalized equilibrium in the generalized Nash game through fixed point formulation*, Math. Program., 132 (2012), pp. 99–123.

[248] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, *An architecture for compressive imaging*, in IEEE Int. Conf. on Imag. Process. (ICIP), Oct 2006, pp. 1273–1276.

[249] Y. Wang, J. Yang, W. Yin, and Y. Zhang, *A new alternating minimization algorithm for total variation image reconstruction*, SIAM J. Imaging Sci., 1 (2008), pp. 248–272.

[250] Y. Wang and L. Zhang, *Properties of equation reformulation of the Karush-Kuhn-Tucker condition for nonlinear second order cone optimization problems*, Math. Methods Oper. Res., 70 (2009), pp. 195–218.

[251] Z. Wen, A. Milzarek, M. Ulbrich, and H. Zhang, *Adaptive regularized self-consistent field iteration with exact Hessian for electronic structure calculation*, SIAM J. Sci. Comput., 35 (2013), pp. A1299–A1324.

[252] Z. Wen and W. Yin, *A feasible method for optimization with orthogonality constraints*, Math. Program., 142 (2013), pp. 397–434.

[253] Z. Wen, W. Yin, D. Goldfarb, and Y. Zhang, *A fast algorithm for sparse reconstruction based on shrinkage, subspace optimization, and continuation*, SIAM J. Sci. Comput., 32 (2010), pp. 1832–1857.

[254] Z. Wen, W. Yin, H. Zhang, and D. Goldfarb, *On the convergence of an active-set method for ℓ_1 minimization*, Optim. Methods Softw., 27 (2012), pp. 1127–1146.

[255] R. S. Womersley, *Local properties of algorithms for minimizing nonsmooth composite functions*, Math. Program., 32 (1985), pp. 69–89.

[256] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493.

[257] N. Yamashita, K. Taji, and M. Fukushima, *Unconstrained optimization reformulations of variational inequality problems*, J. Optim. Theory Appl., 92 (1997), pp. 439–456.

[258] J. Yang and Y. Zhang, *Alternating direction algorithms for ℓ_1-problems in compressive sensing*, SIAM J. Sci. Comput., 33 (2011), pp. 250–278.

[259] J. Yang, Y. Zhang, and W. Yin, *An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise*, SIAM J. Sci. Comput., 31 (2009), pp. 2842–2865.

[260] L. Yang, D. Sun, and K.-C. Toh, SDPNAL+*: a majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints*, Math. Program. Comput., 7 (2015), pp. 331–366.

[261] Y.-L. Yu, *On decomposing the proximal map*, in Advances in Neural Information Processing Systems 26, Curran Associates, Inc., 2013, pp. 91–99.

[262] M. Yuan and Y. Lin, *Model selection and estimation in regression with grouped variables*, J. R. Stat. Soc. Ser. B Stat. Methodol., 68 (2006), pp. 49–67.

[263] Y. Yuan, *Conditions for convergence of trust region algorithms for nonsmooth optimization*, Math. Program., 31 (1985), pp. 220–228.

[264] ——, *On the superlinear convergence of a trust region algorithm for nonsmooth optimization*, Math. Program., 31 (1985), pp. 269–285.

[265] S. Yun and K.-C. Toh, *A coordinate gradient descent method for ℓ_1-regularized convex minimization*, Comput. Optim. Appl., 48 (2011), pp. 273–307.

[266] E. H. Zarantonello, *Projections on convex sets in Hilbert space and spectral theory*, in Contributions to Nonlinear Functional Analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971), Academic Press, New York, 1971, pp. 237–424.

[267] H. Zhang and W. W. Hager, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. Optim., 14 (2004), pp. 1043–1056 (electronic).

[268] X.-Y. Zhao, D. Sun, and K.-C. Toh, *A Newton-CG augmented Lagrangian method for semidefinite programming*, SIAM J. Optim., 20 (2010), pp. 1737–1765.

[269] D. L. Zhu and P. Marcotte, *Modified descent methods for solving the monotone variational inequality problem*, Oper. Res. Lett., 14 (1993), pp. 111–120.