January 2020

Variational Analysis Of Composite Optimization

Ashkan Mohammadi
Wayne State University

VARIATIONAL ANALYSIS OF COMPOSITE OPTIMIZATION

by

ASHKAN MOHAMMADI

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2020

MAJOR: (Applied) MATHEMATICS

Approved By:

Advisor Date

DEDICATION

To My Parents

ACKNOWLEDGEMENTS

It would be impossible to adequately express my gratitude and appreciation for my advisor, Professor Boris Mordukhovich. His ceaseless devotion, support, and encouragement has been the driving force behind all my accomplishments during my doctoral studies. I am not sure the quality of my research would remain same if I had any other advisor in my field. I would like specially thank Dr. Ebrahim Sarabi who has been a coauthor of my papers during my PhD study. I value and appreciate the time we spend on research problems. In numerous moments his brilliant ideas could resolve obstacles I would face during research. I would like to thank the rest of my thesis committee: Professors Rohini Kumar, Alper Murat, George Yin and Sheng Zhang for taking their time to serve on my dissertation committee. It is sad to not see Professor Bert Schreiber anymore among the committee members. He passed away in Summer 2019. He was my mentor in my minor field, Functional Analysis. I learned a lot from him and he would always encourage me to peruse the functional analysis at the research level. My very recent paper, "Varitional analysis in normed spaces", took his advise to the action. I should add that I was fortunate that I had fruitful research discussion with Professor Jonathan Borwein. He directed me to some helpful references during my very early study in Wayne State. Sadly he is not among us anymore, he passed way in August 2016.

A mathematician whose name is constantly repeated throughout my dissertation is Tyrrell Rockafellar. My research is the continuation of the research line in the second-order theory started by Rockafellar. I thank him for promoting the second-order variational analysis and I hope this research line gets more attention from continuous optimization community as it truly deserves.

My sincere thanks also go to the entire Department of Mathematics at Wayne State University for their kind support in a number of ways. Specially, I wish to thank Christopher Leirstein for his invaluable feedback and help with my teaching. I would like thank Professor Alexander Shapiro and Dr. Ying Cui for their support on my research and postdoc position.

I end this section with thanking a man who made me interested in the Optimization subject when I was an undergraduate math student, Dr. Majid Soleimani-damaneh. I truly miss his classes and his challenging star problems.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

Many optimization problems appearing in applications can be formulated in the following format of the composite optimization:

$$\text{minimiz} \ \ \phi(x) + g(F(x)) \quad \text{over all} \ \ x \in \mathbb{X}, \tag{1.1}$$

where $\phi : \mathbb{X} \to \mathbb{R}$ and $F : \mathbb{X} \to \mathbb{Y}$ are continuously twice differentiable around the reference points and $g : \mathbb{Y} \to \overline{\mathbb{R}} := (-\infty, +\infty]$ is a lower semicontinuous (l.s.c.) convex function and where $\mathbb{X}$ and $\mathbb{Y}$ are two finite-dimensional Hilbert spaces. By allowing g to take $+\infty$, we can incorporate constraints into problem (1.1). By letting g different convex function problem (1.1) can cover different class of optimization problems. In the following we mention a sequence of such class of optimization problems. Letting $g(y) := \delta_\Theta(y)$ where δ_Θ is the indicator function of the closed convex set Θ, which takes 0 on the set Θ and ∞ otherwise, problem (1.1) comes out as the (conic) constrained optimization.

$$\text{minimiz} \ \ \phi(x) \quad \text{subject to} \ \ F(x) \in \Theta, \tag{1.2}$$

In particular, (1.1) covers three major classes of constrained optimization problems; 1-Nonlinear programming, 2-Second-order cone programming and 3-Semidefinite cone programming. For example if $\Theta := \mathbb{R}^r_- \times \{0\}^{m-r}$ and $F = (f_1, f_2, ..., f_m)$ problem (1.1) reduces to the nonlinear programming in the conventional format with equality and inequality constraints:

$$\begin{aligned} \text{minimize} \quad & \phi(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, .., r \\ & f_i(x) = 0, \quad i = r+1, ..., m \end{aligned} \tag{1.3}$$

Beyond the constrained optimization (1.2), there are many (nonsmooth) composite optimization that can be formulated into framework (1.1), but not into framework (1.2). For example by letting $g(y) = \max\{y_1, ..., y_m\} + \delta_\Theta(y)$, $F = (f_1, ..., f_m)$ and $\phi = 0$ problem (1.1) reduces to the (constrained) min-max problem:

$$\text{minimize } \max\{f_1, ..., f_m\} \quad \text{subject to } F(x) \in \Theta,$$

By letting

$$g(y) =: \theta_{ZQ}(y) = \sup_{z \in Z}\{\langle z, Qz\rangle - \langle z, y\rangle\}$$

where Q is positive semidefinite matrix and Z is a polyhedral convex set, problem (1.1) reduces to the *extended nonlinear programming* (ENLP). In particular case, if F is affine and ϕ is quadratic functions we get *extended linear-quadratic programming* (ELQP);

$$\text{minimize } \langle c, x\rangle + \frac{1}{2}\langle x, Px\rangle + \theta_{ZQ}(b - Ax) \quad \text{subject to } x \in \mathbb{R}^n \tag{1.4}$$

This kind of model goes back to Rockafellar and Wets [53], where it was introduced for the sake of penalty modeling and algorithm development in stochastic programming. The topic was expanded in [50], where many special cases of ELQP were worked out and applications were made to continuous-time optimal control. Another class of problems which falls into (1.1) framework is minimal norm problems and problems having norm-penalty;

$$\text{minimiz } \phi(x) + \|Ax - b\| \quad \text{subject to } G(x) \in C \tag{1.5}$$

where $\|\|$ can be any norm on $\mathbb{R}^n$. Indeed by setting $F(x) := \big(G(x), Ax - b\big)$ and $g(y, z) = \|z\|$, problem (1.1) reduces to the above optimization problems. Besides penalty techniques in constrained optimization the latter problems covers two important optimization problems: *Least absolute shrinkage and selection operator* (LASSO) which has applications in statistics is formulated through following composite optimization problem:

$$\text{minimize } \alpha\|b - Ax\|_2^2 + \beta\|x\|_1 \quad \text{subject to } x \in \mathbb{R}^n$$

here $g(y) = \beta\|y\|_1$ and $\phi = \alpha\|b - Ax\|_2^2$. The second example is *Support vector machine*

(SVM) where we set $g:\mathbb{R}\to\mathbb{R}$ with $g(y):=\max\{0,y\}$:

$$\text{minimize } \alpha\|x\|_2^2+\frac{1}{m}\sum_{i=1}^{m}\max\{0\ ,\ \omega_i(\langle x,z^i\rangle-\beta)\}\quad \text{s.t }\ (x,\beta)\in\mathbb{R}^{n+1}$$

In last four previous optimization problems the function g shares an interesting geometric feature. Indeed in last four previous examples function g is convex piecewise linear-quadratic. Recall that $g\colon\mathbb{R}^m\to\overline{\mathbb{R}}$ is *piecewise linear-quadratic* (PWLQ) if $\operatorname{dom} g=\cup_{i=1}^{s}C_i$ with C_i being polyhedral convex sets for $i=1,\ldots,s$, and if g has a representation of the form

$$g(x)=\langle x,B_ix\rangle+\langle b_i,x\rangle+\beta_i\quad\text{for all}\quad x\in C_i,$$

where each B_i is symmetric n by n matrix, $b_i\in\mathbb{R}^m$, and $\beta_i\in\mathbb{R}$ for all $i=1,\ldots,s$. This great observation has been done by Rockafellar, [54, Theorem 11.14], who deeply investigated the first- and second-order variational analysis of this class of problems. In this dissertation we take a big step outside of this class of problems and develop first- and second-order varitional analysis for composite optimization problem (1.1). Another important class of optimization problems covered by (1.1) is eigenvalue optimization problem: denote $\lambda_r(X)$ by r^{th} largest eigenvalue of the symmetric matrix $X\in\mathcal{S}^{n\times n}$. Then the optimization problem in the following form

$$\text{minimize } \lambda_1(X)+...+\lambda_r(X)\quad\text{subject to }\ G(X)\in C$$

falls into (1.1) by letting $F(X):=(X,G(X))$, $g(X,Y):=\lambda_1(X)+...+\lambda_r(X)$, and $\Theta:=S^n\times C$ There are many other classes of optimization problems which fall into the composite optimization framework (1.1), thus the analysis of problem (1.1) not only covers a broad class of optimization problems but also it unifies all other (first/second-order) analysis for different class of optimization problems. This dissertation carries out the latter analysis and thus the obtained results are new even in each individual class of optimiztion problems. Although we use variational analysis tools, our approach is different from ones in [33] and [54] which leads us to stronger results in first-order theory and new results in second-order theory.

Looking at the objective function of the optimization problem (1.1), the cumbersome comes from the nonsmooth part, i.e. $g \circ F$. The main aim of this paper this paper is developing first- and second-order theory for the composition $g \circ F$ and then by applying it we aim to obtain optimality conditions and numerical algorithm for solving the composite optimization problems (1.1).

The rest of the dissertation is organized as follows. In Chapter 2 we review the basic generalized differential tools of variational analysis used in formulations and proofs of the main results. We introduce the *metric subregularity constraint qualification* for the problem (1.1). We investigate the relationship between popular constraint qualifications. It turns out that the metric subregularity constraint qualification is (so far) the weakest constraint qualification under which the first-and second-order chain rules hold for the composition $g \circ F$. Indeed in Chapter 2 we establish first-order chain rules via both subderivative and subdifferential under metric subregularity constraint qualification. We define subamenable composition and verify its prox-regularity property. At the end of chapter 2 we apply the obtained first-order calculus to derive first order optimality condition for the composite optimization problem (1.1).

In Chapter 3 we recall the parabolic regularity from [54] which happens to be the key for establishing the second-order chain rule for the strongly subamenable composition $g \circ F$ via second-order subderivatives. In parallel way, we establish the second-order chain rule via parabolic subderivative. This chapter also aims to provide a systematic study of the twice epi-differentiability of extend-real-valued functions in finite dimensional spaces. In particular, we pay special attention to the strongly subamenable compositions. As we mentioned earlier the composite optimization problem (1.1) encompasses major classes of constrained and composite optimization problems including classical nonlinear programming problems, second-order cone and semidefinite programming problems, eigenvalue optimizations problems [57], and fully amenable composite optimization problems [51], see Example 3.20 for more detail. Consequently, the composite problem (1.1) provides a unified framework to study second-order variational properties, including the twice epi-differentiability and second-order optimality conditions, of the aforementioned optimization problems. As argued below, the twice epi-differentiability carries vital second-order information for extend-real-valued functions and therefore plays an important role in modern second-order variational analysis. A lack of an appropriate second-order generalized derivative for nonconvex extended-real-valued functions was the main driving force for Rockafellar to introduce in [48] the concept of the twice epi-differentiability for such functions. Later, in his landmark paper [51], Rockafellar justified this property for an important class of functions, called *fully amenable*, that includes nonlinear programming problems but does not go far enough to cover other major classes of constrained and composite optimization problems. Rockafellar's results were extended in [10, 23] for composite functions appearing in (1.1). However, these extensions were achieved under a restrictive assumption on the second subderivative, which does not hold for constrained optimization problems. Nor does this condition hold for other major composite functions related to eigenvalue optimization problems; see [57, Theorem 1.2] for more detail. Levy in [32] obtained upper and lower estimates for the second subderivative of the composite function from (1.1), but fell short of establishing the twice epi-differentiability for this framework.

The author, Mordukhovich and Sarabi observed recently in [39] that a second-order regularity, called *parabolic regularity* (see Definition 3.3), can play a major role toward the establishment of the twice epi-differentiability for constraint systems, namely when the outer function g in (1.1) is the indicator function of a closed convex set. This vastly alleviated the difficulty that was often appeared in the justification of the twice epi-differentiability for the latter framework and opened the door for crucial applications of this concept in theoretical and numerical aspects of optimization. Among these applications, we can list the following:

- the calculation of proto-derivatives of subgradient mappings via the connection between the second subderivative of a function and the proto-derivative of its subgradient mapping (see equation (4.2));
- the calculation of the second subderivative of the augmented Lagrangian function associated with the composite problem (1.1), which allows us to characterize the second-order growth condition for the augmented Lagrangian problem (cf. [39, Theorems 8.3 & 8.4]);
- the validity of the derivative-coderivative inclusion (cf. [54, Theorem 13.57]), which has important consequences in parametric optimization; see [40, Theorem 5.6] for a recent application in the convergence analysis of the sequential quadratic programming (SQP) method for constrained optimization problems.

Also In this chapter 3, we show that the twice epi-differentiability of the objective function in (1.1) can be guaranteed under parabolic regularity. To achieve this goal, we demand that the outer function g from (1.1) be locally Lipschitz continuous relative to its domain; see the next section for the precise definition of this concept. Shapiro in [55] used a similar condition but in addition assumed that this function is finite-valued. The latter does bring certain restrictions for (1.1) by excluding constrained problems as well as piecewise linear-quadratic composite problems. As shown in Example 3.20, major classes of constrained and composite optimization problems, including ones mentioned in introduction, satisfy this (relative) Lipschitzian condition. However, some composite problems such as the spectral abcissa minimization (cf. [6]), namely the problem of minimizing the largest real parts of eigenvalues, can not be covered by (1.1).

In Chapter 4 the main attention is given to study of the strong metric subregularity of the KKT system of the constrained optimization problem

$$\min_{x\in\mathbb{R}^n} \quad \varphi(x) \quad \text{subject to} \quad F(x)\in\Theta, \tag{1.6}$$

where $\varphi\colon \mathbb{R}^n \to \mathbb{R}$ and $F\colon \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable functions around reference points, where Θ is a closed and convex set in $\mathbb{R}^m$. Following Chapter 3 the second-order analysis of optimization problems requires certain second-order regularity condition on the constraint set Θ. We will assume in this chapter that the set Θ is parabolically regular; see Definition 3.3. Our first goal in this chapter is to provide a systematic study of strong metric subregularity of the KKT system of (4.1) using the second-order analysis conducted for parabolically regular sets in chapter 3.It is well-known [12] that for nonlinear programs the latter property amounts to the uniqueness of Lagrange multipliers as well as the second-order sufficient condition. Later it was observed in [13] that a similar result for $\mathcal{C}^2$-cone reducible constrained problems, namely (4.1) with Θ being $\mathcal{C}^2$-cone reducible in the sense of [4, Definition 3.135], can be established if the uniqueness of Lagrange multipliers is replaced by the strong Robinson constraint qualification (see condition (4.9)). We show that the well-known result for NLPs can be retrieved for parabolically regular constrained problems if we assume further the multiplier mapping is calm. Moreover, our results reveal that the combination of uniqueness of Lagrange multipliers and the calmness of the multiplier mapping amounts to the strong Robinson constraint qualification. These illustrate that the calmness of multiplier mapping, being automatically satisfied for NLPs, is a property that is required in order to achieve a similar result as those in NLPs for the constrained problem (4.1) in general. Such a calmness property was recently used in [42] in order to characterize noncriticality of Lagrange multipliers for generalized KKT systems that encompass the KKT system for the constrained optimization problem (4.1). Our second goal is to provide an important application of the established characterizations of the strong metric subregularity of the KKT system of (4.1) in the *basic sequential programming method* (SQP) for this problem. For the NLPs framework, the sharpest result was achieved by Bonnans [1] in which he showed that the combination of the uniqueness of Lagrange multipliers and the second-order sufficient condition ensures that the basic SQP method can generate a sequence that is convergent and the rate of convergence is superlinear. We will show that Bonnans' result can be extended for the parabolically regular constrained optimization problems if we further assume that the multiplier mapping is calm.

CHAPTER 2 FIRST-ORDER VARIATIONAL ANALYSIS

This chapter is devoted to the first-order variational analysis by paying essential attention to the calculus rules for certain class of nonsmooth functions. The main sources of this chapter are [35, 36, 38].

2.1 Tools of Variational Analysis

In this section we briefly overview some basic constructions of generalized differentiation in variational analysis, which are widely used in what follows. Then we define the metric subregularity qualification condition associated to the composition $f := g \circ F$. We provide some sufficient conditions of this concept, in particular the Robinson Constraint Qualification. In what follows, $\mathbb{X}$ and $\mathbb{Y}$ are finite-dimensional Hilbert spaces equipped with a scalar product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$. By $I\!B$ we denote the closed unit ball in the space in question and by $I\!B_r(x) := x + rI\!B$ the closed ball centered at x with radius $r > 0$. we denote by $\mathbb{R}_+$ (respectively, $\mathbb{R}_-$) the set of non-negative (respectively, non-positive) real numbers. Two important spaces we usually deal with in this thesis are the 1- Euclidean $\mathbb{R}^m$ and 2-the space of symmetric matrices $\mathcal{S}^m$ equipped with the *trace* inner product.

Given a nonempty set Ω in $\mathbb{X}$, in what follows we denote by $\text{dist}(x, \Omega)$ the distance between $x \in \mathbb{X}$ and the set Ω. The notation $\text{co}\,\Omega$ stands for the convex hull of Ω. Furthermore, $x \xrightarrow{\Omega} \bar{x}$ indicates that $x \to \bar{x}$ with $x \in \Omega$. For any set Ω in $\mathbb{X}$, its indicator function is defined by $\delta_\Omega(x) = 0$ for $x \in \Omega$ and $\delta_\Omega(x) = \infty$ otherwise. We write $x(t) = o(t)$ with $x(t) \in \mathbb{X}$ for all $t \in \mathbb{R}_+$ to mean that $\frac{\|x(t)\|}{t}$ goes to 0 as $t \downarrow 0$. Finally, we denote by $\mathbb{R}_+$ (respectively, $\mathbb{R}_-$) the set of non-negative (respectively, non-positive) real numbers. For a function $f : \mathbb{X} \to \mathbb{Y}$, we denote by $\nabla f(\bar{x})\colon \mathbb{X} \to \mathbb{Y}$ the Fréchet derivative (Jacobian) of f at $\bar{x}$, which is a bounded linear operator. If $F : \mathbb{X} \to \mathbb{Y}$ is twice differentiable at $\bar{x}$, the second derivative of F which is actually a linear operator from $\mathbb{X}$ to $L(\mathbb{X}, \mathbb{Y})$, can be equivalently identified by a continuous bilinear map $\nabla^2 F(\bar{x}) : \mathbb{X} \times \mathbb{X} \to \mathbb{Y}$. If F is either continuously twice differentiable around

$\bar{x}$ or $\mathbb{X}$ and $\mathbb{Y}$ are finite-dimensional, then the bilinear map $\nabla^2 F(\bar{x})(.,.)$ is symmetric, i.e, $\nabla^2 F(\bar{x})(u,v) = \nabla^2 F(\bar{x})(v,u)$ for all $u,\ v \in \mathbb{X}$. In the latter case we have

$$F(\bar{x}+h) = F(\bar{x}) + \langle \nabla F(\bar{x}), h\rangle + \frac{1}{2}\nabla^2 F(\bar{x})(h,h) + o(\|h\|^2)$$

see [4, Lemma 2.51] and [54, Theorem 13.2]. For the finite -dimensional case of $\mathbb{X} = \mathbb{R}^n$ and $\mathbb{Y} = \mathbb{R}^m$ if F is twice differentiable then

$$\nabla^2 F(\bar{x})(w,v) = \big(\langle \nabla^2 f_1(\bar{x})w, v\rangle, \ldots, \langle \nabla^2 f_m(\bar{x})w, v\rangle\big) \quad \text{for all } \ v,w \in \mathbb{R}^n,$$

where $F = (f_1, \ldots, f_m)$ and $\nabla^2 f_i(\bar{x})$ stands for the Hessian of f_i at $\bar{x}$. We begin with recalling some of well-known tools of variational analysis that will be utilized throughout of this paper. Given $(\Omega_t)_{t>0}$ a the parameterized family of subsets of $\mathbb{X}$ define the outer and inner limit of Ω_t as $t \downarrow 0$ respectively by

$$\begin{aligned}
\operatorname*{Lim\,sup}_{t\downarrow 0} \Omega_t \quad &:= \quad \{x \in \mathbb{X} \mid \exists\, t_k \downarrow 0,\ \exists\, x_k \to x \text{ with } x_k \in \Omega_{t_k}\},\\
\operatorname*{Lim\,inf}_{t\downarrow 0} \Omega_t \quad &:= \quad \{x \in \mathbb{X} \mid \forall\, t_k \downarrow 0,\ \exists\, x_k \to x,\ \exists N \in I\!N,\ \forall k \geq N \text{ with } x_k \in \Omega_{t_k}\}.
\end{aligned}$$

Following the above definitions, the parameterized family of sets Ω_t is said to be convergent to the (closed) set Ω as $t \downarrow 0$, denoted by $\Omega_t \to \Omega$ or $\operatorname{Lim}_{t\downarrow 0} \Omega_t = \Omega$ if the following equality holds

$$\operatorname*{Lim\,sup}_{t\downarrow 0} \Omega_t = \operatorname*{Lim\,inf}_{t\downarrow 0} \Omega_t = \Omega. \tag{2.1}$$

Given a nonempty set $\Omega \subset \mathbb{X}$ with $\bar{x} \in \Omega$, the *tangent/contingent cone* $T_\Omega(\bar{x})$ to Ω at $\bar{x} \in \Omega$ is defined by

$$T_\Omega(\bar{x}) := \big\{u \in \mathbb{X} \mid \exists\, t_k \downarrow 0, \ \ u_k \to u \quad \text{as } \ k \to \infty \ \text{ with } \ \bar{x} + t_k u_k \in \Omega\big\}.$$

We say that a tangent vector $u \in T_\Omega(\bar{x})$ is *derivable* if there exists $\xi : [0,\varepsilon] \to \Omega$ with $\varepsilon > 0$, $\xi(0) = \bar{x}$, and $\xi'_+(0) = w$, where ξ'_+ signifies the right derivative of ξ at 0 defined by

$$\xi'_+(0) := \lim_{t\downarrow 0} \frac{\xi(t) - \xi(0)}{t}.$$

The set Ω is geometrically derivable at $\bar{x}$ if every tangent vector u to Ω at $\bar{x}$ is derivable. The geometric derivability can be equivalently described as when the outer and inner limits of the parameterized family sets $(\frac{\Omega-\bar{x}}{t})_{t>0}$ agree as $t \downarrow 0$, in another words $\operatorname{Lim}_{t\downarrow 0} \frac{\Omega-\bar{x}}{t} = T_\Omega(\bar{x})$, see [31, Theorem 4.1.24]. Let $\bar{x} \in \Omega$, define the *Dini-Hadamard* normal cone by

$$N_\Omega^-(\bar{x}) := T_\Omega^*(\bar{x}) = \{v \in \mathbb{X} \mid \langle v, u\rangle \leq 0, \text{ for all } u \in T_\Omega(\bar{x})\}$$

The Dini-Hadamard normal cone is convex and closed subset of $\mathbb{X}$. It is well-known that in finite dimensions the Dini-Hadamard normal cone agrees with the well-known Fréchet normal cone denoted by $\widehat{N}$, and defined by

$$\widehat{N}_\Omega(\bar{x}) := \left\{ v \in \mathbb{X} \middle| \limsup_{x \xrightarrow{\Omega} \bar{x}} \frac{\langle v, x-\bar{x}\rangle}{\|x-\bar{x}\|} \leq 0 \right\}$$

Note that in a general infinite-dimensional space we only have the inclusion $\widehat{N}_\Omega(\bar{x}) \subseteq N_\Omega^-(\bar{x})$. If Ω is convex then the both above normal cones reduce to the standard normal cone in convex analysis. We simply use the notation $N_\Omega(\bar{x})$ for the normal cone to the case convex set Ω at $\bar{x}$.

Given the function $f : \mathbb{X} \to \overline{\mathbb{R}} := (-\infty, \infty]$, its domain and epigraph are defined, respectively, by

$$\operatorname{dom} f = \big\{x \in \mathbb{X} |\ f(x) < \infty\big\} \quad \text{and} \quad \operatorname{epi} f = \big\{(x,\alpha) \in \mathbb{X} \times \mathbb{R} |\ f(x) \leq \alpha\big\}.$$

Recall that a set-valued mapping $S\colon \mathbb{X} \rightrightarrows \mathbb{Y}$ is *metrically regular* around $(\bar{x}, \bar{y}) \in S$ if there is $\ell \geq 0$ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that we have the distance estimate

$$\operatorname{dist}\big(x\,;\ S^{-1}(y)\big) \leq \ell \operatorname{dist}\big(y\,;\ F(x)\big) \qquad \text{for all} \quad (x,y) \in U \times V. \tag{2.2}$$

If $y = \bar{y}$ in (2.2), the mapping S is called to be *metrically subregular* at $(\bar{x}, \bar{y})$. We end this section by mentioning a beautiful result by Hoffman [21] in 1952, which shows metric subregularity is automatic in many important situations. For a proof the Hoffman lemma see [25, Theorem 8.33].

Lemma 2.1 (Hoffman lemma) *Let th be a convex polyhedral subset of* $\mathbb{Y}$*. Let* $F : \mathbb{X} \to \mathbb{Y}$ *be a linear mapping. Then the set-valued mapping* $x \to F(x) - \Theta$ *in (uniformly) metrically subregular at all points of* $\mathbb{X} \times \{0\}_{\mathbb{Y}}$*. In another words, there exists* $\kappa > 0$ *such that for all* $x \in X$ *we have:*

$$\operatorname{dist}\big(x\ ;\ F^{-1}(\Theta)\big) \leq \kappa \operatorname{dist}\big(F(x)\ ;\ \Theta\big)$$

2.2 Qualification Conditions

In this section we investigate the qualification conditions associated to the composite function

$$f := g \circ F, \tag{2.3}$$

where g is convex and lower semicontinous (l.s.c) and F is continuously differentiable around the point $\bar{x} \in \mathbb{R}^n$. I the next proposition investigate their relationships. We will see that the generalized Abadie qualification condition serves as the weakest constraint qualification. However, later we will see that the more suitable qualification condition is metric subregularity qualification condition which enables us to remove the bar in (2.6). Another drawback of Abadie qualification is that it is too weak to provide the required metric estimations in analyzing of numerical algorithms.

Proposition 2.2 (relationships between qualification conditions) *In the composition* (3.4)*, let* $F : \mathbb{X} \to \mathbb{Y}$ *be countinuously differentiable around* $\bar{x} \in \mathbb{X}$ *and let* $g : \mathbb{Y} \to \overline{\mathbb{R}}$ *be convex, proper and continuous relative to its domain. Let* $f(\bar{x})$ *be finite, then consider the following conditions:*

(i) *Robinson qualification condition (RQC) holds at* $\bar{x}$*:*

$$0 \in \operatorname{int}\{F(\bar{x}) + \nabla F(\bar{x})\mathbb{X} - \operatorname{dom} g\}.$$

(ii) *The Basic qualification condition (BQC) holds at* $\bar{x}$*:*

$$N_{\operatorname{dom} g} \cap \ker \nabla F(\bar{x})^* = \{0\} \tag{2.4}$$

(iii) *The set-valued mapping* $x \to F(x) - \operatorname{dom} g$ *is metrically regular at* $(\bar{x}, 0)$*.*

(iv) *The set-valued mapping* $x \to F(x) - \operatorname{dom} g$ *is metrically subregular at* $(\bar{x}, 0)$*.*

(v) *The Abadie qualification condition (AQC) holds at* $\bar{x}$ *that is*

$$T_{\operatorname{dom} f}(\bar{x}) = \{u \in \mathbb{X} \mid \nabla F(\bar{x})u \in T_{\operatorname{dom} g}(F(\bar{x}))\} \tag{2.5}$$

(vi) *The limiting Guignard qualification condition holds at* $\bar{x}$*:*

$$N^-_{\operatorname{dom} f}(\bar{x}) = \overline{\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x}))} \tag{2.6}$$

Then we always have (i)$\Longleftrightarrow$(ii)$\Longleftrightarrow$(iii) $\Longrightarrow$ (iv) $\Longrightarrow$ (v) $\Longrightarrow$ (vi)*.*

Proof.

The equivalence (i) $\Longrightarrow$ (ii) can be obtained through convex separation argument in finite dimensions see [4, Proposition 2.97]. For the proof (i) $\Longrightarrow$ (iii) see [4, Proposition 2.89]. The implication (iii) $\Longrightarrow$ (iv) is trivial. To prove the implication (iv) $\Longrightarrow$ (v) assume the set-valued mapping $x \to F(x) - \operatorname{dom} g$ is metrically subregular at $(\bar{x}, 0)$ which means the following inequality holds for all x close enough to $\bar{x}$;

$$\operatorname{dist}(x\ ;\ \operatorname{dom} f) \leq \kappa \operatorname{dist}(F(x)\ ;\ \operatorname{dom} g).$$

Take $u \in T_{\operatorname{dom} f}(\bar{x})$. Therefore, $x(t) := \bar{x} + tu + o(t) \in \operatorname{dom} f$, thus $F(x(t)) \in \operatorname{dom} g$ for all sufficiently small $t > 0$. Since F is continuously differentiable around $\bar{x}$, using the Taylor expansion, we write

$$F(x(t)) = F(\bar{x}) + t\nabla F(\bar{x})u + o(t) \in \operatorname{dom} g,$$

which tells us $\nabla F(\bar{x})u \in T_{\operatorname{dom} g}(F(\bar{x}))$. To prove the other inclusion we use the metric subregularity assumption. Pick $u \in \mathbb{X}$ such that $\nabla F(\bar{x})u \in T_{\operatorname{dom} g}(F(\bar{x}))$. Therefore, we have

$$F(\bar{x}) + t\nabla F(\bar{x})u + o_1(t) \in \operatorname{dom} g.$$

Using the Taylor expansion of F at $\bar{x}$, we can write

$$F(\bar{x} + tu) = F(\bar{x}) + t\nabla F(\bar{x})u + o_2(t).$$

Now by plugging $\bar{x}+tu$ into x in subregularity inequality for sufficiently small $t > 0$ we have:

$$\begin{aligned}
\operatorname{dist}(\bar{x} + tu\ ;\ \operatorname{dom} f) &\leq \kappa \operatorname{dist}(F(\bar{x} + tu)\ ;\ \operatorname{dom} g) \\
&\leq \kappa \,\|(F(\bar{x}) + t\nabla F(\bar{x})u + o_1(t)) - (F(\bar{x}) + t\nabla F(\bar{x})u + o_2(t)) \\
&\leq o(t).
\end{aligned}$$

Therefore, for all $t > 0$ sufficiently small, there exists $x(t) \in \operatorname{dom} f$ such that $\|\bar{x}+tu-x(t)\| \leq o(t)$. This tells us that $x(t)$ can be written in the form $x(t) = \bar{x} + tu + o(t) \in \operatorname{dom} f$, thus $u \in T_{\operatorname{dom} f}(\bar{x})$. To prove the last part of the proposition assume that (v) holds. First, let $v \in N_{\operatorname{dom} g}(F(\bar{x}))$ and $u \in T_{\operatorname{dom} f}(\bar{x})$. From (v) we have $\nabla F(\bar{x})u \in T_{\operatorname{dom} g}(\bar{x})$, this implies that

$$\langle \nabla F(\bar{x})u, v\rangle = \langle u, \nabla F(\bar{x})^* v\rangle \leq 0.$$

Since the above holds for any arbitrary $u \in T_{\operatorname{dom} f}(\bar{x})$, we get that $v \in N^-_{\operatorname{dom} f}(\bar{x})$. This tells us

$$\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x})) \subseteq N^-_{\operatorname{dom} f}(\bar{x}) \quad \Longrightarrow \quad \overline{\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x}))} \subseteq N^-_{\operatorname{dom} f}(\bar{x}).$$

To prove the opposite inclusion, pick $v \in N^-_{\operatorname{dom} f}(\bar{x})$. Then $\langle v, u\rangle \leq 0$ for all $u \in T_{\operatorname{dom} f}(\bar{x})$. Hence, by part (v) we obtain that

$$\langle v, u\rangle \leq 0 \qquad \text{for all } u \in \mathbb{X} \text{ with } \nabla F(\bar{x})u \in T_{\operatorname{dom} g}(F(\bar{x})). \tag{2.7}$$

Now we claim that $v \in \overline{\nabla F(\bar{x})^* N_{\operatorname{dom} g} F(\bar{x})}$. To prove the claim, assume by contrary that $v \notin \overline{\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x}))}$. Since the set $\overline{\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x}))}$ is closed and convex in $\mathbb{X}^*$, by the convex separation theorem, there exist $\xi \in \mathbb{X} \setminus \{0\}$ and $\varepsilon > 0$ such that for all $y \in N_{\operatorname{dom} g}(F(\bar{x}))$ we have

$$\langle \nabla F(\bar{x})\xi, y \rangle = \langle \xi, \nabla F(\bar{x})^* y \rangle \leq \langle \xi, v \rangle - \varepsilon.$$

Since $N_{\operatorname{dom} g}(F(\bar{x}))$ is a cone, we get that $0 \leq \langle \xi, v \rangle - \varepsilon$ and

$$\langle \nabla F(\bar{x})\xi, y \rangle \leq 0 \qquad \text{for all } y \in N_{\operatorname{dom} g}(F(\bar{x})).$$

The above implies that $\nabla f(\bar{x})\xi \in T_{\operatorname{dom} g}(F(\bar{x}))$. Therefore, by (2.7) we get $\langle v, \xi \rangle \leq 0$, which is contradicting with $\varepsilon \leq \langle v, \xi \rangle$. □

Looking at the proof of the Proposition 2.2, In the following example we show that the closure in (2.6) cannot be removed. However, one sufficient condition that allows us to remove the closure in (2.6) is when $\operatorname{dom} g$ is a polyhedral convex set in $\mathbb{Y}$. In the latter case, $\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x}))$ becomes closed, even a polyhedral convex set.

Example 2.3 (Abadie QC is strictly weaker than metric subregularity QC) Let $\Theta \subset \mathbb{R}^m$ be closed convex cone, and let A be a $n \times m$ matrix such that $A\Theta$ is not closed in $\mathbb{R}^m$ (Soon we will show that there exist such a Θ and A). In composition (3.4), let $g = \delta_{\Theta^*}$ and $F(x) = A^* x$, where $\Theta^* := \{y \in \mathbb{R}^m \mid \langle y, c \rangle \leq 0 \ \forall c \in \Theta\}$. Then the followings hold.

(i) The Abadie qualification condition (2.5) holds at $\bar{x} := 0$.

(ii) $\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x}))$ is not closed.

(iii) The set-valued mapping $x \mapsto F(x) - \operatorname{dom} g$ is not metrically subregular at $\bar{x} = 0$.

Note that $\nabla F(\bar{x}) = A^*$, and $T_{\Theta^*}(F(\bar{x})) = T_{\Theta^*}(0) = \Theta^* = \operatorname{dom} g$. To verify (i) we set $f := g \circ F$ then

$$\operatorname{dom} f = \{x \mid A^* x \in \Theta^*\} = \{x \mid \nabla F(\bar{x})x \in T_{\Theta^*}(0)\}$$

Clearly $\operatorname{dom} f$ is a closed, convex cone. Therefore we have

$$T_{\operatorname{dom} f}(\bar{x}) = \operatorname{dom} f = \{x \mid \nabla F(\bar{x})x \in T_{\Theta^*}(0) = T_{\operatorname{dom} g}(F(\bar{x}))\}.$$

This proves that AQC holds at $\bar{x} = 0$ for composition (3.4). To verify that $\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x}))$ is not closed, first note that since Θ is closed convex cone, from standard polarity argument we obtain that $N_{\operatorname{dom} g}(F(\bar{x})) = N_{\Theta^*}(0) = \Theta$. Therefore,

$$\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x})) = AN_{\Theta^*}(0) = A\Theta,$$

which is not closed due to the choice of A and Θ. Turning to ((iii)), it is well known that in finite dimensions under metric subregularity of the mapping $x \mapsto F(x) - \operatorname{dom} g$ at $\bar{x} = 0$, we have $N_{\operatorname{dom} f}(\bar{x}) = \nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x}))$, see [17, Lemma 2.1] or Corollary 2.17 in the present paper. Therefore, under metric subregularity condition we have the closedness of the set $A\Theta = N_{\operatorname{dom} f}(\bar{x})$, which is contradicting with the choice of A and Θ.

In order to show the correctness of the Example 2.3, it remains to construct a closed, convex cone Θ, and a matrix A satisfying the requirements in the Example 2.3. We do it in two stages. First note that there are closed convex cones Θ_1, Θ_2 in $\mathbb{R}^3$ such $\Theta_1 + \Theta_2$ is not closed. Take $\Theta_1 := \{(x, r) \in \mathbb{R}^2 \times \mathbb{R} \mid \|x\| \leq r\}$, the second-order cone (ice cream) in $\mathbb{R}^3$, and $\Theta_2 = \{t(1, 0, -1) \in \mathbb{R}^3 \mid t \geq 0\}$. In this case $\Theta_1 + \Theta_2$ is not closed because it does not contain $(0, 1, 0)$, but it contains

$$\overbrace{(-t, 1 + t^{-1}, t + \frac{t^{-1} + 2t^{-2} + t^{-3}}{2})}^{\in\Theta_1} + \overbrace{(t, 0, -t)}^{\in\Theta_2} = (0, 1 + t^{-1}, \frac{t^{-1} + 2t^{-2} + t^{-3}}{2})$$

for all $t > 0$. Now define the linear transformation $T : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ by $T(x, y) := x + y$. Observe that $\Theta_1 \times \Theta_2$ is closed in $\mathbb{R}^3 \times \mathbb{R}^3$, but $T(\Theta_1 \times \Theta_2) = \Theta_1 + \Theta_2$ is not a closed set. This completes Example 2.3. The following result is immediate from Proposition 2.2.

Corollary 2.4 *In the framework of Proposition 2.2, if* $\nabla F(\bar{x})^* N_{\operatorname{dom} g}(F(\bar{x}))$ *is closed subset* *of $\mathbb{X}$, then the validity of Abadie qualification condition at $\bar{x}$ yields*

$$N^{-}_{\operatorname{dom} f}(\bar{x}) = \nabla F(\bar{x})^{*} N_{\operatorname{dom} g}(F(\bar{x})).$$

Example (2.3) shows that the (generalized) Abadie qualification condition is not a good candidate as a constraint qualification beyond nonlinear programming problems, even in the framework of nonlinear programming, it can not serve as a constraint qualification for second-order necessary optimality condition. In this regard, we use the metric subregularity property as a suitable qualification condition for our analysis in this thesis.

Definition 2.5 (metric subregularity qualification condition) *We say that the composition function* (3.4) *satisfies the metric subregularity qualification condition* (*MSCQ*) *at* $\bar{x} \in \operatorname{dom} f$ *with constant* $\kappa \in \mathbb{R}_+$ *if the set-valued mapping* $x \mapsto F(x) - \operatorname{dom} g$ *is metrically subregular at* $(\bar{x}, 0)$ *with constant* κ.

The introduced MSCQ with constant κ for the composite function (3.4) can be equivalently described as the existence of a neighborhood U of $\bar{x}$ such that the distance estimate

$$\operatorname{dist}(x\,;\ \operatorname{dom} f) \le \kappa \operatorname{dist}\big(F(x)\,;\ \operatorname{dom} g\big) \tag{2.8}$$

The metric subregularity qualification condition for the composition function (3.4) introduced by author in [38] where authors investigated a comprehensive first-order and second-order analysis of problem (1.1) for the case g is a convex piecewise linear-quadratic function. Later this study was generalized to the present framework of this thesis in [37]. As mentioned in Proposition 2.2, while metric regularity qualification can be fully characterized by Basic qualification condition(2.4), finding conditions under which metric subregularity holds is rather challenging. When, however, the outer function ϑ is convex piecewise linear-quadratic that yields the polyhedrality of $\operatorname{dom} g$, Gfrerer [16, Theorem 2.6] achieved a rather simple and verifiable second-order condition to ensure the validity of the metric subregularity (2.8).

His second-order condition in the framework of Definition 2.5 demonstrates that if in addition F is twice differentiable at $\bar{x}$ and if for every $w \in \mathbb{X} \setminus \{0\}$ with $\nabla F(\bar{x})w \in T_{\mathrm{dom}\, g}(F(\bar{x}))$ and every $\lambda \in N_{\mathrm{dom}\, g}\big(F(\bar{x})\big) \cap \ker \nabla F(\bar{x})^*$ the implication

$$\langle \lambda, \nabla^2 F(\bar{x})(w,w) \rangle \geq 0 \implies \lambda = 0 \tag{2.9}$$

holds, then the mapping $x \mapsto F(x) - \mathrm{dom}\, g$ is metrically subregular at $(\bar{x}, 0)$.

The following example provides two cases of the composite functions (3.4) with g being convex piecewise linear-quadratic such that the metric regularity qualification condition fails while the metric subregularity qualification condition is satisfied.

Example 2.6 (failure of metric regularity for composite functions) Consider the class of extended-real-valued functions $g\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ defined by

$$g(y) := \sup_{z \in Z} \big\{ \langle y, z \rangle - \langle Bz, z \rangle \big\}, \tag{2.10}$$

where Z is a polyhedral convex set, and where B is an $m \times m$ positive-semidefinite symmetric matrix. This type of penalty functions was introduced by Rockafellar [47], where he formulated an important class of composite optimization problems under the name of *extended nonlinear programming* (ENLP). It is not hard to see that g is a convex piecewise linear-quadratic function.

(a) Consider the function g from (2.10) with $m = 2$,

$$Z := \mathbb{R}^2, \quad \text{and} \quad B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

and define the constraint mapping $F\colon \mathbb{R}^2 \to \mathbb{R}^2$ by $F(x_1, x_2) := (x_1 - x_2, 0)$. It is easy to check that $\mathrm{dom}\, g = \{(y_1, y_2) |\ y_2 = 0\}$. For $\bar{x} := (0,0) \in \mathbb{R}^2$ we have

$$N_{\mathrm{dom}\, g}\big(F(\bar{x})\big) \cap \ker \nabla F(\bar{x})^* = \big\{ (\lambda_1, \lambda_2) \in \mathbb{R}^2 \big|\ \lambda_1 = 0 \big\},$$

which shows that the metric regularity qualification condition (2.4) fails at $\bar{x}$. On the other hand, MSQC (2.8) holds at $\bar{x}$ since the mapping $x \mapsto F(\bar{x}) - \operatorname{dom} g$ is metrically subregular at $(\bar{x}, 0)$. This follows from the aforementioned Hoffman lemma since F is an affine mapping and $\operatorname{dom} g$ is a polyhedral convex set.

(b) Consider the function g from (2.10) with $m = 3$, $Z := \mathbb{R}^3_+$, and $B := 0 \in \mathbb{R}^{3\times 3}$. Define the constraint mapping $F : \mathbb{R}^3 \to \mathbb{R}^3$ by $F = (f_1, f_2, f_3)$ with $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ and

$$f_1(x) := x_1 - x_2^2, \quad f_2(x) := x_1 - x_3^2, \quad f_3(x) := -x_1 - x_1^2 - x_2^2 - x_3^2.$$

Letting $\bar{x} := (0, 0, 0) \in \mathbb{R}^3$, deduce from $\operatorname{dom} g = \mathbb{R}^3_-$ that

$$N_{\operatorname{dom} g}\big(F(\bar{x})\big) \cap \ker \nabla F(\bar{x})^* = \big\{(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3_+ \big|\ \lambda_1 + \lambda_2 - \lambda_3 = 0\big\}, \tag{2.11}$$

which implies that the metric regularity qualification condition (2.4) fails at $\bar{x}$. For any $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ from (2.11) and any $w = (w_1, w_2, w_3) \in \mathbb{R}^3 \setminus \{0\}$ with $\nabla F(\bar{x})w \in T_{\operatorname{dom} g}(F(\bar{x})) = \mathbb{R}^3_-$ we conclude from the conditions

$$\langle \lambda, \nabla^2 F(\bar{x})(w, w)\rangle = -\lambda_3 w_1^2 - (\lambda_1 + \lambda_3) w_2^2 - (\lambda_2 + \lambda_3) w_3^2 \geq 0$$

that $\lambda = 0$, which confirms that (2.9) is satisfied. This tells us that the constraint mapping $x \mapsto F(x) - \operatorname{dom} g$ is metrically subregular at $(\bar{x}, 0)$.

We close this section by the following lemma taken from [15, Proposition 2.1] which will become handy later in the subdifferential section.

Lemma 2.7 *In the composition function* (3.4)*, let the mapping* $x \to F(x) - \operatorname{dom} g$ *be metrically subregular at* $(\bar{x}, 0)$*. Then the mapping* $u \to \nabla F(\bar{x})u - T_{\operatorname{dom} g}(\bar{x})$ *is metrically subrugular at* $(0, 0)$ *with same constant.*

Proof.

The metric subregularity of the mapping $x \to F(x) - \operatorname{dom} g$ at $(\bar{x}, 0)$ with constant $\kappa > 0$

amounts to the following inequality

$$\operatorname{dist}\big(x\ ;\ \operatorname{dom} f\big) \le \kappa \operatorname{dist}\big(F(x)\ ;\ \operatorname{dom} g\big)$$

for all x sufficiently close to $\bar{x}$. Now by letting $x = \bar{x} + tu$ for all $t > 0$ small enough, we can write $F(x) = F(\bar{x}) + t\nabla F(\bar{x})u + o(t)$, thus we have

$$\begin{aligned}
\operatorname{dist}(\bar{x} + tu\ ;\ \operatorname{dom} f) &\le \kappa \operatorname{dist}(F(\bar{x}) + t\nabla F(\bar{x})u + o(t)\ ;\ \operatorname{dom} g) \\
t\operatorname{dist}\big(u\ ;\ \frac{\operatorname{dom} f - \bar{x}}{t}\big) &\le t\kappa \operatorname{dist}\big(\nabla F(\bar{x})u + \frac{o(t)}{t}\ ;\ \frac{\operatorname{dom} g - F(\bar{x})}{t}\big) \\
\operatorname{dist}\big(u\ ;\ \frac{\operatorname{dom} f - \bar{x}}{t}\big) &\le \kappa \operatorname{dist}\big(\nabla F(\bar{x})u\ ;\ \frac{\operatorname{dom} g - F(\bar{x})}{t}\big) + \frac{o(t)}{t}
\end{aligned} \tag{2.12}$$

Now letting $t \downarrow 0$ (taking liminf) in both sides of above inequality we arrive at

$$\operatorname{dist}\big(u\ ;\ T_{\operatorname{dom} f}(\bar{x})\big) \le \kappa \operatorname{dist}\big(\nabla F(\bar{x})u\ ;\ T_{\operatorname{dom} g}(F(\bar{x}))\big).$$

This finishes the proof. □

2.3 Subderivatives

In this section we recall the subderitives and the concept of epi-differentiability of functions which are the functional counterpart of tangent vectors to sets and their derivability. The main objective of this section is establishing an exact calculus for subderivatives under metric subregularity qualification conditions. We begin with the definitions of the subderivative and epi-differentiability of proper functions, and we investigate some of their properties which will be used later. For a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ and a point $\bar{x}$ with $f(\bar{x})$ finite, the subderivative function $\mathrm{d}f(\bar{x})\colon \mathbb{X} \to [-\infty, \infty]$ is defined by

$$\mathrm{d}f(\bar{x})(\bar{u}) := \liminf_{\substack{t\downarrow 0 \\ u\to\bar{u}}} \frac{f(\bar{x} + tu) - f(\bar{x})}{t}.$$

The above generalized directional derivative is sometimes called lower Dini directional derivative or the Dini-Hadamard directional derivative in literature. However, we keep the name ”subderivative" in order to be compatible with [54, Definition 8.1], where this construction is deeply investigated in finite dimensions.It is not difficult to verify that $T_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x})) = \mathrm{epi}\, \mathrm{d}f(\bar{x})$ and $\mathrm{dom}\, \mathrm{d}f(\bar{x}) \subseteq T_{\mathrm{dom}\, f}(\bar{x})$. The latter inclusion holds as equality when f is Lipschitz continuous around around $\bar{x}$ relative to its domain; in the mathematical language there are $\ell > 0$ and a neighborhood U of $\bar{x}$ such that for all $x,\ y \in \mathrm{dom}\, f \cap U$ we have

$$|f(x) - f(y)| \leq \ell\, \|x - y\|. \tag{2.13}$$

Indicator functions and *piecewise linear-quadratic* functions are examples of relatively Lipschitz continuous functions. Recall that $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is *piecewise linear-quadratic* (PWLQ) if $\mathrm{dom}\, f = \cup_{i=1}^{s} \Omega_i$ with Ω_i being polyhedral convex sets for $i = 1, \ldots, s$, and if f has a representation of the form

$$f(x) = \langle x, B_i x\rangle + \langle b_i, x\rangle + \beta_i \quad \text{for all} \quad x \in \Omega_i, \tag{2.14}$$

where each B_i is symmetric n by n matrix, $b_i \in \mathbb{R}^n$, and $\beta_i \in \mathbb{R}$ for all $i = 1, \ldots, s$.

Lemma 2.8 (domain of subderivatives) *Let $f\colon \mathbb{X} \to \overline{\mathbb{R}}$ be Lipschitz continuous relative to* $\mathrm{dom}\, f$ *around $\bar{x}$. Then* $\mathrm{d}f(\bar{x}) : \mathbb{X} \to \overline{\mathbb{R}}$ *is finite, lower semicontinuous, and homogeneous function with* $\mathrm{dom}\, \mathrm{d}f(\bar{x}) = T_{\mathrm{dom}\, f}(\bar{x})$.

Proof.

homogeneity of $\mathrm{d}f(\bar{x})$ is clear from definition. To prove the statement regarding the domain of subderivative note that we always have $\mathrm{dom}\, \mathrm{d}f(\bar{x}) \subseteq T_{\mathrm{dom}\, f}(\bar{x})$. To prove the other direction, take $u \in T_{\mathrm{dom}\, f}(\bar{x})$, note that from Lipschitzian property we get that for all $t > 0$ sufficiently small and $u' \in \mathbb{X}$ with $\bar{x} + tu' \in \mathrm{dom}\, f$ we have

$$\left|\frac{f(\bar{x} + tu') - f(\bar{x})}{t}\right| \leq \ell \|u'\|,$$

where ℓ is taken from (2.13). Thus we get $|\mathrm{d}f(\bar{x})(u)| \leq \ell \|u\|$ for all $u \in T_{\mathrm{dom}\, f}(\bar{x})$, in particular $\mathrm{d}f(\bar{x})(0) = 0$, this proves that f is finite and $\mathrm{dom}\, \mathrm{d}f(\bar{x}) = T_{\mathrm{dom}\, f}(\bar{x})$. The lower semicontinuity of $\mathrm{d}\phi(\bar{x})$ comes from the fact that $T_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x})) = \mathrm{epi}\, \mathrm{d}f(\bar{x})$. $\square$

The function f is called epi-differentiable at $\bar{x}$ if the parameterized family of sets $\mathrm{epi}\, \Delta_t f(\bar{x})(.)$ converges to $\mathrm{epi}\, \mathrm{d}f(\bar{x})$ as $t \downarrow 0$ in the sense of set-convergence (2.1), where

$$\Delta_t f(\bar{x})(u) := \frac{f(\bar{x} + tu) - f(\bar{x})}{t}.$$

The function f is called properly epi-differentiable at $\bar{x}$, if f is epidifferentiable at $\bar{x}$ and $\mathrm{d}f(\bar{x})(.)$ is a proper function. It is not difficult to observe that $T_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x})) = \mathrm{epi}\, \mathrm{d}f(\bar{x})$, and that the epi-differentiabilty of f at $\bar{x}$ amounts to geometric derivability of $\mathrm{epi}\, f$ at $(\bar{x}, f(\bar{x}))$. Also observe that if f is epi-differentiable at $\bar{x}$, then the set $\mathrm{dom}\, f$ is geometrically derivable at $\bar{x}$. Any convex function is epi-differentiable at all points of its domain. More generally if a function is *Clarke regular* at $\bar{x}$, then it is epi-differentiable at $\bar{x}$; see [54, Theorem 6.26]. Still the epi-differentiability is a much weaker property than Clarke regularity. For example, take the absolute value function in $\mathbb{R}$, it is not difficult to observe that this function is eppi-differentiable at $\bar{x} = 0$ but clearly is not Clarke regular at $\bar{x}$.

The following lemma provides a simple characterization of proper epi-differentiability.

Lemma 2.9 *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ be finite at $\bar{x} \in \mathbb{X}$, and let $\mathrm{d}f(\bar{x})(.)$ be a proper function. Then f is properly epi-differentiable at $\bar{x}$ if and only if for every $u \in \mathbb{X}$ there exists a path $u(.) : [0, \varepsilon] \to \mathbb{X}$ (not necessarily continuous) with the properties $\lim_{t \downarrow 0} u(t) = u(0) = u$ and*

$$\mathrm{d}f(\bar{x})(u) = \lim_{t \downarrow 0} \frac{f(\bar{x} + tu(t)) - f(\bar{x})}{t} \tag{2.15}$$

Proof.

First we assume that f is epi-differentiable at $\bar{x}$. Pick $u \in \mathbb{X}$, if $\mathrm{d}f(\bar{x})(u) = \infty$, set $u(t) = u$ for all $t \geq 0$ and observe that (2.15) holds. If $\mathrm{d}f(\bar{x})(u)$ is a finite number, then $(u, \mathrm{d}f(\bar{x})(u)) \in T_{\mathrm{epi}\, f}(\bar{x}, f(\bar{x}))$. Since f is epi-differentiable at $\bar{x}$, the pair $(u, \mathrm{d}f(\bar{x})(u))$ is a derivable tangent vector. This tells us that there exists a path $\xi : [0, \varepsilon] \to \mathrm{epi}\, f$

with components $\xi(t)=(\xi_1(t),\xi_2(t))$ for all $t\in[0,\varepsilon]$, and properties $\xi(0)=(\bar{x},f(\bar{x}))$, and $\xi'_+(0)=(u,\mathrm{d}f(\bar{x})(u))$. Now setting $u(t):=\frac{\xi_1(t)-\bar{x}}{t}$ for all $t\in(0,\varepsilon)$ and $u(0):=u$, from $\xi(t)\in\operatorname{epi} f$, we get that

$$\frac{f(\bar{x}+tu(t))-f(\bar{x})}{t}\le\frac{\xi_2(t)-f(\bar{x})}{t}.$$

Clearly, we have that $\frac{\xi_2(t)-f(\bar{x})}{t}\to \mathrm{d}f(\bar{x})(u)$ as $t\downarrow 0$, this forces the limit of the left side exists and it is equal to $\mathrm{d}f(\bar{x})(u)$. In order to prove the opposite implication, assume that (2.15) holds. We need to show that every vector in $T_{\operatorname{epi} f}(\bar{x},f(\bar{x}))$ is a derivable tangent. Take $(u,\alpha)\in T_{\operatorname{epi} f}(\bar{x},f(\bar{x}))=\operatorname{epi}\mathrm{d}f(\bar{x})$. Let $u:[0,\varepsilon]\to\mathbb{X}$ be the path taken from (2.15). By setting $\xi\colon[0,\varepsilon]\to\mathbb{X}$ with

$$\xi(t):=(\bar{x}+tu(t),f(\bar{x}+tu(t))+t(\alpha-\mathrm{d}f(\bar{x})(u))),$$

it is easy to check that $\xi(t)\in\operatorname{epi} f$ for all $t\in[0,\varepsilon]$, $\xi(0)=(\bar{x},f(\bar{x}))$, and $\xi'_+(0)=(u,\mathrm{d}f(\bar{x})(u))$. This completes the proof of the Lemma. $\square$

The following Theorem is the main result of this section in which we establish a subderivative chain rule for composition (3.4) under the Abadie qulaification condition. We also establish the epi-differentiability of composition (3.4) under metric subregularity qualification condition.

Theorem 2.10 (subderivative chain rule under metric subregularity) *Let $F:\mathbb{X}\to\mathbb{Y}$ be continuously differentiable around $\bar{x}\in\mathbb{X}$, and let $g:\mathbb{Y}\to\overline{\mathbb{R}}$ be Lipschitz continuous around $F(\bar{x})$ relative to its domain and it is epi-differentiable . Assume the Abadie qualification condition(AQC) in* (2.5) *holds, then the following subderivative chain rule holds:*

$$\mathrm{d}(g\circ F)(\bar{x})(u)=\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u). \tag{2.16}$$

If we replace the AQC with the metric subregularity qualification condition (2.8)*, then* (2.16) *holds and $g\circ F$ is epi-differentiable at $\bar{x}$.*

Proof.

To prove (2.16), pick $u \in \mathbb{X}$ and observe that $\nabla F(\bar{x})u' + \frac{o(t\|u'\|)}{t} \to \nabla F(\bar{x})u$ as $t \downarrow 0$ and $u' \to u$. This yields the relationships

$$\begin{aligned}
\mathrm{d}(g \circ F)(\bar{x})(u) &= \liminf_{\substack{t \downarrow 0 \\ u \to \bar{u}}} \frac{g(F(\bar{x} + tu')) - g(F(\bar{x}))}{t} \\
&= \liminf_{\substack{t \downarrow 0 \\ u \to \bar{u}}} \frac{g(F(\bar{x}) + t\nabla F(\bar{x})u' + o(t\|w'\|)) - g(F(\bar{x}))}{t} \\
&= \liminf_{\substack{t \downarrow 0 \\ u \to \bar{u}}} \frac{g(F(\bar{x}) + t(\nabla F(\bar{x})u' + \frac{o(t\|w'\|)}{t})) - g(F(\bar{x}))}{t} \\
&\geq \mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u).
\end{aligned}$$

Turing to the proof of the opposite inequality, we conclude from the Lemma 2.8 that $\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u) > -\infty$. Moreover, the latter inequality is obvious if $\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u) = \infty$. So assume that $\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u)$ is finite. Note that g is epi-differentiable at $F(\bar{x})$ and it is Lipschitz relative to $\operatorname{dom} g$ around $F(\bar{x})$ these together yield that g is properly epi-differentiable at $F(\bar{x})$. Therefore, by Lemma (2.9) there exists a path $y(.)$ in $\mathbb{Y}$ with $\lim_{t \downarrow 0} y(t) = \nabla F(\bar{x})u$ such that

$$\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u) = \lim_{t \downarrow 0} \frac{g(F(\bar{x}) + ty(t)) - g(F(\bar{x}))}{t}. \tag{2.17}$$

Since $\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u)$ is finite, we can assume without lost of generality that $F(\bar{x}) + ty(t) \in \operatorname{dom} \nu$ for all $t \in [0, \varepsilon]$. Moreover, $\nabla F(\bar{x})u \in \operatorname{dom} \mathrm{d}g(F(\bar{x})) = T_{\operatorname{dom} g}(F(\bar{x}))$. Thus by the assumption AQC (2.5) at $\bar{x}$, we have $u \in T_{\operatorname{dom} f}(\bar{x})$. This also gives us sequences $\{u_k\}$ in $\mathbb{X}$ and $t_k > 0$ converging to u and 0 respectively such that $\bar{x} + t_k u_k \in \operatorname{dom} f$ for all $k \in I\!N$. Using these two along with (2.17), leads us to

$$\begin{aligned}
\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u) &= \lim_{k \to \infty} \Big[\frac{g(F(\bar{x} + t_k u_k)) - g(F(\bar{x}))}{t_k} + \frac{g(F(\bar{x}) + t_k y(t_k)) - g(F(\bar{x} + t_k u_k))}{t_k} \Big] \\
&\geq \liminf_{k \to \infty} \frac{g(F(\bar{x} + t_k u_k)) - g(F(\bar{x}))}{t_k} - \ell \limsup_{k \to \infty} \| \frac{F(\bar{x} + t_k u_k) - F(\bar{x})}{t_k} - y(t_k) \| \\
&\geq \mathrm{d}(g \circ F)(\bar{x})(u) - \ell \limsup_{k \to \infty} \| \nabla F(\bar{x}) u_k + \frac{o(t_k)}{t_k} - y(t_k) \|,
\end{aligned}$$

where ℓ is a Lipschtiz constant of g around $F(\bar{x})$ relative to its domain. Since $y(t_k) \to \nabla F(\bar{x})u$, we get $\limsup_{k\to\infty} \|\nabla F(\bar{x})u_k + \frac{o(t_k)}{t_k} - y(t_k)\| = 0$. This proves the subderivative chain rule (2.16).

Turing to the proof of the last part, we are going to show that f is epi-differentiable at $\bar{x}$. Assume that MSQC (2.8) holds at $\bar{x}$. Since MSQC implies AQC at $\bar{x}$ the subderivative chain rule in (2.16) holds at $\bar{x}$. Moreover, we conclude from Lemma 2.8, that $\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x}).) = \mathrm{d}f(\bar{x})(.)$ is a proper function. To finish the proof by using Lemma 2.9 we only need verify (2.15) for all $u \in \mathbb{X}$. Thus we take $u \in \mathbb{X}$ with $\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u)$ be finite. This yields $\nabla F(\bar{x})u \in T_{\mathrm{dom}\, g}$. Choose the path $y(.)$ in $\mathbb{Y}$ such that (2.17) holds. Now applying MSQC (2.8), with $x := \bar{x} + tu$ for sufficiently small $t > 0$ we have

$$\mathrm{dist}(\bar{x} + tu\ ;\ \mathrm{dom}\, f) \le \kappa\, \mathrm{dist}(F(\bar{x} + tu)\ ;\ \mathrm{dom}\, g),$$

which in turn results in the relationships

$$\begin{aligned} \mathrm{dist}(u\ ;\ \frac{\mathrm{dom}\, f - \bar{x}}{t}) \quad &\le \quad \frac{\kappa}{t}\, \mathrm{dist}(F(\bar{x}) + t\nabla F(\bar{x})u + o(t)\ ;\ \mathrm{dom}\, g) \\ &\le \quad \frac{\kappa}{t}\, \|F(\bar{x}) + t\nabla F(\bar{x})u + o(t) - F(\bar{x}) - ty(t)\| \\ &= \quad \kappa\, \|\nabla F(\bar{x})u - y(t) + \frac{o(t)}{t}\|. \end{aligned}$$

Thus for all $t > 0$ sufficiently small, we find $u(t) \in \frac{\mathrm{dom}\, f - \bar{x}}{t}$ such that

$$\|u - u(t)\| \le \kappa\, \|\nabla F(\bar{x})u - y(t) + \frac{o(t)}{t}\| + t.$$

This tells us that $\bar{x} + tu(t) \in \Omega$ for all $t > 0$ sufficiently small, and that $\lim_{t\downarrow 0} u(t) = u$. Furthermore since g is Lipschitz continuous around $F(\bar{x})$ relative to $\mathrm{dom}\, g$ we have

$$\|\frac{g(F(\bar{x}) + ty(t)) - g(F(\bar{x} + tu(t)))\|}{t} = \ell\, \|\nabla F(\bar{x})u(t) + \frac{o(t)}{t} - y(t)\| \to 0. \quad \text{as } t \downarrow 0.$$

Taking the above into consideration we arrive at

$$\begin{aligned}
\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})u) &= \lim_{t\downarrow 0}\Big[\frac{g(F(\bar{x}+tu(t)))-g(F(\bar{x}))}{t}+\frac{g(F(\bar{x})+ty(t))-g(F(\bar{x}+tu(t)))}{t}\Big]\\
&= \lim_{t\downarrow 0}\frac{g(F(\bar{x}+tu(t)))-g(F(\bar{x}))}{t}=\lim_{t\downarrow 0}\frac{f(\bar{x}+tu(t))-f(\bar{x})}{t}\le \mathrm{d}f(\bar{x})(u)
\end{aligned}$$

All above inequalities hold as equality due to (2.16), in particular this ensures (2.15) holds. Thus the proof of the theorem is complete. □

As it can be seen in Theorem 2.10 the metric subregularity qualification condition, in addition to the chain rule formula (2.16), ensures the epi-differentiability of the composite function (4.35). In the next section we will see that the metric subregularity reveals a stronger regularity property for the composition (3.4), called *prox-regularity*, which is a crucial regularity for the second-order theory.

Corollary 2.11 (subderivative sum rule) *Let $f:\mathbb{X}\to\overline{\mathbb{R}}$ and $h:\mathbb{X}\to\overline{\mathbb{R}}$ be finite at $\bar{x}$ and Lipschitz around $\bar{x}$ relative to their domains. Let the following metric qualification condition hold: there exist $\kappa>0$ and U a neighborhood of $\bar{x}$ such that for all $x\in U$ one has*

$$\operatorname{dist}(x\,;\,\operatorname{dom} f\cap\operatorname{dom} h)\le\ \kappa\left(\operatorname{dist}(x;\operatorname{dom} f)+\operatorname{dist}(x,\operatorname{dom} h)\right)\tag{2.18}$$

Then following subderivative sum rule holds at $\bar{x}$

$$\mathrm{d}(f+h)(\bar{x})(u)=\mathrm{d}f(\bar{x})(u)+\mathrm{d}h(\bar{x})(u)\qquad \textit{for all}\ \ u\in\mathbb{X}$$

Furthermore, if both f and h are epi-differentiable at $\bar{x}$, then $f+h$ is epi-differentiable at $\bar{x}$.

Proof.

Define $g:\mathbb{X}\times\mathbb{X}\to\overline{\mathbb{R}}$ with $g(x,y):=f(x)+h(y)$, and $F:\mathbb{X}\to\mathbb{X}\times\mathbb{X}$, by $F(x):=(x,x)$. It is not difficult to check that the metric qualification condition in (2.18) is reduced to the MSQC (2.8) at $\bar{x}$ for the new composition $g\circ F$. We claim g is epi-differentiable at $(\bar{x},\bar{x})$.

Pick $(u,z)\in\mathbb{X}\times\mathbb{X}$ clearly we have

$$\Delta_t g(\bar{x},\bar{x})(u,z)=\Delta_t f(\bar{x})(u)+\Delta_t h(\bar{x})(z).$$

Let $u(.)$ and $z(.)$ be the two paths satisfying epi-differentibility criteria in Lemma 2.9 for functions f and h respectively. Then we have:

$$\begin{aligned}
\mathrm{d}g(\bar{x},\bar{x})(u,z)\le\liminf_{t\downarrow 0}\Delta_t\psi(\bar{x},\bar{x})(u(t),z(t)) &= \liminf_{t\downarrow 0}\{\Delta_t f(\bar{x})(u(t))+\Delta_t h(\bar{x})(z(t))\}\\
&= \lim_{t\downarrow 0}\Delta_t f(\bar{x})(u(t))+\lim_{t\downarrow 0}\Delta_t h(\bar{x})(z(t))\\
&= \mathrm{d}f(\bar{x})(u)+\mathrm{d}h(\bar{x})(z).
\end{aligned}$$

Actually the inequality $\le$ in above holds as equality because we always have

$$\mathrm{d}f(\bar{x})(u)+\mathrm{d}h(\bar{x})(z)\le\mathrm{d}g(\bar{x},\bar{x})(u,z).$$

This proves that ψ is epi-differentibal at $(\bar{x},\bar{x})$. Now by applying Theorem 2.10 (ii), for all $u\in\mathbb{X}$ we get that

$$\begin{aligned}
\mathrm{d}(f+h)(\bar{x})(u)=\mathrm{d}(\psi\circ F)(\bar{x})(u) &= \mathrm{d}\psi(F(\bar{x}))(\nabla F(\bar{x})u)\\
&= \mathrm{d}\psi(\bar{x},\bar{x})(u,u)=\mathrm{d}f(\bar{x})(u)+\mathrm{d}h(\bar{x})(u).
\end{aligned}$$

This finishes the proof the Corollary. $\square$

Remark 2.12 We proved Theorem 2.10 under epi-differentiablity assumption on g. In many applications g is a convex function which yields its epi-differentiability everywhere on its domain. It is worth mentioning that in the composition structure (3.4) if g is convex and $F(\bar{x})\in\operatorname{ri}\operatorname{dom}g$ then g is Lipschitz continuous around $F(\bar{x})$ relative to its domain. We will see that the convexity of g plays a crucial role to obtaining the exact first and second-order chain roule for the composition (3.4). For this reason and that the convexity of g is not a restriction in applications, for the rest of the thesis we assume g is convex and lower semicontinuous. We will see that the convexity of g together with the metric subregularity qualification (2.8) ensures some useful regularity and the exact first-order chain rule for the composition $f = g \circ F$. Therefore, it is nice if we assign a name for such compositions.

We close this section with following definition.

Definition 2.13 (subamenabl functions) *Let $f : \mathbb{X} \to \mathbb{R}$ be finite at $\bar{x}$. We say f is subamenable at $\bar{x}$ if there exist functions $g : \mathbb{Y} \to \mathbb{R}$ and $F : \mathbb{X} \to \mathbb{Y}$ and an open neighborhood U of $\bar{x}$ such that the following properties hold:*

(i) *$f(x) = g(F(x))$ for all $x \in U$,*

(ii) *F is continuously differentiable on U.*

(iii) *g is convex, l.s.c on U and Lipschitz relative to its domain around $\bar{x}$.*

(iv) *The metric subregularity qualification* (2.8) *holds at $\bar{x}$.*

We say the set Ω is subamenable at $\bar{x}$ if δ_Ω is subamenable function at $\bar{x}$. The functions f is called strongly subamenable at $\bar{x}$ if f is subamenable fucntion and F is twice-continuously differentiable in (ii), furthermore, f is called fully subamenable at $\bar{x}$ if f is strongly subamenable where additionally g is piecewise linear quadratic function.

It is easy to observe that the subamenability is a robust property, in the sense that if f is subamenable at $\bar{x}$ then there exists a neighborhood U of $\bar{x}$ such that f is subamenable at all points in $U \cap \operatorname{dom} f$ with the same Lipschitz and metric subregularity constants. By Theorem 2.10 the subderivative of any subamenable function f can be fully written in terms of it decomposition g and F. In particular, the subderivative function, $\mathrm{d}f(\bar{x}) : \mathbb{X} \to \mathbb{R}$ is convex and lower semicontinuous function. Furthermore it is finite everywhere if g is finite around $\bar{x}$.

The amenable, strongly amenable and fully amenable functions introduced by Rockafellar [54, Definition 10.23], are defined in a similar way, the main difference is qualification condition. Indeed Rockafellar used the basic qualification (2.4) for amenable structures which is strictly stronger than metric subregularity qualification condition. Luckily the composition (3.4) is strongly subamenable in many applications such composite optimization examples we mentioned in the introduction chapter.

2.4 Subdifferential

The main objective of this section is developing subdifferential calculus via Dini-Hadamard subdifferential. The approach we use to develop first-order calculus fundamentally differs from the exteremal principle used in [33] or penalization method in [54]. We assume the reader is familiar with sum and chain rules of convex function via convex subdifferential, denoted by ∂. We basically first establish this calculus under assumption the Lipschitzian, then we relax this assumption to the relatively Lipschitzian. In particular, we establish exact subdifferential chain rule for the composition the subamenable composition functions. We start by recalling the definition of Dini-Hadamard subdifferential from [25, Definition 4.21]. Given the function $f\colon \mathbb{X} \to \overline{\mathbb{R}}$, finite at $\bar{x}$, the Dini-Hadamard subdifferential is defined by

$$\partial^{-} f(\bar{x}) := \{v \in \mathbb{X} \mid \langle v, u\rangle \leq \mathrm{d}f(\bar{x})(u) \text{ for all } u \in X\}.$$

It is not difficult to see that

$$\partial^{-} f(\bar{x}) = \{v \in \mathbb{X} \mid (v, -1) \in N^{-}_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))\}$$

Dini-Hadamard subdifferential agrees with the well known *Fréchet subdifferential* in finite dimensional spaces. It is clear that from definition we have

$$\sup\{\langle v, u\rangle \mid v \in \partial^{-} f(\bar{x})\} \leq \mathrm{d}f(\bar{x})(u)$$

under some regularity conditions, such as Clarke regularity, equality holds. In particular if f is convex then the above inequality holds as equality. It is not diffucult to show that for any set Ω with $\bar{x} \in \Omega$ we have $\partial^{-}\delta_{\Omega}(\bar{x}) = N^{-}_{\Omega}(\bar{x})$. For convex cases the Dini-Hadamard subdifferential and normal cone reduce to thier counterpart in convex analysis. Throughout this dissertation we use the notations ∂ and N_Ω for the convex subdifferential and convex normal cone respectively.

Example 2.14 (subdifferential of distance function) Let Ω be a closed set with $\bar{x} \in \Omega$. Define $f(x) := \operatorname{dist}(x; \Omega)$. Then we have

$$\mathrm{d}f(\bar{x})(u) = \operatorname{dist}(u\ ;\ T_\Omega(\bar{x})) \qquad\qquad \partial^- f(\bar{x}) = N_\Omega^-(\bar{x}) \cap I\!\!B.$$

Proof.

Proof is straightforward; see [54, Example 8.53] □

The following lemma enables us to calculate the Dini-Hadamard subdifferential of subamenable functions through the convex subdifferential.

Lemma 2.15 *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ is subamenable at $\bar{x}$. Then $\mathrm{d}f(\bar{x}) : \mathbb{X} \to \overline{\mathbb{R}}$ is convex and lower semicontinous function and we have:*

$$\partial^- f(\bar{x}) = \partial[\mathrm{d}f(\bar{x})](0)$$

Proof.

As we discussed earlier the convexity of $\mathrm{d}f(\bar{x})$ follows from subamenability of f at $\bar{x}$ and Theorem 2.10. To prove the formula we proceed by definition of Dini-Hadamard subdifferential in terms of normal cone.

$$\begin{aligned} v \in \partial^- f(\bar{x}) \iff (v, -1) \in N^-((\bar{x}, f(\bar{x}))\ ;\ \operatorname{epi} f) \quad &= \quad T^*((\bar{x}, f(\bar{x}))\ ;\ \operatorname{epi} f) \qquad (2.19) \\ &= \quad \big(\operatorname{epi} \mathrm{d}f(\bar{x})\big)^* \\ &= \quad N((0,0)\ ;\ \operatorname{epi} \mathrm{d}f(\bar{x})) \\ &\iff \quad v \in \partial[\mathrm{d}f(\bar{x})](0). \end{aligned}$$

This completes the proof. □

In the following we prove the chain rule subdifferential for subamenable composition when the outer function g is finite around the reference point. Later in Theorem 2.19 we drop the finitness assumption for g.

Theorem 2.16 *Let $f := g \circ F$ be a subamenable composition at $\bar{x}$. Furthure assume that $g : \mathbb{Y} \to \overline{\mathbb{R}}$ is finite around $F(\bar{x})$. Then the following chain rule formula holds:*

$$\partial^-(g \circ F)(\bar{x}) = \nabla F(\bar{x})^* \partial g(F(\bar{x}))$$

Proof.

To prove this chain rule we get help from convex subdifferential chain rule. Indeed we are going to apply convex subdifferential chain rule for composition of the finite convex function $\mathrm{d}g(F(\bar{x}))$ and the linear function $u \to \nabla F(\bar{x})u$. By Lemma 2.15 and Theorem (2.10) we have:

$$\begin{aligned} \partial^-(g \circ F)(\bar{x}) = \partial[\mathrm{d}(g \circ F)(\bar{x})](0) &= \partial[\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})\ .)](0) \\ &= \nabla F(\bar{x})^* \partial[\mathrm{d}g(F(\bar{x})](0) \\ &= \nabla F(\bar{x})^* \partial g(F(\bar{x})) \end{aligned} \tag{2.20}$$

This completes the proof of the Theorem. □

Corollary 2.17 (normal cone chain rule for constraint sets) *Let the set $\Omega := \{x \mid F(x) \in \Theta\}$ contain $\bar{x}$. Assume Θ is closed convex set and F is continuously differentiable around $\bar{x}$. Furthur assume that the metric subregularity constraint qualification holds at $\bar{x}$, in the sense that there exist a $\kappa > 0$ and a neighborhood U of $\bar{x}$ such that for all $x \in U$ we have*

$$\mathrm{dist}(x\ ;\ \Omega) \leq \kappa\ \mathrm{dist}(F(x)\ ;\ \Theta). \tag{2.21}$$

Then following normal cone chain rule holds:

$$N^-_\Omega(\bar{x}) = \nabla F(\bar{x})^* N_\Theta(F(\bar{x}))$$

In fact for each $v \in N^-_\Omega(\bar{x})$ *there exists* $\lambda \in N_\Theta(F(\bar{x}))$ *with* $\|\lambda\| \leq \kappa\, \|v\|$.

Proof.

By defining $f(x) = \text{dist}(x\ ;\ \Omega)$ and $h(x) = \kappa\ \text{dist}(F(x)\ ;\ \Theta)$ then the metric subregularity constraint qualification (2.21) comes out $f(x) \leq h(x)$ for all $x \in U$. This yields $\partial^- f(\bar{x}) \subseteq \partial^- h(\bar{x})$. Therefore by Example 2.14, the latter inclusion comes out as

$$N^-_\Omega(\bar{x}) \cap I\!B \subseteq N_\Theta(F(\bar{x})) \cap \kappa I\!B, \tag{2.22}$$

which implies $N^-_\Omega(\bar{x}) \subseteq N_\Theta(F(\bar{x}))$. The proof of the other inclusion is straightforward and it does not need any constraint qualification. The proof of the last statement follows from (2.22). $\square$

In Theorem 2.16 the assumption that requires g be finite function is restrictive. Indeed, to involve the constraint set in composite optimization we must let g get infinity value. Our goal is to dropping the finiteness assumption on g in Theorem 2.16. In this regard we use following lemma taken from [38].

Lemma 2.18 (extension of Lipschitz continuity) *Let* $f\colon \mathbb{X} \to \overline{\mathbb{R}}$ *be a Lipschitz continuous function around* $\bar{x}$ *relative to its domain with constant* $\ell \in \mathbb{R}_+$. *Then there exist a number* $\varepsilon > 0$ *and a function* $h\colon \mathbb{X} \to \mathbb{R}$, *which agrees with* f *on* $I\!B_\varepsilon(\bar{x}) \cap \text{dom}\, f$ *and which is Lipschitz continuous on* X *with the same constant* ℓ. *If in addition* f *is convex, then the function* h *can be chosen to be convex as well.*

Proof.

The local Lipschitz continuity of f relative to dom f gives us $\varepsilon > 0$ such that f is Lipschitz continuous on the set $\Omega := I\!B_\varepsilon(\bar{x}) \cap \operatorname{dom} f$ with constant ℓ. Considering the function

$$h(x) := \inf_{u \in \Omega} \big\{ f(u) + \ell \|x - u\| \big\}, \quad x \in \mathbb{X},$$

we can easily check (see, e.g., Rockafellar and Wets [54, Exercise 9.12]) that h agrees with f on Ω while being Lipschitz continuous on $\mathbb{X}$ with the same constant ℓ. Furthermore, the convexity of f clearly yields the convexity of the function

$$\psi(x,u) := f(u) + \delta_\Omega(u) + \ell\|x-u\|, \quad (x,u) \in \mathbb{X} \times \mathbb{X},$$

with respect to both variables. Having the representation $h(x) = \inf_{u \in \mathbb{X}} \psi(x,u)$, we deduce directly from the definition that h is convex on $\mathbb{X}$. $\square$

Now we are ready establish the chain rule formula for subamenable compositions.

Theorem 2.19 (chain rule for subamenable composition) *Let $f := g \circ F$ be a subamenable composition at $\bar{x}$. then following chain rule composition holds:*

$$\partial^-(g \circ F)(\bar{x}) = \nabla F(\bar{x})^* \partial g(F(\bar{x}))$$

*Proof.*Recalling definition 2.13, g is Lipschitz continuous around $\bar{x}$ relative to $\operatorname{dom} g$. Therefore, by Lemma 2.18 there exist a convex Lipscitz function h and a neighborhood V of $F(\bar{x})$ such that $g(y) = h(y) + \delta_{\operatorname{dom} g}(y)$ for $y \in V$. Hence, by applying the lemma 2.15, Corollary 2.11 and standard convex subdifferential sum rule we get

$$\begin{aligned} \partial^-(g \circ F)(\bar{x}) = \partial\left[\mathrm{d}(g \circ F)(\bar{x})\right](0) &= \partial[\mathrm{d}(h \circ F)(\bar{x}) + \mathrm{d}(\delta_{\operatorname{dom} g} \circ F)(\bar{x})](0) \qquad (2.23) \\ &= \partial[\mathrm{d}(h \circ F)(\bar{x})](0) + \partial[\mathrm{d}(\delta_{\operatorname{dom} g} \circ F)(\bar{x})](0) \\ &= \partial^-(h \circ F)(\bar{x}) + N^-_{\operatorname{dom} f}(\bar{x}) \\ &= \nabla F(\bar{x})^* \partial h(F(\bar{x})) + \nabla F(\bar{x})^* N^-_{\operatorname{dom} g}(F(\bar{x})) \\ &= \nabla F(\bar{x})^* \partial\big(h + \delta_{\operatorname{dom} g}\big)(F(\bar{x})) = \nabla F(\bar{x})^* \partial g(F(\bar{x})) \end{aligned}$$

This finishes the proof of the Theorem. □

Below is a consequence of the obtained chain rules, where the metric subregularity qualification condition (2.8) is automatically satisfied. These results cannot be deduced from the other known qualification conditions formulated in [24, 25, 33, 54].

Corollary 2.20 (chain rules for convex piecewise linear-quadratic functions) *Let $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ be defined by $f(x) := g(Ax + a)$, where $g\colon \mathbb{R}^m \to \overline{\mathbb{R}}$ is a convex piecewise linear-quadratic function, A is an $m \times n$ matrix, and $a \in \mathbb{R}^n$. Then for any point $x \in \operatorname{dom} f$ we have*

$$\mathrm{d}f(x)(w) = \mathrm{d}g(Ax + a)(Aw) \quad \textit{and} \quad \partial f(x) = A^*\partial g(Ax + a).$$

*Proof.*Since g is a convex piecewise linear-quadratic function, its domain is a polyhedral convex set. Then the Hoffman lemma 2.1, tells us that metric subregularity (2.8) with $f(x) := Ax + a$ holds automatically at any point $x \in \operatorname{dom} f$. The claimed chain rules follow now from Theorems 2.10 and 2.19. □

below we obtain the sum rule for subamenable functions from the chain rule in Theorem 2.19.

Theorem 2.21 (subdifferential sum rule for subamenable functions) *Let f and h be subamenable at $\bar{x}$. Assume the metric qualification* (2.18) *holds at $\bar{x}$. Then $f + h$ is subamenable at $\bar{x}$ and we have:*

$$\partial^-(f + h)(\bar{x}) = \partial^- f(\bar{x}) + \partial^- h(\bar{x})$$

Proof.

Suppose f and h admit subamenable composition representations $f = g_1 \circ F_1$ and $h = g_2 \circ F_2$ respectively. Defining $F := (F_1, F_2)$ and $g := g_1 + g_2$. Using (2.18), it is not difficult to check the metric subregulaity qualification condition (2.8) holds at $\bar{x}$. This makes $g \circ F$

a subamenable composition at $\bar{x}$. Now by applying Theorem 2.19 we get that

$$\begin{aligned} \partial^-(g \circ F)(\bar{x}) &= \nabla F(\bar{x})^* \partial g(F(\bar{x})) && (2.24) \\ &= (\nabla F_1(\bar{x})^*, \nabla F_2(\bar{x})^*) \left[\partial^- g_1(F_1(\bar{x}))\,,\ \partial^-(F_2(\bar{x}))\right]^\top \\ &= \nabla F(\bar{x})^* \partial^- g_1(F_1(\bar{x})) + \nabla F_2(\bar{x})^* \partial^- g_2(F_2(\bar{x})) \\ &= \partial^-(g_1 \circ F_1)(\bar{x}) + \partial^-(g_2 \circ F_2)(\bar{x}) \\ &= \partial^- f(\bar{x}) + \partial^- h(\bar{x}). \end{aligned}$$

This completes the proof. $\square$

The next corollary plays a significant role in deriving subsequent second-order results. It establishes the boundedness (with quantitative estimates) of dual elements under MSQC (2.8). The latter is well known and rather easy to check under metric regularity.

Corollary 2.22 (bounded multipliers) *Let $f := g \circ F$ be a subamenable composition at $\bar{x}$ in which the Lipschitz and metric subregularity constants are $\ell > 0$ and $\kappa > 0$ respectively. Then for every vector $v \in \partial^-(g \circ F)(\bar{x})$ there exists $\lambda \in \partial g(F(\bar{x}))$ such that*

$$v = \nabla F(\bar{x})^* \lambda \quad \textit{with} \quad \|\lambda\| \leq \ell + \kappa\|v\| + \kappa\ell\|\nabla F(\bar{x})\|. \tag{2.25}$$

Proof.

Assume without lost of generality that g is Lipschitz continuous relative to its entire domain. Applying Lemma 2.18 to the convex outer function g in composition (3.4), we find a convex Lipschitz continuous function $h\colon \mathbb{Y} \to \mathbb{R}$ such that $g = h + \delta_{\operatorname{dom} g}$. Pick $v \in \partial^-(g \circ F)(\bar{x})$. It follows from (2.23) that there exist $\lambda_1 \in \partial h(F(\bar{x}))$ and $\lambda_2 \in N_{\operatorname{dom} g}(F(\bar{x}))$ such that $v = \nabla F(\bar{x})^*(\lambda_1 + \lambda_2)$. Since h is Lipschitz continuous with the same constant ℓ due to Lemma 2.18, we have $\|\lambda_1\| \leq \ell$. On the other hand, we can deduce from Corollary 2.17 that the following condition

$$\|\lambda_2\| \leq \kappa \|\nabla F(\bar{x})^* \lambda_2\| = \kappa \|v - \nabla F(\bar{x})^* \lambda_1\|.$$

holds. Setting now $\lambda := \lambda_1 + \lambda_2$ leads us to

$$\lambda \in \partial h\big(F(\bar{x})\big) + N_{\mathrm{dom}\, g}\big(F(\bar{x})\big) = \partial g\big(F(\bar{x})\big).$$

Furthermore, it follows from the above discussion that

$$\|\lambda\| = \|\lambda_1 + \lambda_2\| \leq \|\lambda_1\| + \kappa\|v - \nabla F(\bar{x})^*\lambda_1\| \leq \ell + \kappa\|v\| + \kappa\ell\,\|\nabla F(\bar{x})\|,$$

which readily verifies representation (2.25).

□

In the next result we study the robustness of the Dini-Hadamard subdifferential of subamenable functions. We show that the subdifferential of subamenable functions at a given point agrees with its limiting version, thus it caries information from neiborhood of points. The *limiting Dini-Hadamard subdifferential* of function $f : \mathbb{X} \to \overline{\mathbb{R}}$ at $\bar{x} \in \mathrm{dom}\, f$ is defined by

$$\partial^{L-} f(\bar{x}) := \{v \in X \mid \exists \text{ seq } v_k \to v,\ x_k \xrightarrow{f} \bar{x} \text{ with } v_k \in \partial^- f(x_k)\} \tag{2.26}$$

Proposition 2.23 (robustness property of subamenable functions) *Let f be a subamenable function at $\bar{x}$. Then $\partial^- f(\bar{x}) = \partial^{L-} f(\bar{x})$*

Proof.

It is clear that $\partial^- f(\bar{x})(\bar{x}) \subset \partial^{L-} f(\bar{x})$, to prove the opposite inclusion pick $v \in \partial^{L-} f(\bar{x})$, and find sequences $x_k \xrightarrow{f} \bar{x}$ and $v_k \to v$ such that $v_k \in \partial^- f(\bar{x})(x_k)$. Suppose that f admits the subamenable representation $f = g \circ F$ around $\bar{x}$. Then by Corollary 2.22 for each $k \in I\!N$ sufficiently large, we can find the multiplier $\lambda_k \in \partial g(F(x_k))$ such that $v_k = \nabla F(\bar{x})^*\lambda$ and $\|\lambda_k\| \leq \ell + \kappa\|v_k\| + \kappa\ell\,\|\nabla F(x_k)\|$. Moreover, the sequence $\{v_k\}$ is bounded so is $\{\lambda_k\}$. By passing to a subsequence we may assume $\{\lambda_k\}$ is convergent to a vector λ. Hence,

$v = \nabla F(\bar{x})^*\lambda$. Now we show that $\lambda \in \partial g(F(\bar{x}))$, to this end first note that $\lambda_k \in \partial g(F(x_k))$ for all indexes k. This tells us that for each k we have

$$\langle \lambda_k, y - F(x_k)\rangle \leq g(y) - g(F(x_k)) \qquad \text{for all } y \in \mathbb{Y}.$$

Letting $k \to \infty$, we get

$$\langle \lambda, y - F(\bar{x})\rangle \leq g(y) - g(F(\bar{x})) \qquad \text{for all } y \in \mathbb{Y}$$

thus $\lambda \in \partial g(F(\bar{x}))$. Finally, by applying again the Theorem 2.19, we have

$$v = \nabla F(\bar{x})^*\lambda \in \nabla F(\bar{x})^*\partial g(F(\bar{x})) = \partial^- f(\bar{x}).$$

This completes the proof. □

Remember that (in finite dimensions) Fréchet subdifferential agrees with Dini-Hadamard subdifferential thus the limiting Dini-Hadamard subdifferential is same as the well known *Limiting Subdifferential*, denoted by ∂_M, introduced by Mordukhovich; see [33, Definition 1.18]. Latter is no longer true in infinite dimensions, indeed in any normed space the inclusion $\partial_M f(\bar{x}) \subseteq \partial^{L-} f(\bar{x})$ holds and there are examples where the inclusion is strict.

In the next proposition we show that the strongly subamenable functions are prox-regular at (around) reference points. Recall that a function f is strongly subamenable at $\bar{x}$ if it admits the subamenable composition $f = g \circ F$ around the point $\bar{x}$ additionally F is twice continuously differentiable around $\bar{x}$. Also a function f is called prox-regular at $\bar{x}$ for $\bar{v} \in \partial^- f(\bar{x})$ if there exist $r > 0$ and $\varepsilon > 0$ such that for all $x,\ u \in \operatorname{dom} f \cap I\!B_\varepsilon(\bar{x})$ and $v \in \partial^- f(x) \cap I\!B_\varepsilon(\bar{v})$ with $f(x) \leq f(\bar{x}) + \varepsilon$ we have

$$-\frac{r}{2}\|u - x\|^2 + \langle v, u - x\rangle \leq f(u) - f(x)$$

In many applications the condition $f(x) \leq f(\bar{x}) + \varepsilon$ in above definition holds for free, more generally if we have the *Subdifferential Continuity* property for many functions appearing in application. Recall from [54, Definition 13.28], a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is called subdifferentially continuous at $\bar{x}$ if $v \in \partial^- f(\bar{x})$ and whenever $(x_k, v_k) \to (\bar{x}, v)$ with $v_k \in \partial^- f(x_k)$, one has $f(x_k) \to f(\bar{x})$.

Prox-regularity together with subdifferential continuity is a very important regularity for the second-order analysis. It was first introduced by Poliquin and Rockafellar in [45]. Since then it has played a significant role in second-order variational analysis, from this point of view it is very important to see when this regularity is present. Note that the orginal definition of prox/subdifferential regularity is slightly different from what we defined above. Indeed the orginal definition uses the ∂_M, however, two definitions are equivalent for the class of strongly subamenable function which is studied in this chapter.

Proposition 2.24 (strongly subamenable functions are prox-regular) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ be strongly subamenable at $\bar{x}$. then f is prox-regular at $\bar{x}$ for all $v \in \partial^- f(\bar{x})$.*

Proof.

Since f is strongly subamenbale at $\bar{x}$ then it is strongly subamenable for all $x \in U \cap \text{dom}\, f$, where U is an open neighborhood of $\bar{x}$. Fix $\bar{v} \in \partial^- f(\bar{x})$, and let V be a bounded and open neighborhood of $\bar{v}$. Suppose f admits the strongly subamenable composition $f = g \circ F$ around $\bar{x}$. Then the Corollary 2.22 tells us that there is $M > 0$ such that for all $x \in \Omega \cap U$ and $v \in \partial^- f(x) \cap V$, there exists $\lambda \in \partial g(F(x))$ such that

$$v = \nabla F(x)^* \lambda \quad , \quad \|\lambda\| \leq M. \tag{2.27}$$

Since F is C^2 around $\bar{x}$, we may assume that U is sufficiently small such that on which F can be represented by

$$F(u) - F(x) = \nabla F(x)(u - x) + \eta(u, x),$$

for some mapping $\eta : \mathbb{X}^2 \to \mathbb{Y}$ and constant $M' > 0$ with $\|\eta(u, x)\| \leq M' \; \|u - x\|^2$ for all

$x, u \in U$. Indeed by mean value theorem we can choose

$$\eta(u,x) := \nabla^2 F(c_{xu})\big((u-x),(u-x)\big), \qquad \text{where} \quad c_{xu} \in [x,u].$$

Now, by taking any $x, u \in \operatorname{dom} f \cap U$, and $v \in \partial^- f(x) \cap V$ with corresponding λ selected in (2.27) we have

$$\begin{aligned} g(F(u)) - g(F(x)) &\geq \langle \lambda, F(u) - F(x) \rangle \\ &= \langle \lambda, \nabla F(x)(u-x) \rangle + \langle \lambda, \eta(u,x) \rangle \\ &= \langle v, u-x \rangle + \langle \lambda, \eta(u,x) \rangle \\ &\geq \langle v, u-x \rangle - MM' \, \|u-x\|^2. \end{aligned} \tag{2.28}$$

by setting $r := 2MM'$ we arrive at

$$-\frac{r}{2}\|u-x\|^2 + \langle v, u-x \rangle \leq f(u) - f(x).$$

This finishes the proof of prox-regularity part. To the prove subdifferential continuity take the sequence $\{(x_k, v_k)\}_{k=1}^{\infty}$ in gph $\partial^- f$ such that $(x_k, v_k) \to (\bar{x}, \bar{v}) \in$ gph $\partial^- f$. Assume, without lost of generality, by (2.27) $\lambda_k \in \partial g(F(x_k))$ is a convergent sequence to $\lambda \in \partial g(F(\bar{x}))$ such that $v_k = \nabla F(x_k)^* \lambda_k$. Then by letting $k \to \infty$ we get

$$v_k = \nabla F(x_k)^* \lambda_k \to \nabla F(\bar{x})\lambda \in v_k = \nabla F(\bar{x})^* \partial g(F(\bar{x})) = \partial^- f(\bar{x}).$$

This finishes the proof of the Proposition.

□

One immediate consequence of Proposition 2.24 is that if f is strongly subamenable at $\bar{x}$ then $\partial^- f(\bar{x}) = \partial^p f(\bar{x})$, where ∂^p stands for the *proximal subdifferential*. Recall that $v \in \partial^p f(\bar{x})$ if and only if there exist $r > 0$ such that for all $x \in \mathbb{X}$

$$-\frac{r}{2}\|x-\bar{x}\|^2 + \langle v, x-\bar{x} \rangle \leq f(x) - f(\bar{x}) \tag{2.29}$$

Therefore when one is dealing with strongly subamenable functions (in finite dimensions), the choice of subdifferential does not really matter as all well known subdifferentials coincide in this class of functions; such as limiting, proximal, Clarke and Fréchet. Thus if is strongly subamenable at $\bar{x}$, the we have

$$\partial^p f(\bar{x}) = \partial^- f(\bar{x}) = \partial^{L-} f(\bar{x}) = \partial^c f(\bar{x}).$$

2.5 First-Order Optimality Conditions

In this section we are going to derive the first-order necessary optimality condition for the composite optimization problem (1.1). The main idea is as follow, if $\bar{x} \in X$ is a local optimal solution for the problem

$$\text{minimiz } \; f(x) \quad \text{over all } \; x \in \mathbb{X},$$

then it is not difficult to check that $0 \in \partial^- f(\bar{x})$. Hence if $f := \phi + g \circ F$ for some functions ϕ, g and F, then by subdifferential calculus developed in previous sections we can write ∂^- fully in terms of ϕ, g and F. The constraint qualification conditions in optimization problems are nothing but he qualification conditions ensuring the aforementioned calculus. In this section we consider following fairly general composite optimization problem;

$$\text{minimize } \; \phi(x) + g(F(x)) \quad \text{over all } \; x \in \mathbb{X}, \tag{2.30}$$

where $\phi : \mathbb{X} \to \mathbb{R}$ and $F : \mathbb{X} \to \mathbb{Y}$ are continuously differentiable around $\bar{x} \in \mathbb{X}$ and $g : \mathbb{Y} \to \overline{\mathbb{R}}$ is a lower semicontinuous (l.s.c.) convex function.

Theorem 2.25 (first-order necessary optimalty conditions) *Let $\bar{x}$ be a local solution for the composite problem* (2.30). *Further, assume the metric subregularity constraint qualification holds at $\bar{x}$ which amounts to the existence of $\kappa > 0$ such that* (2.8) *holds. Then the necessary optimality conditions holds as follow:*

$$\text{there exists } \lambda \in \mathbb{Y} \text{ with } \begin{cases} 0 \in \nabla\phi(\bar{x}) + \nabla F(\bar{x})^*\lambda \\ \lambda \in \partial g(F(\bar{x})) \\ \|\lambda\| \le \ell + \kappa\|\nabla\phi(\bar{x})\| + \kappa\ell\|\nabla F(\bar{x})\| \end{cases} \tag{2.31}$$

where ℓ is a relatively Lipschiz constant for g taken from (2.13).

Proof.

Find U an open neighborhood of $\bar{x}$ on which ϕ is continuosly differentiable and $g \circ F$ is subamenable composition. Since $\bar{x}$ is a local optimal solution we have

$$0 \in \partial^-\big(\phi + g \circ F\big)(\bar{x})$$

Hence, by sum rule, Theorem 2.21, we get

$$0 \in \nabla\phi(\bar{x}) + \partial^-(g \circ F)(\bar{x}) \quad \Rightarrow \quad -\nabla\phi(\bar{x}) \in \partial^-(g \circ F)(\bar{x}).$$

By Corollary 2.22 we can find $\lambda \in \partial g(F(\bar{x}))$ such that $-\nabla\phi(\bar{x}) = \nabla F(\bar{x})\lambda_k$ and

$$\|\lambda\| \le \ell + \kappa\|\nabla\phi(\bar{x})\| + \kappa\ell\|\nabla F(\bar{x})\|.$$

□

Now we look at the constraint optimization problems as special case of problem (2.30). We consider following general conic programming problem:

$$\text{minimize } \ \phi(x) \quad \text{over all } \ F(x) \in \Theta, \tag{2.32}$$

where ϕ and F are continuously differentiable around $\bar{x}$ and Θ is a closed convex set. We denote the feasible set of the problem (2.32) by Ω that is $\Omega := \{x \mid F(x) \in \Theta\}$.

Corollary 2.26 (first-order optimality condition of conic programming) *Let $\bar{x}$ be a local solution for the conic programming* (2.32). *Further, assume the metric subregularity* *constraint qualification holds at $\bar{x}$ which amounts to the existence of $\kappa > 0$ such that:*

$$\mathrm{dist}(x\ ;\ \Omega) \leq \kappa\ \mathrm{dist}(F(x)\ ;\ \Theta).$$

Then the necessary optimality conditions holds as follow:

$$\textit{there exists } \lambda \in \mathbb{Y} \textit{ with } \begin{cases} 0 \in \nabla\phi(\bar{x}) + \nabla F(\bar{x})^*\lambda \\ \lambda \in N_\Theta(F(\bar{x})) \\ \|\lambda\| \leq \kappa\|\nabla\phi(\bar{x})\| \end{cases}$$

Proof.

In Theorem 2.25 take $g = \delta_\Theta$. Then clearly g is Lipschitz relative to its domain with constant $\ell = 0$. □

Nonlinear programming is an important special case of conic programming which is formulated as follows:

$$\begin{aligned} \text{minimize} \quad & \phi(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, .., r \\ & f_i(x) = 0, \quad i = r+1, ..., m \end{aligned} \tag{2.33}$$

where each f_i and ϕ is continuouslt differentiabl around $\bar{x}$. Denote the set of feasible solution by Ω. and define $f_i^+(x) := \max\{0, f_i(x)\}$ for each $x \in \mathbb{R}^n$.

Corollary 2.27 (KKT conditions for nonlinear programming) *Let $\bar{x}\mathbb{R}^n$ be a local solution for the nonlinear programming* (2.33)*. Further, assume the metric subregularity constraint qualification holds at $\bar{x}$ which amounts to the existence of $\kappa > 0$ such that:*

$$\mathrm{dist}(x\ ;\ \Omega) \leq \kappa \left(\sum_{i=1}^{r} f_i^+(x) + \sum_{i=r+1}^{m} |f_i(x)| \right) \tag{2.34}$$

Then the necessary optimality conditions holds as follow:

$$\text{there exist scalars } \lambda_1, ..., \lambda_m \in \mathbb{R} \ s.t \begin{cases} 0 = \nabla\phi(\bar{x}) + \sum_{i=1}^{r} \lambda_i \nabla f_i(x) + \sum_{i=r+1}^{m} \lambda_i \nabla f_i(x) \\ \lambda_i \geq 0, \quad \lambda_i f_i(\bar{x}) = 0, \quad i = 1, ..., r \\ \sum_{i=1}^{m} |\lambda_i| \leq \sqrt{m}\kappa \|\nabla\phi(\bar{x})\| \end{cases} \tag{2.35}$$

Proof.

In Corollary 2.26 set $F := (f_1, ..., f_2)$ and $\Theta := \mathbb{R}_+^r \times \{0\}^{m-r}$. Then It is not difficult to check that

$$\operatorname{dist}\big(F(x) \ ; \ \Theta\big) = \sum_{i=1}^{r} f_i^+(x) + \sum_{i=r+1}^{m} |f_i(x)|.$$

Thus the error estimate in (2.34) reduces to metric subregularity constraint qualification. Now by applying the Corollary 2.26 we come up to the KKT conditions. Note that

$$N_\Theta(F(\bar{x})) = \{\lambda = (\lambda_1, ..., \lambda_r, ..., \lambda_m) \in \mathbb{R}^m \mid \lambda_i \geq 0, \quad \lambda_i f_i(\bar{x}) = 0, \ i = 1, ..., r\}$$

and

$$\sum_{i=1}^{m} |\lambda_i| \leq \sqrt{m} \sqrt{\sum_{i=1}^{m} |\lambda_i|} = \sqrt{m} \ \|\lambda\|.$$

This completes the proof. ☐

Note that the bounded multiplier condition is not included in traditional KKT condition, it is in fact a byproduct of metric subregularity constraint qualification. It is worth mentioning that metric subregularity constraint qualification (2.34) is implied by the well-recognized Mangasarian–Fromovitz constraint qualification. Moreover Hoffman lemma 2.1 tells us that the metric subregularity holds for free if each f_i in (2.33) is affine function. It is also possible to write the KKT optimality conditions under the weaker constraint qualification, Abadie constraint qualification, however, in this case the bounded multiplier condition is missing which is crucial from second-order analysis and algorithmic point of view; see [36].

CHAPTER 3 SECOND-ORDER VARIATIONAL ANALYSIS

This chapter addresses the composite optimization problem (1.1), i.e :

$$\text{minimize} \;\; \varphi(x) + g \circ F(x) \;\; \text{subject to} \;\; x \in \mathbb{X}. \tag{3.1}$$

With the focus on the second-order variational analysis on the nonsmooth part $f = g \circ F$. In this chapter we mostly assume the composition $g \circ F$ makes a strongly subamenable composition; see Definition 2.13. We recall the parabolic regularity from [54] which happens to be the key for establishing the second-order chain rule for the strongly subamenable composition $g \circ F$ via second-order subderivatives. In parallel way we establish the second-order chain rule via parabolic subderivative. The chapter also aims to provide a systematic study of the twice epi-differentiability of extend-real-valued functions in finite dimensional spaces. In particular, we pay special attention to the strongly subamenable compositions. As we mentioned in introduction section the composite optimization problem (1.1) encompasses major classes of constrained and composite optimization problems including classical nonlinear programming problems, second-order cone and semidefinite programming problems, eigenvalue optimizations problems [57], and fully amenable composite optimization problems [51], see Example 3.20 for more detail. Consequently, the composite problem (1.1) provides a unified framework to study second-order variational properties, including the twice epi-differentiability and second-order optimality conditions, of the aforementioned optimization problems. As argued below, the twice epi-differentiability carries vital second-order information for extend-real-valued functions and therefore plays an important role in modern second-order variational analysis.

A lack of an appropriate second-order generalized derivative for nonconvex extended-real-valued functions was the main driving force for Rockafellar to introduce in [48] the concept of the twice epi-differentiability for such functions. Later, in his landmark paper [51], Rockafellar justified this property for an important class of functions, called *fully amenable*, that includes nonlinear programming problems but does not go far enough to cover other major classes of constrained and composite optimization problems. Rockafellar's results were extended in [10, 23] for composite functions appearing in (1.1). However, these extensions were achieved under a restrictive assumption on the second subderivative, which does not hold for constrained optimization problems. Nor does this condition hold for other major composite functions related to eigenvalue optimization problems; see [57, Theorem 1.2] for more detail. Levy in [32] obtained upper and lower estimates for the second subderivative of the composite function from (1.1), but fell short of establishing the twice epi-differentiability for this framework.

The author, Sarabi and Mordukhovich observed recently in [39] that a second-order regularity, called *parabolic regularity* (see Definition 3.3), can play a major role toward the establishment of the twice epi-differentiability for constraint systems, namely when the outer function g in (1.1) is the indicator function of a closed convex set. This vastly alleviated the difficulty that was often appeared in the justification of the twice epi-differentiability for the latter framework and opened the door for crucial applications of this concept in theoretical and numerical aspects of optimization. Among these applications, we can list the following:

- the calculation of proto-derivatives of subgradient mappings via the connection between the second subderivative of a function and the proto-derivative of its subgradient mapping (see equation (4.2));
- the calculation of the second subderivative of the augmented Lagrangian function associated with the composite problem (1.1), which allows us to characterize the second-order growth condition for the augmented Lagrangian problem (cf. [39, Theorems 8.3 & 8.4]);
- the validity of the derivative-coderivative inclusion (cf. [54, Theorem 13.57]), which has

important consequences in parametric optimization; see [40, Theorem 5.6] for a recent application in the convergence analysis of the sequential quadratic programming (SQP) method for constrained optimization problems.

In this chapter, we show that the twice epi-differentiability of the objective function in (1.1) can be guaranteed under parabolic regularity. To achieve this goal, we demand that the outer function g from (1.1) be locally Lipschitz continuous relative to its domain; see the next section for the precise definition of this concept. Shapiro in [55] used a similar condition but in addition assumed that this function is finite-valued. The latter does bring certain restrictions for (1.1) by excluding constrained problems as well as piecewise linear-quadratic composite problems. As shown in Example 3.20, major classes of constrained and composite optimization problems , including ones mentioned in introduction, satisfy this (relative) Lipschitzian condition. However, some composite problems such as the spectral abcissa minimization (cf. [6]), namely the problem of minimizing the largest real parts of eigenvalues, can not be covered by (1.1). The key sources of this chapter are [37, 39].

3.1 Second-Order Tools in Variational Analysis

In this section we first briefly review basic constructions of variational analysis and generalized differentiation employed in this chapter; see [33,54] for more detail. Given a nonempty set $\Omega \subset \mathbb{X}$ with $\bar{x} \in \Omega$, recall the tangent cone $T_\Omega(\bar{x})$ to Ω at $\bar{x}$ is defined by

$$T_\Omega(\bar{x}) = \big\{ w \in \mathbb{X} |\ \exists\, t_k \downarrow 0,\ \ w_k \to w \ \ \text{as}\ \ k \to \infty\ \ \text{with}\ \ \bar{x} + t_k w_k \in \Omega \big\}.$$

We say a tangent vector $w \in T_\Omega(\bar{x})$ is *derivable* if there exist a constant $\varepsilon > 0$ and an arc $\xi : [0, \varepsilon] \to \Omega$ such that $\xi(0) = \bar{x}$ and $\xi'_+(0) = w$, where ξ'_+ signifies the right derivative of ξ at 0, defined by

$$\xi'_+(0) := \lim_{t \downarrow 0} \frac{\xi(t) - \xi(0)}{t}.$$

The set Ω is called geometrically derivable at $\bar{x}$ if every tangent vector w to Ω at $\bar{x}$ is derivable. The geometric derivability of Ω at $\bar{x}$ can be equivalently described by the sets $[\Omega - \bar{x}]/t$ converging to $T_\Omega(\bar{x})$ as $t \downarrow 0$. Convex sets are important examples of geometrically derivable sets. The second-order tangent set to Ω at $\bar{x}$ for a tangent vector $w \in T_\Omega(\bar{x})$ is given by

$$T^2_\Omega(\bar{x}, w) = \big\{u \in \mathbb{X} |\, \exists\, t_k \downarrow 0, \;\; u_k \to u \quad \text{as} \;\; k \to \infty \;\; \text{with} \;\; \bar{x} + t_k w + \frac{1}{2} t_k^2 u_k \in \Omega\big\}.$$

A set Ω is said to be parabolically derivable at $\bar{x}$ for w if $T^2_\Omega(\bar{x}, w)$ is nonempty and for each $u \in T^2_\Omega(\bar{x}, w)$ there are $\varepsilon > 0$ and an are $\xi : [0, \varepsilon] \to \Omega$ with $\xi(0) = \bar{x}$, $\xi'_+(0) = w$, and $\xi''_+(0) = u$, where

$$\xi''_+(0) := \lim_{t \downarrow 0} \frac{\xi(t) - \xi(0) - t\xi'_+(0)}{\frac{1}{2}t^2}.$$

It is well known that if $\Omega \subset \mathbb{X}$ is convex and parabolically derivable at $\bar{x}$ for w, then the second-order tangent set $T^2_\Omega(\bar{x}, w)$ is a nonempty convex set in $\mathbb{X}$ (cf. [4, page 163]).

Recall form chapter one that $v \in \mathbb{X}$ is a proximal subgradient of f at $\bar{x}$ if there exists $r \in \mathbb{R}_+$ and a neighborhood U of $\bar{x}$ such that for all $x \in U$ we have

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle - \frac{r}{2} \|x - \bar{x}\|^2. \tag{3.2}$$

The set of all such v is called the proximal subdifferential of f at $\bar{x}$ and is denoted by $\partial^p f(\bar{x})$. By definitions, it is not hard to obtain the inclusions $\partial^p f(\bar{x}) \subseteq \partial^- f(\bar{x})$ and equality holds if f is strongly subamenable at $\bar{x}$; see proposition 2.24.

Consider a set-valued mapping $S : \mathbb{X} \rightrightarrows \mathbb{Y}$ with its domain and graph defined, respectively, by

$$\operatorname{dom} S = \big\{x \in \mathbb{X} |\; S(x) \neq \emptyset\big\} \quad \text{and} \quad \operatorname{gph} S = \big\{(x, y) \in \mathbb{X} \times \mathbb{Y} |\; y \in S(x)\big\}.$$

The graphical derivative of S at $(\bar{x},\bar{y}) \in \text{gph}\, S$ is defined by

$$DS(\bar{x},\bar{y})(w) = \big\{v \in \mathbb{Y} |\ (w,v) \in T_{\text{gph}S}(\bar{x},\bar{y})\big\}, \quad w \in \mathbb{X}.$$

The set-valued mapping S is called strongly metrically subregular at $(\bar{x},\bar{y})$ if there are a constant $\kappa \in \mathbb{R}_+$ and a neighborhood U of $\bar{x}$ such that the estimate

$$\|x - \bar{x}\| \le \kappa\, d(\bar{y}, S(x)) \ \text{ for all } \ x \in U$$

holds. It is known (cf. [11, Theorem 4E.1]) that the set-valued mapping S is strongly metrically subregular at $(\bar{x},\bar{y})$ if and only if we have

$$0 \in DS(\bar{x},\bar{y})(w) \implies w = 0. \tag{3.3}$$

Given a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ and a point $\bar{x}$ with $f(\bar{x})$ finite, define the parametric family of second-order difference quotients for f at $\bar{x}$ for $\bar{v} \in \mathbb{X}$ by

$$\Delta_t^2 f(\bar{x},\bar{v})(w) = \frac{f(\bar{x}+tw) - f(\bar{x}) - t\langle \bar{v},\, w\rangle}{\frac{1}{2}t^2} \qquad \text{with } \ w \in \mathbb{X}, \ \ t > 0.$$

If $f(\bar{x})$ is finite, then the second subderivative of f at $\bar{x}$ for $\bar{v}$ is given by

$$\mathrm{d}^2 f(\bar{x},\bar{v})(w) = \liminf_{\substack{t\downarrow 0 \\ w' \to w}} \Delta_t^2 f(\bar{x},\bar{v})(w'), \ \ w \in \mathbb{X}.$$

Below, we collect some important properties of the second subderivative that are used throughout this paper. The proof is elementary and straightforward; parts (i) and (ii) were taken from [54, Proposition 13.5] and part (iii) was recently observed in [38, Theorem 4.1(i)].

Proposition 3.1 (properties of second subderivative) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ and $(\bar{x},\bar{v}) \in \mathbb{X} \times \mathbb{X}$ with $f(\bar{x})$ finite. Then the following conditions hold:*

(i) *the second subderivative $\mathrm{d}^2 f(\bar{x},\bar{v})$ is a lower semicontinuous (l.s.c.) function;*

(ii) *if $\mathrm{d}^2 f(\bar{x},\bar{v})$ is a proper function, meaning that $\mathrm{d}^2 f(\bar{x},\bar{v})(w) > -\infty$ for all $w \in \mathbb{X}$ and* *its effective domain, defined by*

$$\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) = \big\{ w \in \mathbb{X} |\, \mathrm{d}^2 f(\bar{x}, \bar{v})(w) < \infty \big\},$$

is nonempty, then we always have the inclusion

$$\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) \subset \big\{ w \in \mathbb{X} |\, \mathrm{d} f(\bar{x})(w) = \langle \bar{v}, w \rangle \big\};$$

(iii) *if $\bar{v} \in \partial^p f(\bar{x})$, then for any $w \in \mathbb{X}$ we have $\mathrm{d}^2 f(\bar{x}, \bar{v})(w) \geq -r\|w\|^2$, where $r \in \mathbb{R}_+$ is taken from* (3.2). *In particular, $\mathrm{d}^2 f(\bar{x}, \bar{v})$ is a proper function.*

Following [54, Definition 13.6], a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is said to be twice epi-differentiable at $\bar{x}$ for $\bar{v} \in \mathbb{X}$, with $f(\bar{x})$ finite, if the sets $\operatorname{epi} \Delta_t^2 f(\bar{x}, \bar{v})$ converge to $\operatorname{epi} \mathrm{d}^2 f(\bar{x}, \bar{v})$ as $t \downarrow 0$. The latter means by [54, Proposition 7.2] that for every sequence $t_k \downarrow 0$ and every $w \in \mathbb{X}$, there exists a sequence $w_k \to w$ such that

$$\mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \lim_{k \to \infty} \Delta_{t_k}^2 f(\bar{x}, \bar{v})(w_k). \tag{3.4}$$

We say f is *properly epi-differetiable* at $\bar{x} \in \operatorname{dom} f$ for $\bar{v}$ if f is epi-differetiable at $\bar{x}$ for $\bar{v}$ and the mapping $w \to \mathrm{d}^2 f(\bar{x}, \bar{v})(w)$ is proper. Since we are mostly interested to proper epi-differetiability we provide following characterization of it whose proof is straightforward and goes same idea in the proof of Lemma 2.9.

Lemma 3.2 (charecterization of properly epi-diffferentiable functions) *Let f be finite at $\bar{x}$ and $w \to \mathrm{d}^2 f(\bar{x}, \bar{v})(w)$ is a proper function. Then f is properly epi-differentiable at $\bar{x}$ for $\bar{v}$ if and only if for each $w \in \mathbb{X}$ there exists a path $w(.) : [0, 1] \to \mathbb{X}$, not necessarily continuous, such $w_t := w(t) \to w$ as $t \downarrow 0$ and*

$$\mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \lim_{t \downarrow 0} \frac{f(\bar{x} + t w_t) - f(\bar{x}) - t\langle \bar{v}, w_t \rangle}{\frac{1}{2} t^2}.$$

It is easy to observe that if f is twice differentiable in classical sense at $\bar{x}$, [54, Definition a] then f is both twice epi-differentiable at $\bar{x}$ (for $\nabla f(\bar{x})$) and for all $w \in \mathbb{R}^n$ one has

$$\mathrm{d}^2 f(\bar{x}, \nabla f(\bar{x}))(w) = \langle w, \nabla^2 f(\bar{x}) w \rangle.$$

Unlike the first order subderivative $\mathrm{d}f(\bar{x})$ the second-order subderivative $\mathrm{d}^2 f(\bar{x}, \bar{v})$ does not have a geometric tangent set counterpart via the known tangent sets so far. This is one of the main reasons why getting the calculus rules for $\mathrm{d}^2 f(\bar{x}, \bar{v})$ is a hard task and it requires pure analytic approach rather than geometric approach. In this regard the *parabolic subderivative*, introduced by Ben-Tal and Zowe in [2], fills this gap. For $f : \mathbb{X} \to \overline{\mathbb{R}}$, a point $\bar{x}$ with $f(\bar{x})$ finite, and a vector w with $\mathrm{d}f(\bar{x})(w)$ finite, the parabolic subderivative of f at $\bar{x}$ for w with respect to z is defined by

$$\mathrm{d}^2 f(\bar{x})(w \mid z) := \liminf_{\substack{t \downarrow 0 \\ z' \to z}} \frac{f(\bar{x} + tw + \frac{1}{2}t^2 z') - f(\bar{x}) - t\mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t^2}.$$

Recall from [54, Definition 13.59] that f is called *parabolically epi-differentiable* at $\bar{x}$ for w if

$$\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \mid \cdot) = \big\{ z \in \mathbb{X} \,|\, \mathrm{d}^2 f(\bar{x})(w \mid z) < \infty \big\} \neq \emptyset,$$

and for every $z \in \mathbb{X}$ and every sequence $t_k \downarrow 0$ there exists a sequences $z_k \to z$ such that

$$\mathrm{d}^2 f(\bar{x})(w \mid z) = \lim_{k \to \infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2}. \tag{3.5}$$

Parabolic subderivative has a nice geometric interpretation in terms of second-order tangent set. Indeed by [54, Example 13.62] we have

$$\operatorname{epi} \mathrm{d}^2 f(\bar{x})(w \mid .) = T^2_{\operatorname{epi} f}\big((\bar{x}, f(\bar{x})), (w, \mathrm{d}f(\bar{x})(w)\big)$$

and parabolic epi-differentiablity f at $\bar{x}$ for w coincides with parabolic derivability $\operatorname{epi} f$ at $(\bar{x}, f(\bar{x}))$ for $(w, \mathrm{d}f(\bar{x})(w))$. Thus parabolic epi-differentiability guarantees the existence of a vector z such that $\mathrm{d}^2 f(\bar{x})(w \mid z) < \infty$, this however does not rule out the possibility that $\mathrm{d}^2 f(\bar{x})(w \mid .) \equiv -\infty$. It is easy to observe that for any function f we have

$\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \,|\, .) \subseteq T^2_{\operatorname{dom} f}(\bar{x}, w)$ where equality holds under some conditions. We will clarify this issue in proposition 3.12.

The main interest in parabolic subderivatives in this paper lies in its nontrivial connection with second subderivatives. Indeed, it was shown in [54, Proposition 13.64] that if the function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is finite at $\bar{x}$, then for any pair $(\bar{v}, w) \in \mathbb{X} \times \mathbb{X}$ with $\mathrm{d}f(\bar{x})(w) = \langle w, \bar{v} \rangle$ we always have

$$\mathrm{d}^2 f(\bar{x}, \bar{v})(w) \leq \inf_{z \in \mathbb{X}} \big\{ \mathrm{d}^2 f(\bar{x})(w \,|\, z) - \langle z, \bar{v} \rangle \big\}. \tag{3.6}$$

As it will be observed in the next section, equality in this estimate amounts to a very important regularity in second-order variational analysis.

3.2 Twice Epi-Differetiability of Parabolically Regular Functions

This section aims to delineate conditions under which the twice epi-differenibility of extend-real-valued functions can be established. To this end, we appeal to an important second-order regularity condition, called parabolic regularity, which was recently exploited in [39] to study a similar property for constraint systems. We begin with the definition of this regularity condition.

Definition 3.3 (parabolic regularity) *A function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is parabolically regular at $\bar{x}$ for $\bar{v} \in \mathbb{X}$ if $f(\bar{x})$ is finite and if for any w such that $\mathrm{d}^2 f(\bar{x}, \bar{v})(w) < \infty$, there exist, among the sequences $t_k \downarrow 0$ and $w_k \to w$ with $\Delta^2_{t_k} f(\bar{x}, \bar{v})(w_k) \to \mathrm{d}^2 f(\bar{x}, \bar{v})(w)$, those with the additional property that*

$$\limsup_{k \to \infty} \frac{\|w_k - w\|}{t_k} < \infty. \tag{3.7}$$

A nonempty set $\Theta \subset \mathbb{X}$ is said to be parabolically regular at $\bar{x}$ for $\bar{v}$ if the indicator function δ_Θ is parabolically regular at $\bar{x}$ for $\bar{v}$.

Although the notion of parabolic regularity was introduced first in [54, Definition 13.65], its origin goes back to [8, Theorem 4.4], where Chaney observed a duality relationship between his second-order generalized derivative and the parabolic subderivative, defined in [2] by Ben-Tal and Zowe. This duality relationship was derived later by Rockafellar [49, Proposition 3.5] for convex piecewise linear-quadratic functions. As shown in Proposition 3.8 below, the latter duality relationship is equivalent to the concept of parabolic regularity from Definition 3.3 provided that $\bar{v}$, appearing in Definition 3.3, is a proximal subgradient. A different second-order regularity was introduced by Bonnans, Comminetti, and Shapiro [5, Definition 3] for sets, which was later extended in [4, Definition 3.93] for functions. It is not difficult to see that parabolic regularity is implied by the second-order regularity in the sense of [5]; see [4, Proposition 3.103] for a proof of this result. Moreover, the example from [4, page 215] shows that the converse implication may not hold in general.

We showed in [39] that important sets appearing in constrained optimization problems, including polyhedral convex sets, the second-order cone, and the cone of positive semidefinite symmetric matrices, are parabolically regular. Below, we add two important classes of functions for which this property automatically fulfill. We begin first by convex piecewise-linear quadratic functions and then consider eigenvalues functions. While the former was justified in [54, Theorem 13.67], we provide below a different and simpler proof.

Example 3.4 (piecewise linear-quadratic functions) Assume that the function $f : \mathbb{X} \to \overline{\mathbb{R}}$ with $\mathbb{X} = \mathbb{R}^n$ is convex piecewise linear-quadratic. Recall that f is called piecewise linear-quadratic if $\operatorname{dom} f = \cup_{i=1}^{s} C_i$ with $s \in I\!N$ and C_i being polyhedral convex sets for $i = 1, \ldots, s$, and if f has a representation of the form

$$f(x) = \langle A_i x, x\rangle + \langle a_i, x\rangle + \alpha_i \quad \text{for all} \quad x \in C_i,$$

where A_i is an $n \times n$ symmetric matrix, $a_i \in \mathbb{R}^n$, and $\alpha_i \in \mathbb{R}$ for $i = 1, \cdots, s$. It was proven in [54, Propsoition 13.9] that the second subderivative of f at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ can be calculated by

$$\mathrm{d}^2 f(\bar{x},\bar{v})(w) = \begin{cases} \langle A_i w, w\rangle & \text{if } \; w \in T_{C_i}(\bar{x}) \cap \{\bar{v}_i\}^{\perp}, \\ \\ \infty & \text{otherwise,} \end{cases} \tag{3.8}$$

where $\bar{v}_i := \bar{v} - A_i\bar{x} - a_i$. To prove the parabolic regularity of f at $\bar{x}$ for $\bar{v}$, pick a vector $w \in \mathbb{R}^n$ with $\mathrm{d}^2 f(\bar{x},\bar{v})(w) < \infty$. This implies that there is an i with $1 \le i \le s$ such that $w \in T_{C_i}(\bar{x}) \cap \{\bar{v}_i\}^{\perp}$. Since C_i is a polyhedral convex set, we conclude from [54, Exercise 6.47] that there exists an $\varepsilon > 0$ such that $\bar{x} + tw \in C_i$ for all $t \in [0,\varepsilon]$. Pick a sequence $t_k \downarrow 0$ such that $t_k \in [0,\varepsilon]$ and let $w_k := w$ for all $k \in I\!N$. Thus a simple calculation tells us that

$$\begin{aligned} \Delta^2_{t_k} f(\bar{x},\bar{v})(w_k) &= \frac{f(\bar{x}+t_k w_k) - f(\bar{x}) - t_k\langle \bar{v},\, w_k\rangle}{\frac{1}{2}t_k^2} \\ &= \frac{\frac{1}{2}\langle A_i(\bar{x}+t_k w_k), \bar{x}+t_k w_k\rangle + \langle a_i, \bar{x}+t_k w_k\rangle + \alpha_i - \frac{1}{2}\langle A_i\bar{x},\bar{x}\rangle - \langle a_i,\bar{x}\rangle - \alpha_i - t_k\langle \bar{v},\, w_k\rangle}{\frac{1}{2}t_k^2} \\ &= \langle A_i w, w\rangle + \frac{t_k\langle w_k, \bar{v} - A_i\bar{x} - a_i\rangle}{\frac{1}{2}t_k^2} = \langle A_i w, w\rangle, \end{aligned}$$

which in turn implies by (3.8) that $\Delta^2_{t_k} f(\bar{x},\bar{v})(w_k) \to \mathrm{d}^2 f(\bar{x},\bar{v})(w)$ as $k \to \infty$. Since (3.7) is clearly holds, f is parabolic regular at $\bar{x}$ for $\bar{v}$.

Example 3.5 (eigenvalue functions) Let $\mathbb{X} = \mathcal{S}^n$ be the space of $n \times n$ symmetric real matrices, which is conveniently treated via the inner product

$$\langle A, B\rangle := \operatorname{tr} AB$$

with $\operatorname{tr} AB$ standing for the sum of the diagonal entries of AB. For a matrix $A \in \mathcal{S}^n$, we denote by $A^{\dagger}$ the Moore-Penrose pseudo-inverse of A and by $\lambda(A) = (\lambda_1(A),\ldots,\lambda_n(A))$ the vector of eigenvalues of A in nonincreasing order with eigenvalues repeated according to their multiplicity. Given $i \in \{1,\ldots,n\}$, denote by $\ell_i(A)$ the number of eigenvalues that are equal to $\lambda_i(A)$ but are ranked before i including $\lambda_i(A)$. This integer allows us to locate $\lambda_i(A)$ in the group of the eigenvalues of A as follows:

$$\lambda_1(A) \geq \cdots \geq \lambda_{i-\ell_i(A)} > \lambda_{i-\ell_i(A)+1}(A) = \cdots = \lambda_i(A) \geq \cdots \geq \lambda_n(A).$$

The eigenvalue $\lambda_{i-\ell_i(A)+1}(A)$, ranking first in the group of eigenvalues equal to $\lambda_i(A)$, is called the *leading* eigenvalue. For any $i \in \{1, \ldots, n\}$, define now the function $\alpha_i : \mathcal{S}^n \to \mathbb{R}$ by

$$\alpha_i(A) = \lambda_{i-\ell_i(A)+1}(A) + \cdots + \lambda_i(A), \quad A \in \mathcal{S}^n. \tag{3.9}$$

It was proven in [57, Theorem 2.1] that $\widehat{\partial}\alpha_i(A) = \partial\alpha_i(A)$ and that the second subderivative of α_i at A for any $V \in \partial\alpha_i(A)$ is calculated for every $W \in \mathcal{S}^n$ by

$$\mathrm{d}^2\alpha_i(A, V)(W) = \begin{cases} 2\langle V, W(\lambda_i(A)I_n - A)^{\dagger}W\rangle & \text{if } \mathrm{d}\alpha_i(A)(W) = \langle X, W\rangle, \\ \infty & \text{otherwise}, \end{cases} \tag{3.10}$$

where I_n stands for the $n \times n$ identity matrix. Moreover, for any $W \in \mathcal{S}^n$ with $\mathrm{d}^2\alpha_i(A, H)(W) < \infty$ and any sequence $t_k \downarrow 0$, the proof of [57, Theorem 2.1] confirms that

$$\Delta^2_{t_k}\alpha_i(A, V)(W_k) \to \mathrm{d}^2\alpha_i(A, V)(W) \quad \text{with } W_k := W - t_k W(\lambda_i(A)I_n - A)^{\dagger}W.$$

This readily verifies (3.7) and thus the functions α_i, $i \in \{1, \ldots, n\}$, are parabolically regular at A for any $V \in \partial\alpha_i(A)$. In particular, for $i = 1$, the function α_i from (3.9) boils down to the maximum eigenvalue function of a matrix, namely

$$\lambda_{\max}(A) := \alpha_1(A) = \lambda_1(A), \quad A \in \mathcal{S}^n. \tag{3.11}$$

So the maximum eigenvalue function $\lambda_{\max}$ is parabolically regular at A for any $V \in \partial\lambda_{\max}(A)$. This can be said for any leading eigenvalue $\lambda_{i-\ell_i(A)+1}(A)$ since we have $\alpha_i(B) = \lambda_{i-\ell_i(A)+1}(B)$ for every matrix $B \in \mathcal{S}^n$ sufficiently close to A. Another important function related to the eigenvalues of a matrix $A \in \mathcal{S}^n$ is the sum of the first i components of $\lambda(A)$ with

$i \in \{1, \ldots, n\}$, namely

$$\sigma_i(A) = \lambda_1(A) + \cdots + \lambda_i(A). \tag{3.12}$$

It is well-known that the functions σ_i are convex (cf. [54, Exercise 2.54]). Moreover, we have $\sigma_i(A) = \alpha_i(A) + \sigma_{i-\ell_i(A)}(A)$. It follows from [57, Proposition 1.3] that $\sigma_{i-\ell_i(A)}$ is twice continuously differentiable ($\mathcal{C}^2$-smooth) on $\mathcal{S}^n$. This together with the parabolic regularity of α_i ensures that σ_i are parabolically regular at A for any $V \in \partial\sigma_i(A)$.

To proceed, let $f : \mathbb{X} \to \overline{\mathbb{R}}$ and pick $(\bar{x}, \bar{v}) \in \operatorname{gph} \partial f$. The *critical cone* of f at $(\bar{x}, \bar{v})$ is defined by

$$K_f(\bar{x}, \bar{v}) := \big\{ w \in \mathbb{X} \,|\, \mathrm{d}f(\bar{x})(w) = \langle \bar{v}, w \rangle \big\}. \tag{3.13}$$

When f is the indicator function of a set, this definition boils down to the classical definition of the critical cone for sets; see [11, page 109]. It is not difficult to see that the set $K_f(\bar{x}, \bar{v})$ is a cone in $\mathbb{X}$. Taking into account Proposition 3.1(ii), we conclude that the domain of the second subderivative $\mathrm{d}^2 f(\bar{x}, \bar{v})$ is always included in the critical cone $K_f(\bar{x}, \bar{v})$ provided that $\mathrm{d}^2 f(\bar{x}, \bar{v})$ is a proper function. The following result provides conditions under which the domain of the second subderivative is the entire critical cone.

Proposition 3.6 (domain of second subderivatives) *Assume that $f : \mathbb{X} \to \overline{\mathbb{R}}$ is finite at $\bar{x}$ with $\bar{v} \in \partial^p f(\bar{x})$ and that for every $w \in K_f(\bar{x}, \bar{v})$ we have* $\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \,|\, \cdot) \neq \emptyset$*. Then for all $w \in K_f(\bar{x}, \bar{v})$ we have*

$$-\, r\|w\|^2 \le \mathrm{d}^2 f(\bar{x}, \bar{v})(w) \le \inf_{z \in \mathbb{X}} \big\{ \mathrm{d}^2 f(\bar{x})(w \,|\, z) - \langle z, \bar{v} \rangle \big\} < \infty, \tag{3.14}$$

where $r \in \mathbb{R}_+$ is a constant satisfying (3.2)*. In particular, we have* $\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) = K_f(\bar{x}, \bar{v})$.

Proof.

The lower estimate of $\mathrm{d}^2 f(\bar{x}, \bar{v})$ in (3.14) results from Proposition 3.1(iii), which readily implies that $\mathrm{d}^2 f(\bar{x},\bar{v})(0)=0$. This tells us that the second subderivative $\mathrm{d}^2 f(\bar{x},\bar{v})$ is proper. Employing now Proposition 3.1(ii) gives us the inclusion $\mathrm{dom}\, \mathrm{d}^2 f(\bar{x},\bar{v}) \subset K_f(\bar{x},\bar{v})$. The upper estimate of $\mathrm{d}^2 f(\bar{x},\bar{v})(w)$ in (3.14) directly comes from (3.6). By assumptions, for any $w \in K_f(\bar{x},\bar{v})$, there exists a z_w so that $\mathrm{d}^2 f(\bar{x})(w \mid z_w) < \infty$. This guarantees that the infimum term in (3.14) is finite. Pick $w \in K_f(\bar{x},\bar{v})$ and observe from (3.14) that $\mathrm{d}^2 f(\bar{x},\bar{v})(w)$ is finite. This yields the inclusion $K_f(\bar{x},\bar{v}) \subset \mathrm{dom}\, \mathrm{d}^2 f(\bar{x},\bar{v})$, which completes the proof. $\square$

The following example, taken from [54, page 636], shows the domain of the second subderivative can be the entire set $K_f(\bar{x},\bar{v})$ even if the assumption on the domain of the parabolic subderivative in Proposition 3.6 fails. As shown in the next section, however, this condition is automatically satisfied for composite functions appearing in (1.1).

Example 3.7 (domain of second subderivative) Define the function $f:\mathbb{X}\to\overline{\mathbb{R}}$ with $\mathbb{X}=\mathbb{R}^2$ by $f(x_1,x_2)=|x_2-x_1^{4/3}|-x_1^2$. As argued in [54, page 636], the subderivative and subdifferential of f at $\bar{x}=(0,0)$, respectively, are

$$\mathrm{d}f(\bar{x})(w)=|w_2| \quad \text{and} \quad \partial f(\bar{x})=\big\{v=(v_1,v_2)\in\mathbb{R}^2 \,|\, v_1=0,\ |v_2|\le 1\big\},$$

where $w=(w_1,w_2)\in\mathbb{R}^2$. It is not hard to see that $\bar{v}=(0,0)\in\partial^p f(\bar{x})$. Moreover, the second subderivative of f at $\bar{x}$ for $\bar{v}$ has a representation of the form

$$\mathrm{d}^2 f(\bar{x},\bar{v})(w)=\begin{cases} -2w_1^2 & \text{if } w_2=0,\\ \infty & \text{if } w_2\neq 0.\end{cases}$$

Using the above calculation tells us that $K_f(\bar{x},\bar{v})=\{w=(w_1,w_2)\,|\, w_2=0\}$. Thus we have $\mathrm{dom}\, \mathrm{d}^2 f(\bar{x},\bar{v})=K_f(\bar{x},\bar{v})$. However, for any $w=(w_1,w_2)\in K_f(\bar{x},\bar{v})$ with $w_1\neq 0$ we have

$$\mathrm{d}^2 f(\bar{x})(w \mid z)=\infty \quad \text{for all } \ z\in\mathbb{R}^2,$$

which confirms that the assumption related to the domain of the parabolic subderivative in Proposition 3.6 fails.

We proceed next by providing an important characterization of the parabolic regularity that plays a key role in our developments in this paper.

Proposition 3.8 (characterization of parabolic regularity) *Assume that* $f : \mathbb{X} \to \overline{\mathbb{R}}$ *is finite at* $\bar{x}$ *with* $\bar{v} \in \partial^p f(\bar{x})$. *Then the function* f *is parabolically regular at* $\bar{x}$ *for* $\bar{v}$ *if and only if we have*

$$\mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \inf_{z \in \mathbb{X}} \left\{ \mathrm{d}^2 f(\bar{x})(w \,|\, z) - \langle z, \bar{v} \rangle \right\} \tag{3.15}$$

for all $w \in K_f(\bar{x}, \bar{v})$. *Furthermore, for any* $w \in \operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v})$, *there exists a* $\bar{z} \in \operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \,|\, \cdot)$ *such that*

$$\mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \mathrm{d}^2 f(\bar{x})(w \,|\, \bar{z}) - \langle \bar{z}, \bar{v} \rangle. \tag{3.16}$$

Proof.

It follows from $\bar{v} \in \partial^p f(\bar{x})$ and Proposition 3.1(ii)-(iii) that the second subderivative $\mathrm{d}^2 f(\bar{x}, \bar{v})$ is a proper function and

$$\operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v}) \subset K_f(\bar{x}, \bar{v}). \tag{3.17}$$

Assume now that f is parabolically regular at $\bar{x}$ for $\bar{v}$. If there exists a $w \in K_f(\bar{x}, \bar{v}) \setminus \operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v})$, then (3.15) clearly holds due to (3.6). Suppose now that $w \in \operatorname{dom} \mathrm{d}^2 f(\bar{x}, \bar{v})$. By Definition 3.3, there are sequences $t_k \downarrow 0$ and $w_k \to w$ for which we have

$$\Delta^2_{t_k} f(\bar{x}, \bar{v})(w_k) \to \mathrm{d}^2 f(\bar{x}, \bar{v})(w) \quad \text{and} \quad \limsup_{k \to \infty} \frac{\|w_k - w\|}{t_k} < \infty.$$

Since the sequence $z_k := 2[w_k - w]/t_k$ is bounded, we can assume by passing to a subsequence if necessary that $z_k \to \bar{z}$ as $k \to \infty$ for some $\bar{z} \in \mathbb{X}$. Thus we have $w_k = w + \frac{1}{2}t_k z_k$ and

$$\begin{aligned} d^2 f(\bar{x}, \bar{v})(w) &= \lim_{k\to\infty} \frac{f(\bar{x} + t_k w_k) - f(\bar{x}) - t_k\langle \bar{v}, w_k\rangle}{\frac{1}{2}t_k^2} \\ &= \lim_{k\to\infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k\langle \bar{v}, w\rangle}{\frac{1}{2}t_k^2} - \langle \bar{v}, z_k\rangle \\ &\geq \liminf_{k\to\infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k df(\bar{x})(w)}{\frac{1}{2}t_k^2} - \langle \bar{v}, \bar{z}\rangle \\ &\geq d^2 f(\bar{x})(w \mid \bar{z}) - \langle \bar{v}, \bar{z}\rangle. \end{aligned}$$

Combining this and (3.6) implies that (3.15) and (3.16) hold for all $w \in \operatorname{dom} d^2 f(\bar{x}, \bar{v})$. To obtain the opposite implication, assume that (3.15) holds for all $w \in K_f(\bar{x}, \bar{v})$. To prove the parabolic regularity of f at $\bar{x}$ for $\bar{v}$, let $d^2 f(\bar{x}, \bar{v})(w) < \infty$, which by (3.17) yields $w \in K_f(\bar{x}, \bar{v})$. Employing now [54, Proposition 13.64] results in

$$d^2 f(\bar{x}, \bar{v})(w) = \inf_{z\in\mathbb{X}} \left\{ d^2 f(\bar{x})(w \mid z) - \langle z, \bar{v}\rangle \right\} = \liminf_{\substack{t\downarrow 0,\, w'\to w \\ [w'-w]/t \text{ bounded}}} \Delta_t^2 f(\bar{x}, \bar{v})(w').$$

The last equality clearly justifies (3.7), and thus f is parabolically regular at $\bar{x}$ for $\bar{v}$. This completes the proof. □

We next show that the indicator function of the cone of $n \times n$ positive semidefinite symmetric matrices, denoted by $\mathcal{S}_+^n$, is parabolic regular. This can be achieved via [39, Theorem 6.2] using the theory of $\mathcal{C}^2$-cone reducible sets but below we give an independent proof via Proposition 3.8.

Example 3.9 (parabolic regularity of $\mathcal{S}_+^n$) Let $\mathcal{S}_-^n$ stand for the cone of $n \times n$ negative semidefinite symmetric matrices. For any $A \in \mathcal{S}_-^n$, we are going to show that $f := \delta_{\mathcal{S}_-^n}$ is parabolic regular at A for any $V \in N_{\mathcal{S}_-^n}(A)$. Since we have $\mathcal{S}_+^n = -\mathcal{S}_-^n$, this clearly yields the same property for $\mathcal{S}_+^n$. Using the notation in Example 3.5, we can equivalently write

$$\mathcal{S}_-^n = \left\{ A \in \mathcal{S}^n \,|\, \lambda_1(A) \leq 0 \right\}, \tag{3.18}$$

which in turn implies that $\delta_{\mathcal{S}_-^n}(A) = \delta_{\mathbb{R}_-}(\lambda_1(A))$ for any $A \in \mathcal{S}^n$. If A is negative definite, i.e.,

$\lambda_1(A) < 0$, then our claim immediately follows from $N_{\mathcal{S}^n_-}(A) = \{0\}$ for this case. Otherwise, we have $\lambda_1(A) = 0$. Pick $V \in N_{\mathcal{S}^n_-}(A)$ and conclude from (3.18) and the chain rule from convex analysis that $N_{\mathcal{S}^n_-}(A) = \mathbb{R}_+\partial\lambda_1(A)$, which implies that $V = rB$ for some $r \in \mathbb{R}_+$ and $B \in \partial\lambda_1(A)$. If $r = 0$, we get $V = 0$ and parabolic regularity of $\delta_{\mathcal{S}^n_-}$ at A for V follows directly from the definition. Assume now $r > 0$ and pick $W \in K_f(A, V)$. The latter amounts to

$$\langle V, W\rangle = \mathrm{d}\delta_{\mathcal{S}^n_-}(A)(W) = 0 \quad \text{and} \quad W \in T_{\mathcal{S}^n_-}(A).$$

Employing now [4, Proposition 2.61] tells us that $\mathrm{d}\lambda_1(A)(W) \leq 0$. Since $B \in \partial\lambda_1(A)$ and $\langle B, W\rangle = 0$, we arrive at $\mathrm{d}\lambda_1(A)(W) = 0$. We know from [54, Example 10.28] that

$$\mathrm{d}\lambda_1(A)(W) = \lim_{\substack{t\downarrow 0\\ W'\to W}} \Delta_t\lambda_1(A)(W') \quad \text{with} \quad \Delta_t\lambda_1(A)(W') := \frac{\lambda_1(A + tW') - \lambda_1(A)}{t}.$$

Using direct calculations, we conclude for any $t > 0$ and $W' \in \mathcal{S}^n$ that

$$\Delta_t^2\delta_{\mathcal{S}^n_-}(A, V)(W') = \Delta_t^2\delta_{\mathbb{R}_-}(\lambda_1(A), r)(\Delta_t\lambda_1(A)(W')) + r\Delta_t^2\lambda_1(A, B)(W'),$$

which in turn results in

$$\mathrm{d}^2\delta_{\mathcal{S}^n_-}(A, V)(W) \geq \mathrm{d}^2\delta_{\mathbb{R}_-}(\lambda_1(A), r)(\mathrm{d}\lambda_1(A)(W)) + r\mathrm{d}^2\lambda_1(A, B)(W).$$

Since $r > 0$, $\lambda_1(A) = 0$, and $\mathrm{d}\lambda_1(A)(W) = 0$, we conclude from [39, Example 3.4] that

$$\mathrm{d}^2\delta_{\mathbb{R}_-}(\lambda_1(A), r)(\mathrm{d}\lambda_1(A)(W)) = \delta_{K_{\mathbb{R}_-}(\lambda_1(A),r)}(0) = \delta_{\{0\}}(0) = 0.$$

Using this together with (3.10) brings us to

$$\mathrm{d}^2\delta_{\mathcal{S}^n_-}(A, V)(W) \geq -2r\langle B, WA^\dagger W\rangle = -2\langle V, WA^\dagger W\rangle.$$

On the other hand, we conclude from (3.14) that

$$\mathrm{d}^2\delta_{\mathcal{S}^n_-}(A, V)(W) \leq -\sigma_{T^2_{\mathcal{S}^n_-}(A,W)}(V) = -2\langle V, WA^\dagger W\rangle,$$

where the last equality comes from [4, page 487] with $\sigma_{T^2_{\mathcal{S}^n_-}(A,W)}$ standing for the support function of $T^2_{\mathcal{S}^n_-}(A,W)$. Combining these confirms that

$$\mathrm{d}^2\delta_{\mathcal{S}^n_-}(A,V)(W) = -\sigma_{T^2_{\mathcal{S}^n_-}(A,W)}(V) = -2\langle V, WA^\dagger W\rangle \quad \text{for all } \ W \in K_f(A,V).$$

This together with Proposition 3.8 tells us that $\mathcal{S}^n_-$ is parabolic regular at A for V.

We are now in a position to establish the main result of this section, which states that parabolically regular functions are always twice epi-differentiable.

Theorem 3.10 (twice epi-differenitability of parabolically regular functions) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ be finite at $\bar{x}$ and $\bar{v} \in \partial^p f(\bar{x})$ and let f be parabolically epi-differentiable at $\bar{x}$ for every $w \in K_f(\bar{x},\bar{v})$. If f is parabolically regular at $\bar{x}$ for $\bar{v}$, then it is properly twice epi-differentiable at $\bar{x}$ for $\bar{v}$ with*

$$\mathrm{d}^2 f(\bar{x},\bar{v})(w) = \begin{cases} \min_{z\in\mathbb{X}} \left\{ \mathrm{d}^2 f(\bar{x})(w \,|\, z) - \langle z, \bar{v}\rangle \right\} & \textit{if} \quad w \in K_f(\bar{x},\bar{v}), \\ +\infty & \textit{otherwise.} \end{cases} \tag{3.19}$$

Proof.

It follows from the parabolic epi-differentiability of f at $\bar{x}$ for every $w \in K_f(\bar{x},\bar{v})$ and Proposition 3.6 that $\operatorname{dom} \mathrm{d}^2 f(\bar{x},\bar{v}) = K_f(\bar{x},\bar{v})$. This together with (3.15) and (3.16) justifies the second subderivative formula (3.19). To establish the twice epi-differentiability of f at $\bar{x}$ for $\bar{v}$, we are going to show that (3.4) holds for all $w \in \mathbb{X}$. Pick $w \in K_f(\bar{x},\bar{v})$ and an arbitrary sequence $t_k \downarrow 0$. Since f is parabolically regular at $\bar{x}$ for $\bar{v}$, by Proposition 3.8, we find a $\bar{z} \in \mathbb{X}$ such that

$$\mathrm{d}^2 f(\bar{x},\bar{v})(w) = \mathrm{d}^2 f(\bar{x})(w \,|\, \bar{z}) - \langle \bar{z}, \bar{v}\rangle. \tag{3.20}$$

By the parabolic epi-differentiability of f at $\bar{x}$ for w, we find a sequence $z_k \to \bar{z}$ for which we have

$$\mathrm{d}^2 f(\bar{x})(w \,|\, \bar{z}) = \lim_{k\to\infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2}.$$

Define $w_k := w + \frac{1}{2}t_k z_k$ for all $k \in I\!N$. Using this and $w \in K_f(\bar{x}, \bar{v})$, we obtain

$$\begin{aligned}\Delta^2_{t_k} f(\bar{x}, \bar{v})(w_k) &= \frac{f(\bar{x} + t_k w_k) - f(\bar{x}) - t_k\langle \bar{v}, w_k\rangle}{\frac{1}{2}t_k^2}\\ &= \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2} - \langle \bar{v}, z_k\rangle.\end{aligned}$$

This together with (3.8) results in

$$\lim_{k\to\infty} \Delta^2_{t_k} f(\bar{x}, \bar{v})(w_k) = \mathrm{d}^2 f(\bar{x})(w \mid \bar{z}) - \langle \bar{v}, \bar{z}\rangle = \mathrm{d}^2 f(\bar{x}, \bar{v})(w),$$

which justifies (3.4) for every $w \in K_f(\bar{x}, \bar{v})$. Finally, we are going to show the validity of (3.4) for every $w \notin K_f(\bar{x}, \bar{v})$. For any such a w, we conclude from (3.19) that $\mathrm{d}^2 f(\bar{x}, \bar{v})(w) = \infty$. Pick an arbitrary sequence $t_k \downarrow 0$ and set $w_k := w$ for all $k \in I\!N$. Thus we have

$$\infty = \mathrm{d}^2 f(\bar{x}, \bar{v})(w) \le \liminf_{k\to\infty} \Delta^2_{t_k} f(\bar{x}, \bar{v})(w_k) \le \limsup_{t\downarrow 0} \Delta^2_{t_k} f(\bar{x}, \bar{v})(w_k) \le \infty = \mathrm{d}^2 f(\bar{x}, \bar{v})(w),$$

which again proves (3.4) for all $w \notin K_f(\bar{x}, \bar{v})$. This completes the proof of the Theorem. $\square$

The above theorem provides a very important generalization of a similar result obtained recently by the authors and Mordukhovich in [39, Theorem 3.6] in which the twice epi-differentiability of set indicator functions was established. It is not hard to see that the assumptions of Theorem 3.10 boils down to those in [39, Theorem 3.6]. To the best of our knowledge, the only results related to the twice epi-differentiability of functions, beyond set indicator functions, are [54, Theorem 13.14] and [57, Theorem 3.1] in which this property was proven for the fully amenable and eigenvalue functions, respectively. We will derive these results in Section 3.4 as an immediate consequence of our chain rule for the second subderivative.

We proceed with an important consequence of Theorem 3.10 in which the proto-differentiability of subgradient mappings is established under parabolic regularity. Recall that a set-valued mapping $S : \mathbb{X} \rightrightarrows \mathbb{Y}$ is said to be *proto-differentiable* at $\bar{x}$ for $\bar{y}$ with

$(\bar{x}, \bar{y}) \in S$ if the set S is geometrically derivable at $(\bar{x}, \bar{y})$. When this condition holds for the set-valued mapping S at $\bar{x}$ for $\bar{y}$, we refer to $DS(\bar{x}, \bar{y})$ as the *proto-derivative* of S at $\bar{x}$ for $\bar{y}$. The connection between the twice epi-differentiablity of a function and the proto-differentiability of its subgradient mapping was observed first by Rockafellar in [52] for convex functions and was extended later in [45] for prox-regular functions. Recall that a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ is called prox-regular at $\bar{x}$ for $\bar{v}$ if f is finite at $\bar{x}$ and is locally l.s.c. around $\bar{x}$ with $\bar{v} \in \partial^- f(\bar{x})$ and there are constant $\varepsilon > 0$ and $r \geq 0$ such that for all $x \in I\!\!B_\varepsilon(\bar{x})$ with $f(x) \leq f(\bar{x}) + \varepsilon$ we have

$$f(x) \geq f(u) + \langle v, x - u\rangle - \tfrac{r}{2}\|x - u\|^2 \qquad \text{for all} \quad (u, v) \in (\mathrm{gph}\partial^- f) \cap I\!\!B_\varepsilon(\bar{x}, \bar{v}).$$

Moreover, recall that f is subdifferentially continuous at $\bar{x}$ for $\bar{v}$ if $(x_k, v_k) \to (\bar{x}, \bar{v})$ with $v_k \in \partial^- f(x_k)$, one has $f(x_k) \to f(\bar{x})$. Note that the original definitions of prox-regularity and subdifferentially continuous in [54, Definition 13.27] require v to be taken from $\partial^{L-} f(\bar{x})$. However, since in this chapter we mainly deal with strongly subamenable functions we keep the above definition. For strongly subamenable fucntions we have $\partial^- f(\bar{x}) = \partial^{L-} f(\bar{x})$; see Proposition 2.23.

Corollary 3.11 (proto-differentiability under parabolic regularity) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ be prox-regular and subdifferentially continuous at $\bar{x}$ for $\bar{v}$ and let f be parabolically epi-differentiable at $\bar{x}$ for every $w \in K_f(\bar{x}, \bar{v})$. If f is parabolically regular at $\bar{x}$ for $\bar{v}$, then the following equivalent conditions hold:*

- **(i)** *the function f is twice epi-differentiable at $\bar{x}$ for $\bar{v}$;*
- **(ii)** *the subgradient mapping ∂f is proto-differentiable at $\bar{x}$ for $\bar{v}$.*

Furthermore, the proto-derivative of the subgradient mapping ∂f at $\bar{x}$ for $\bar{v}$ can be calculated by

$$D(\partial^- f)(\bar{x}, \bar{v})(w) = \partial^{L-}\big(\mathrm{d}^2 f(\bar{x}, \bar{v})\big)(w) \quad \text{for all } w \in \mathbb{X}. \tag{3.21}$$

Proof.

Note that $\bar{v} \in \partial^p f(\bar{x})$ since f is prox-regular at $\bar{x}$ for $\bar{v}$. Employing now Theorem 3.10 gives us (i). The equivalence between (i) and (ii) and the validity of (4.2) come from [54, Theorem 13.40]. □

3.3 Variational Properties of Parabolic Subderivatives

This section is devoted to second-order analysis of parabolic subderivatives of extended-real-valued functions that are locally Lipschitz continuous relative to their domains. We pay special attention to functions that are expressed as a composition of a convex function and a twice differentiable function. We begin with the following result that gives us sufficient conditions for finding the domain of the parabolic subderivative.

Proposition 3.12 (properties of parabolic subderivatives) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ be finite at $\bar{x}$ and let f be Lipschitz continuous around $\bar{x}$ relative to its domain with constant $\ell \in \mathbb{R}_+$. Assume that $w \in T_{\operatorname{dom} f}(\bar{x})$ and that f is parabolic epi-differentiable at $\bar{x}$ for w. Then the following conditions hold:*

(i) $\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \,|\, \cdot) = T^2_{\operatorname{dom} f}(\bar{x}, w)$*;*

(ii) $\operatorname{dom} f$ *is parabolically derivable at $\bar{x}$ for w.*

Proof.

Since $w \in T_{\operatorname{dom} f}(\bar{x})$, we conclude from Proposition 2.8 that $\mathrm{d}f(\bar{x})(w)$ is finite. To prove (i), observe first that by definition, we always have the inclusion

$$\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \,|\, \cdot) \subset T^2_{\operatorname{dom} f}(\bar{x}, w). \tag{3.1}$$

To obtain the opposite inclusion, take $z \in T^2_{\operatorname{dom} f}(\bar{x}, w)$. This tells us that there exist sequences $t_k \downarrow 0$ and $z_k \to z$ so that $\bar{x} + t_k w + \frac{1}{2} t_k^2 z_k \in \operatorname{dom} f$. Since f is parabolically epi-differentiable at $\bar{x}$ for w, we have $\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \mid \cdot) \neq \emptyset$. Thus there exists a $z_w \in \mathbb{X}$ such that $\mathrm{d}^2 f(\bar{x})(w \mid z_w) < \infty$. Moreover, corresponding to the sequence t_k, we find another sequence $z'_k \to z_w$ such that

$$\mathrm{d}^2 f(\bar{x})(w \mid z_w) = \lim_{k\to\infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z'_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2}.$$

Since $\mathrm{d}^2 f(\bar{x})(w \mid z_w) < \infty$, we can assume without loss of generality that $\bar{x} + t_k w + \frac{1}{2}t_k^2 z'_k \in \operatorname{dom} f$ for all $k \in I\!N$. Using these together with the Lipschitz continuity of f around $\bar{x}$ relative to its domain, we have for all k sufficiently large that

$$\begin{aligned}
\frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2} &= \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z'_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2} \\
&\quad + \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z'_k)}{\frac{1}{2}t_k^2} \\
&\leq \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z'_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2} \\
&\quad + \ell\|z_k - z'_k\|.
\end{aligned}$$

Passing to the limit results in the inequality

$$\mathrm{d}^2 f(\bar{x})(w \mid z) \leq \mathrm{d}^2 f(\bar{x})(w \mid z_w) + \ell\|z - z_w\|, \tag{3.2}$$

which in turn yields $\mathrm{d}^2 f(\bar{x})(w \mid z) < \infty$, i.e., $z \in \operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \mid \cdot)$. This justifies the opposite inclusion in (3.1) and hence proves (i).

Turning now to (ii), we conclude from (3.1) and the parabolic epi-differentiability of f at $\bar{x}$ for w that the second-order tangent set $T^2_{\operatorname{dom} f}(\bar{x}, w)$ is nonempty. Moreover, it follows from [54, Example 13.62(b)] that the parabolic epi-differentiability of f at $\bar{x}$ for w yields the parabolic derivability of $\operatorname{epi} f$ at $(\bar{x}, f(\bar{x}))$ for $(w, \mathrm{d}f(\bar{x})(w))$. The latter clearly enforces the same property for $\operatorname{dom} f$ at $\bar{x}$ for w and hence completes the proof. $\square$

It is important to notice the parabolic epi-differentiability of f in Proposition 3.12 is essential to ensure that condition (i) therein, namely the characterization of the domain of the parabolic subderivative, is satisfied. Indeed, as mentioned in the proof of this proposition, inclusion (3.1) always holds. If the latter condition fails, this inclusion can be strict. For example, the function f from Example 3.7 is not parabolic epi-differentiable at $\bar{x} = (0,0)$ for any vector $w = (w_1, w_2) \in K_f(\bar{x}, \bar{v})$ with $w_1 \neq 0$ since $\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \,|\, \cdot) = \emptyset$. On the other hand, we have $\operatorname{dom} f = \mathbb{R}^2$ and thus $T^2_{\operatorname{dom} f}(\bar{x}, w) = \mathbb{R}^2$ for any such a vector $w \in K_f(\bar{x}, \bar{v})$, and so condition (i) in Proposition 3.12 fails.

Given a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ finite at $\bar{x}$, in the rest of this paper, we mainly focus on the case when this function has a representation of the form

$$f(x) = (g \circ F)(x) \quad \text{for all } \; x \in \mathcal{O}, \tag{3.3}$$

where $\mathcal{O}$ is a neighborhood of $\bar{x}$ and where the functions F and g are satisfying the following conditions:

- $F : \mathbb{X} \to \mathbb{Y}$ is twice differentiable at $\bar{x}$;
- $g : \mathbb{Y} \to \overline{\mathbb{R}}$ is proper, l.s.c., convex, and Lipschitz continuous around $F(\bar{x})$ relative to its domain with constant $\ell \in \mathbb{R}_+$.

It is not hard to see that the imposed assumptions on g from representation (3.3) implies that $\operatorname{dom} g$ is locally closed around $F(\bar{x})$, namely for some $\varepsilon > 0$ the set $(\operatorname{dom} g) \cap \mathbb{B}_\varepsilon(F(\bar{x}))$ is closed. Taking the neighborhood $\mathcal{O}$ from (3.3), we obtain

$$(\operatorname{dom} f) \cap \mathcal{O} = \big\{ x \in \mathcal{O} |\; F(x) \in \operatorname{dom} g \big\}. \tag{3.4}$$

It has been well understood that the second-order variational analysis of the composite form (3.3) requires a certain qualification condition. we will see the metric subregularity qualification condition (MSCQ) is enough for both first- and second-order analysis. For convenience we recall the metric subregularity qualification condition below. We say that the metric subregularity constraint qualification holds for the constraint set (3.4) (or composition

(3.3)) at $\bar{x}$ if there exist a constants $\kappa \in \mathbb{R}_+$ and a neighborhood U of $\bar{x}$ such that

$$\operatorname{dist}(x\ ,\ \operatorname{dom} f) \le \kappa\, d(F(x)\ ,\ \operatorname{dom} g) \quad \text{for all } \ x \in U. \tag{3.5}$$

Recall that from chapter 1, if the metric subregularity qualification condition holds at $\bar{x}$ for the composition function (3.3) then f is called strongly subamenable at $\bar{x}$ consequently first-order subderivative /subdifferential chain chain rules hold at $\bar{x}$; see Theorems 2.10 and 2.19. Moreover, f is prox-regular at $\bar{x}$ for every $v \in \partial^- f(\bar{x})$; see Theorem 2.24. For the convienence of the reader we gather above first-order result in the following proposition.

Proposition 3.13 (first-order chain rules for strongly subamenable function) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ have the composite representation* (3.3) *at $\bar{x} \in \operatorname{dom} f$ and let the metric subregularity qualification condition* (3.5) *hold for the composition* (3.3) *at $\bar{x}$. Then the following hold:*

(i) *for any $w \in \mathbb{X}$, the following subderivative chain rule for f at $\bar{x}$ holds:*

$$\mathrm{d}f(\bar{x})(w) = \mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})w)$$

(ii) *we have the chain rules*

$$\partial^p f(\bar{x}) = \partial^- f(\bar{x}) = \nabla F(\bar{x})^* \partial g(F(\bar{x})) \quad \textit{and} \quad T_{\operatorname{dom} f}(\bar{x}) = \big\{w \in \mathbb{X} |\ \nabla F(\bar{x})w \in T_{\operatorname{dom} g}(F(\bar{x}))\big\}.$$

To proceed with our tangential second-order analysis, for each $w \in \mathbb{R}^n$ satisfying $\nabla F(\bar{x})w \in T_{\operatorname{dom} g}(F(\bar{x}))$ define the parameterized set-valued mapping $S_w \colon \mathbb{Y} \rightrightarrows \mathbb{X}$ involving the second-order tangent set (3.1) by

$$S_w(p) := \big\{u \in \mathbb{X} \big|\ \nabla F(\bar{x})u + \nabla^2 F(\bar{x})(w,w) + p \in T^2_{\operatorname{dom} g}\big(F(\bar{x}), \nabla F(\bar{x})w\big)\big\}. \tag{3.6}$$

This mapping describes a canonically perturbed *second-order tangential approximation* of the constraint system (3.4). The next result of its own interest proves under metric subregularity constraint qualification the *uniform outer/upper Lipschitz* property of (3.6) in the sense of

Robinson [46] broadly employed below.

Theorem 3.14 (uniform outer Lipschitzian property of $S_w(.)$) *Let* $\mathrm{dom}\, f \cap \mathcal{O}$ *be a constraint system represented by* (3.4) *around* $\bar{x} \in \mathrm{dom}\, f$*, and let* $w \in \mathbb{R}^n$ *be such that* $\nabla F(\bar{x})w \in T_{\mathrm{dom}\, g}(F(\bar{x}))$*. Assume that* (3.5) *holds at* $\bar{x}$ *with modulus* $\kappa > 0$*. Then the approximating mapping* (3.6) *satisfies the inclusion*

$$S_w(p) \subset S_w(0) + \kappa\|p\|I\!B \;\; \text{for all} \;\; p \in \mathbb{R}^m \;\; \text{uniformly in} \;\; w, \tag{3.7}$$

which means the uniform outer Lipschitzian property of S_w *at the origin.*

Proof.

Fixing some $p \in \mathbb{Y}$ and $u \in S_w(p)$, we get by (3.6) that

$$\nabla F(\bar{x})u + \nabla^2 F(\bar{x})(w,w) + p \in T^2_{\mathrm{dom}\, g}\big(F(\bar{x}), \nabla F(\bar{x})w\big).$$

We deduce from definition (3.1) of second-order tangents that there exists a sequence $t_k \downarrow 0$ with

$$F(\bar{x}) + t_k \nabla F(\bar{x})w + \tfrac{1}{2}t_k^2\big(\nabla F(\bar{x})u + \nabla^2 F(\bar{x})(w,w) + p\big) + o(t_k^2) \in \Theta, \quad k \in I\!N.$$

For any k sufficiently large we get by the twice differentiability of F at $\bar{x}$ that

$$F\big(\bar{x} + t_k w + \tfrac{1}{2}t_k^2 u\big) = F(\bar{x}) + t_k \nabla F(\bar{x})w + \tfrac{1}{2}t_k^2\big(\nabla F(\bar{x})u + \nabla^2 F(\bar{x})(w,w)\big) + o(t_k^2),$$

which in turn implies via MSCQ (3.5) that

$$\begin{aligned} \mathrm{dist}\big(\bar{x} + t_k w + \tfrac{1}{2}t_k^2 u \,;\, \mathrm{dom}\, f\big) &\le \kappa\, \mathrm{dist}\big(F(\bar{x} + t_k w + \tfrac{1}{2}t_k^2 u) \,;\, \mathrm{dom}\, g\big) \\ &\le \frac{1}{2}\kappa t_k^2\Big(\|p\| + \frac{o(t_k^2)}{t_k^2}\Big). \end{aligned}$$

Thus there exists a vector $y_k \in \mathrm{dom}\, f$ satisfying

$$\|d_k\| \le \frac{1}{2}\kappa\Big(\|p\| + \frac{o(t_k^2)}{t_k^2}\Big) \;\; \text{with} \;\; d_k := \frac{\bar{x} + t_k w + \frac{1}{2}t_k^2 u - y_k}{t_k^2}.$$

Passing to a subsequence if necessary ensures the existence of $d \in \mathbb{X}$ such that $d_k \to d$ as

$k \to \infty$. This yields the estimate

$$\|d\| \leq \tfrac{1}{2}\kappa\|p\|. \tag{3.8}$$

On the other hand, we can suppose without loss of generality that $\bar{x}+t_k w+\frac{1}{2}t_k^2 u-t_k^2 d_k = y_k \in \Omega\cap\mathcal{O}$ for k sufficiently large, and hence it follows from (3.4) that $f(\bar{x}+t_k w+\frac{1}{2}t_k^2 u-t_k^2 d_k) \in \operatorname{dom} g$. Taking into account the representation

$$D\big(\bar{x}+t_k w+\tfrac{1}{2}t_k^2 u-t_k^2 d_k\big) = F(\bar{x})+t_k\nabla F\big(\bar{x})w+\tfrac{1}{2}t_k^2\big(\nabla F(\bar{x})(u-2d_k)+\nabla^2 F(\bar{x})(w,w)\big)+o(t_k^2),$$

we readily arrive at the inclusion

$$F(\bar{x})+t_k\nabla F(\bar{x})w+\frac{1}{2}t_k^2\Big(\nabla F(\bar{x})(u-2d_k)+\nabla^2 F(\bar{x})(w,w)+\frac{2o(t_k^2)}{t_k^2}\Big)\in \operatorname{dom} g,$$

which in turn implies that $\nabla F(\bar{x})(u-2d)+\nabla^2 F(\bar{x})(w,w)\in T^2_{\operatorname{dom} g}(F(\bar{x}),\nabla F(\bar{x})w)$. The latter reads as $u-2d\in S_w(0)$, which together with (3.8) justifies the claimed inclusion (3.7) that gives us the uniform outer Lipschitzian property of the mapping S_w from (3.6) at $p=0$. $\square$

Let us make some comments to the second-order result obtained in Theorem 3.14.

Remark 3.15 (discussions on the outer Lipschitzian property) The following hold:

(i) The result of Theorem 3.14 reduces to [17, Proposition 3.1] in the case where the set $\Theta := \operatorname{dom} g$ is a closed convex cone and $F(\bar{x}) = 0$; neither of these conditions is in our assumptions. Indeed, we can easily observe that the assumptions of [17] ensure that $T^2_\Theta(F(\bar{x}),\nabla F(\bar{x})w) = T_\Theta(\nabla F(\bar{x})w)$, which allows us to derive the result of [17, Proposition 3.1] from Theorem 3.14.

(ii) Although Theorem 3.14 is verified for vectors $w\in\mathbb{X}$ with $\nabla F(\bar{x})w\in T_\Theta(F(\bar{x}))$, it is clear that the outer Lipschitzian property (3.7) holds in fact for all vectors $w\in\mathbb{X}$. To check (3.7) for w with $\nabla F(\bar{x})w\notin T_\Theta(F(\bar{x}))$, we observe directly from the definition that $T^2_\Theta(F(\bar{x}),\nabla F(\bar{x})w)=\emptyset$ and hence $S_w(p)=\emptyset$ for all $p\in\mathbb{Y}$. This clearly yields (3.7).

(iii) Note finally that Theorem 3.14 implies the outer Lipschitzian property of the mapping

$$p \mapsto \big\{ w \in \mathbb{R}^n \big|\ \nabla F(\bar{x})w + p \in T_\Theta\big(F(\bar{x})\big)\big\}.$$

This can be easily deduced from Theorem 3.14 by letting $w = 0 \in \mathbb{R}^n$ and by observing that $T^2_\Theta(F(\bar{x}), 0) = T_\Theta(F(\bar{x}))$. This was already observed at [15, Proposition 2.1].

We are now ready to provide an application of Theorem 3.14 to establishing the parabolic derivability of constraint systems (3.4) via a chain rule for second-order tangent sets under MSQC (2.8). Such a chain rule for (3.4) was obtained in [54, Proposition 13.13] and also in [4, Proposition 3.33] under the much stronger metric regularity condition for the mapping $x \mapsto F(x) - \mathrm{dom}\, g$ around $(\bar{x}, 0)$.

Proposition 3.16 (second-order tangent set chain rules) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ have the composite representation* (3.3) *at $\bar{x} \in \mathrm{dom}\, f$ where the function g from* (3.3) *is parabolically epi-differentiable at $F(\bar{x})$ for $\nabla F(\bar{x})w \in \mathrm{dom}\, g$. Finally assume the metric subregularity constraint qualification condition* (3.5) *holds at $\bar{x}$. Then $w \in T_{\mathrm{dom}\, f}(\bar{x})$ and we have*

$$z \in T^2_{\mathrm{dom}\, f}(\bar{x}, w) \iff \nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w, w) \in T^2_{\mathrm{dom}\, g}\big(F(\bar{x}), \nabla F(\bar{x})w\big). \tag{3.9}$$

Moreover, $\mathrm{dom}\, f$ *is parabolically derivable at $\bar{x}$ for w.*

Proof.

We have $w \in T_{\mathrm{dom}\, f}(\bar{x})$ by Proposition 3.13 (ii). To prove the second-order tangent set chain rule (3.9), by a close look at the proof of (3.9), which was given in [54, Proposition 13.13] under the metric regularity property of the mapping $x \mapsto F(x) - \mathrm{dom}\, g$ around $(\bar{x}, 0)$, we can observe that it actually utilizes merely MSQC at this point. To verify the claimed parabolic derivability of the constraint system (3.4) under the assumptions made, pick any $w \in T_\Omega(\bar{x})$ and recall that $T^2_\Omega(\bar{x}, w)$ in (3.1) can be reformulated via the outer limit of the sets

$(\operatorname{dom} f - \bar{x} - tw)/\frac{1}{2}t^2$ as $t \downarrow 0$. The first requirement of parabolic derivability is to show that this outer limit is actually achieved as the *full set limit* meaning that the outer and inner limits agree. This again can be done by following the proof of [54, Proposition 13.13], which basically works under MSQC. The second requirement of parabolic derivability is crucial: to show that $T^2_{\operatorname{dom} f}(\bar{x}, w) \neq \emptyset$ for any tangent vector $w \in T_{\operatorname{dom} f}(\bar{x})$. The proof of the latter fact given in [54] heavily exploits the metric regularity of the constraint mapping and does not hold under MSQC. Now we provide a new proof for this property, which needs merely MSQC.

To proceed, employ the imposed parabolic derivability of $\operatorname{dom} g$ at $F(\bar{x})$ for $\nabla F(\bar{x})w$ to conclude that $T^2_{\operatorname{dom} g}(F(\bar{x}), \nabla F(\bar{x})w) \neq \emptyset$. Picking $z \in T^2_{\operatorname{dom} g}(F(\bar{x}), \nabla F(\bar{x})w)$ gives us the inclusion

$$\nabla F(\bar{x})u + \nabla^2 F(\bar{x})(w,w) + p \in T^2_{\operatorname{dom} g}\big(F(\bar{x}), \nabla F(\bar{x})w\big) \;\text{ with }\; p := z - \nabla F(\bar{x})u - \nabla^2 F(\bar{x})(w,w),$$

which can be equivalently expressed as $u \in S_w(p)$ via the mapping S_w from (3.6). Now we apply Theorem 3.14 and deduce from the outer Lipschitzian property (3.7) that there exists a vector $\widetilde{u} \in S_w(0)$ such that $\|u - \widetilde{u}\| \leq \kappa \|p\|$. This tells us that

$$\nabla F(\bar{x})\widetilde{u} + \nabla^2 F(\bar{x})(w,w) \in T^2_{\operatorname{dom} g}\big(F(\bar{x}), \nabla F(\bar{x})w\big).$$

Using the chain rule (3.9) leads us to $\widetilde{u} \in T^2_{\operatorname{dom} f}(\bar{x}, w)$, which verifies the nonemptiness of the second-order tangent set $T^2_{\operatorname{dom} f}(\bar{x}, w)$. This completes the proof. □

We continue by establishing a chain rule for the parabolic subderivative, which is important for our developments in the next section.

Theorem 3.17 (chain rule for parabolic subderivatives) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ have the composite representation* (3.3) *at $\bar{x} \in \operatorname{dom} f$ and $w \in T_{\operatorname{dom} f}(\bar{x})$ and let the metric subregularity constraint qualification hold for the constraint set* (3.4) *at $\bar{x}$. Assume that the function g from* (3.3) *is parabolically epi-differentiable at $F(\bar{x})$ for $\nabla F(\bar{x})w$. Then the fol-*

lowing conditions are satisfied:

(i) *for any $z \in \mathbb{X}$ we have*

$$\mathrm{d}^2 f(\bar{x})(w \mid z) = \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid \nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w,w)); \tag{3.10}$$

(ii) *the domain of the parabolic subderivative of f at $\bar{x}$ for w is given by*

$$\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \mid \cdot) = T^2_{\operatorname{dom} f}(\bar{x}, w);$$

(iii) *f is parabolically epi-differentiable at $\bar{x}$ for w.*

Proof.

Pick $z \in \mathbb{X}$ and set $u := \nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w,w)$. We prove (i)-(iii) in a parallel way. Assume that $z \notin T^2_{\operatorname{dom} f}(\bar{x}, w)$. As mentioned in the proof of Proposition 3.12, inclusion (3.1) always holds. This implies that $\mathrm{d}^2 f(\bar{x})(w \mid z) = \infty$. On the other hand, by (3.9) we get $u \notin T^2_{\operatorname{dom} g}\big(F(\bar{x}), \nabla F(\bar{x})w\big)$. Employing Proposition 3.12(i) for the function g and $\nabla F(\bar{x})w \in T_{\operatorname{dom} g}(F(\bar{x}))$ gives us

$$\operatorname{dom} \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid \cdot) = T^2_{\operatorname{dom} g}(F(\bar{x}), \nabla F(\bar{x})w). \tag{3.11}$$

Combining these tells us that $\mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid u) = \infty$, which in turn justifies (3.10) for every $z \notin T^2_{\operatorname{dom} f}(\bar{x}, w)$. Consider an arbitrary sequence $t_k \downarrow 0$ and set $z_k := z$ for all $k \in I\!N$. Then we have

$$\begin{aligned}
\mathrm{d}^2 f(\bar{x})(w \mid z) &\le \liminf_{k\to\infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2} \\
&\le \limsup_{k\to\infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2} \\
&\le \infty = \mathrm{d}^2 f(\bar{x})(w \mid z),
\end{aligned}$$

which in turn justifies (3.5) for all $z \notin T^2_{\operatorname{dom} f}(\bar{x}, w)$.

Since g is parabolically epi-differentiable at $F(\bar{x})$ for $\nabla F(\bar{x})w$, Proposition 3.12(ii) tells us that $\operatorname{dom} g$ is parabolically derivable at $F(\bar{x})$ for $\nabla F(\bar{x})w$. We conclude from Proposition 3.16

that $\operatorname{dom} f$ is parabolically derivable at $\bar{x}$ for w. In particular, we have

$$T^2_{\operatorname{dom} f}(\bar{x}, w) \neq \emptyset. \tag{3.12}$$

Pick now $z \in T^2_{\operatorname{dom} f}(\bar{x}, w)$ and then consider an arbitrary sequence $t_k \downarrow 0$. Thus, by definition, for the aforementioned sequence t_k, we find a sequence $z_k \to z$ as $k \to \infty$ such that

$$x_k := \bar{x} + t_k w + \frac{1}{2}t_k^2 z_k \in \operatorname{dom} f \quad \text{for all} \;\; k \in I\!N. \tag{3.13}$$

Moreover, since g is parabolically epi-differentiable at $F(\bar{x})$ for $\nabla F(\bar{x})w$, we find a sequence $u_k \to u$ such that

$$\mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid u) = \lim_{k\to\infty} \frac{g(F(\bar{x}) + t_k \nabla F(\bar{x})w + \frac{1}{2}t_k^2 u_k) - g(F(\bar{x})) - t_k \mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})w)}{\frac{1}{2}t_k^2}. \tag{3.14}$$

It follows from (3.9) that $u \in T^2_{\operatorname{dom} g}\big(F(\bar{x}), \nabla F(\bar{x})w\big)$. Combining this with (3.11) tells us that $\mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid u) < \infty$. This implies that $y_k := F(\bar{x}) + t_k \nabla F(\bar{x})w + \frac{1}{2}t_k^2 u_k \in \operatorname{dom} g$ for all k sufficiently large. Remember that g is Lipschitz continuous around $F(\bar{x})$ relative to its domain with constant ℓ. Using this together with Proposition 3.16(i), (3.13), and (3.14), we obtain

$$\begin{aligned}
\mathrm{d}^2 f(\bar{x})(w \mid z) \;\; &\leq \;\; \liminf_{k\to\infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2} \\
&\leq \;\; \limsup_{k\to\infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2} \\
&= \;\; \limsup_{k\to\infty} \frac{g(F(x_k)) - g(F(\bar{x})) - t_k \mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})w)}{\frac{1}{2}t_k^2} \\
&\leq \;\; \lim_{k\to\infty} \frac{g(y_k) - g(F(\bar{x})) - t_k \mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})w)}{\frac{1}{2}t_k^2} + \limsup_{k\to\infty} \frac{g(F(x_k)) - g(y_k)}{\frac{1}{2}t_k^2} \\
&\leq \;\; \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid u) + \limsup_{k\to\infty} \ell \|\nabla F(\bar{x})z_k + \nabla^2 F(\bar{x})(w,w) - u_k + \frac{o(t_k^2)}{t_k^2}\| \\
&= \;\; \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid u).
\end{aligned} \tag{3.15}$$

On the other hand, it is not hard to see that for any $z \in \mathbb{X}$, we always have

$$\mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid u) \leq \mathrm{d}^2 f(\bar{x})(w \mid z).$$

Combining this and (3.15) implies that

$$\mathrm{d}^2 f(\bar{x})(w \mid z) = \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid u)$$

and that

$$\mathrm{d}^2 f(\bar{x})(w \mid z) = \lim_{k\to\infty} \frac{f(\bar{x} + t_k w + \frac{1}{2}t_k^2 z_k) - f(\bar{x}) - t_k \mathrm{d}f(\bar{x})(w)}{\frac{1}{2}t_k^2}.$$

These prove (3.10) and (3.5) for any $z \in T^2_{\mathrm{dom}\, f}(\bar{x}, w)$, respectively, and hence we finish the proof of (i).

Next, we are going to verify (ii). We already know that inclusion (3.1) always holds. To derive the opposite inclusion, pick $z \in T^2_{\mathrm{dom}\, f}(\bar{x}, w)$, which amounts to $u \in T^2_{\mathrm{dom}\, g}\big(F(\bar{x}), \nabla F(\bar{x})w\big)$ due to (3.9). By (i) and (3.11), we obtain

$$\mathrm{d}^2 f(\bar{x})(w \mid z) = \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid u) < \infty.$$

This tells us that $z \in \mathrm{dom}\, \mathrm{d}^2 f(\bar{x})(w \mid \cdot)$ and hence completes the proof of (ii).

Finally, to justify (iii), we require to prove the fulfillment of (3.5) for all $z \in \mathbb{X}$ and to show that $\mathrm{dom}\, \mathrm{d}^2 f(\bar{x})(w \mid \cdot) \neq \emptyset$. The former was proven above and so we proceed with the proof of the latter. This, indeed, follows from (3.12) and the characterization of $\mathrm{dom}\, \mathrm{d}^2 f(\bar{x})(w \mid \cdot)$, achieved in (ii), and thus completes the proof. $\square$

It is worth mentioning that a chain rule for parabolic subderivatives for the composite form (3.3) was achieved in [54, Exercise 13.63] and [4, Proposition 3.42] when g is merely a proper l.s.c. function and the basic constraint qualification (2.4) is satisfied. Replacing the latter condition with the significantly weaker condition (3.5), we can achieve a similar result if we assume further that g is convex and locally Lipschitz continuous relative to its domain. Another important difference between Theorem 3.17 and those mentioned above is that the chain rule (3.10) obtained in [4, 54] does not require the parabolic epi-differentiability of g. Indeed, the usage of the basic constraint qualification (2.4) in [4, 54] allows to achieve (3.10) via a chain rule for the epigraphs of f and g similar to the one in (3.9), which is not conceivable under (3.5). These extra assumptions on g automatically fulfill in many important composite and constrained optimization problems and so do not seem to be restrictive in our developments.

We continue by establishing two important properties for parabolic subderivatives that play crucial roles in our developments in the next section. One notable difference between the following results and those obtained in Proposition 3.12 and Theorem 3.17 is that we require the parabolic subderivative be proper. This can be achieved if the parabolic subderivative is bounded below. In general, we may not be able to guarantee this. It turns out, however, that if the vector w in the pervious results is taken from the critical cone to the function in question, which is a subset of the tangent cone to the domain of that function, this can be accomplished via (3.14). Since we only conduct our analysis in the next section over the critical cone, this will provide no harm. Below, we first show that the parabolic subderivative of an extended-real-valued function, which is locally Lipschitz continuous relative to its domain, is Lipschitz continuous relative to its domain.

Proposition 3.18 (Lipschitz continuity of of parabolic subderivatives) *Let $\psi : \mathbb{X} \to \overline{\mathbb{R}}$ be finite at $\bar{x}$ and $\bar{v} \in \partial^p \psi(\bar{x})$, and let ψ be Lipschitz continuous around $\bar{x}$ relative to its domain with constant $\ell \in \mathbb{R}_+$. Assume that $w \in K_\psi(\bar{x}, \bar{v})$ and that ψ is parabolically epi-differentiable at $\bar{x}$ for w. Then the parabolic subderivative $\mathrm{d}^2\psi(\bar{x})(w \mid \cdot)$ is proper, l.s.c., and Lipschitz continuous relative to its domain with constant ℓ.*

*Proof.*Since ψ is parabolically epi-differentiable at $\bar{x}$ for w, we get $\mathrm{dom}\, \mathrm{d}^2\psi(\bar{x})(w \mid \cdot) \neq \emptyset$. Let $z \in \mathrm{dom}\, \mathrm{d}^2\psi(\bar{x})(w \mid \cdot)$. By Proposition 3.6, we find $r \in \mathbb{R}_+$ such that

$$-\, r\|w\|^2 \leq \mathrm{d}^2\psi(\bar{x}, \bar{v})(w) \leq \mathrm{d}^2\psi(\bar{x})(w \mid z) - \langle z, \bar{v} \rangle. \tag{3.16}$$

This tells us that $\mathrm{d}^2\psi(\bar{x})(w \mid z)$ is finite for every $z \in \mathrm{dom}\, \mathrm{d}^2\psi(\bar{x})(w \mid \cdot)$ and thus the parabolic subderivative $\mathrm{d}^2\psi(\bar{x})(w \,|\, \cdot)$ is proper. Pick now $z_i \in \mathrm{dom}\, \mathrm{d}^2\psi(\bar{x})(w \,|\, \cdot)$ for $i = 1, 2$. By Proposition 3.12(i), we have $z_i \in T^2_{\mathrm{dom}\,\psi}(\bar{x}, w)$ for $i = 1, 2$. Letting $z := z_1$ and $z_w := z_2$ in (3.2) results in

$$\mathrm{d}^2\psi(\bar{x})(w \,|\, z_1) \leq \mathrm{d}^2\psi(\bar{x})(w \,|\, z_2) + \ell\|z_1 - z_2\|.$$

Similarly, we can let $z := z_2$ and $z_w := z_1$ in (3.2) and obtain

$$\mathrm{d}^2\psi(\bar{x})(w \,|\, z_2) \leq \mathrm{d}^2\psi(\bar{x})(w \,|\, z_1) + \ell\|z_1 - z_2\|.$$

Combining these implies that the parabolic subderivative is Lipschitz continuous relative to its domain. By [54, Proposition 13.64], the parabolic subderivative is always an l.s.c. function, which completes the proof. $\square$

We end this section by obtaining an exact formula for the conjugate function of the parabolic subderivative of a convex function.

Proposition 3.19 (conjugate of parabolic subderivatives) *Let $\psi : \mathbb{X} \to \overline{\mathbb{R}}$ be an l.s.c. convex function with $\psi(\bar{x})$ finite, $\bar{v} \in \partial\psi(\bar{x})$, and $w \in K_\psi(\bar{x}, \bar{v})$. Define the function φ by $\varphi(z) := \mathrm{d}^2\psi(\bar{x})(w \,|\, z)$ for any $z \in \mathbb{X}$. If ψ is parabolically epi-differentiable at $\bar{x}$ for w and parabolically regular at $\bar{x}$ for every $v \in \partial\psi(\bar{x})$, then φ is a proper, l.s.c., and convex function and its conjugate function is given by*

$$\varphi^*(v) = \begin{cases} -\mathrm{d}^2\psi(\bar{x}, v)(w) & \text{if } v \in \mathcal{A}(\bar{x}, w), \\ \infty & \text{otherwise}, \end{cases} \tag{3.17}$$

where $\mathcal{A}(\bar{x}, w) := \{v \in \partial\psi(\bar{x}) \,|\, \mathrm{d}\psi(\bar{x})(w) = \langle v, w\rangle\}$.

Proof.

It follows from [54, Proposition 13.64] that φ is l.s.c. Using similar arguments as the beginning of the proof of Proposition 3.18 together with (3.16) tells us that φ is proper.

Also we deduce from [54, Example 13.62] that

$$\operatorname{epi}\varphi = T^2_{\operatorname{epi}\psi}\big((\bar{x},\psi(\bar{x})),(w,\mathrm{d}\psi(\bar{x})(w))\big),$$

and thus the parabolic epi-differentiability of ψ at $\bar{x}$ for w amounts to the parabolic derivability of $\operatorname{epi}\psi$ at $(\bar{x},\psi(\bar{x}))$ for $(w,\mathrm{d}\psi(\bar{x})(w))$. The latter combined with the convexity of ψ tells us that $\operatorname{epi}\varphi$ is a convex set in $\mathbb{X}\times\mathbb{R}$ and so φ is convex.

To verify (3.17), pick $v \in \mathcal{A}(\bar{x},w)$. This yields $v \in \partial\psi(\bar{x}) = \partial^p\psi(\bar{x})$ and $w \in K_\psi(\bar{x},v)$, namely the critical cone of ψ at $(\bar{x},v)$. Using Proposition 3.8 and parabolic regularity of ψ at $\bar{x}$ for v implies that

$$\mathrm{d}^2\psi(\bar{x},v)(w) = \inf_{z\in\mathbb{X}}\big\{\mathrm{d}^2\psi(\bar{x})(w\,|\,z) - \langle z,v\rangle\big\} = -\varphi^*(v),$$

which clearly proves (3.17) in this case. Assume now that $v \notin \mathcal{A}(\bar{x},w)$. This means either $v \notin \partial\psi(\bar{x})$ or $\mathrm{d}\psi(\bar{x})(w) \neq \langle v,w\rangle$. Define the parabolic difference quotients for ψ at $\bar{x}$ for w by

$$\vartheta_t(z) = \frac{\psi(\bar{x}+tw+\frac{1}{2}t^2z) - \psi(\bar{x}) - t\mathrm{d}\psi(\bar{x})(w)}{\frac{1}{2}t^2}, \quad z\in\mathbb{X},\ \ t>0.$$

It is not hard to see that ϑ_t are proper, convex, and

$$\vartheta_t^*(v) = \frac{\psi(\bar{x})+\psi^*(v)-\langle v,\bar{x}\rangle}{\frac{1}{2}t^2} + \frac{\mathrm{d}\psi(\bar{x})(w)-\langle v,w\rangle}{\frac{1}{2}t}, \quad v\in\mathbb{X}.$$

Remember that by [54, Definition 13.59] the parabolic epi-differentiability of ψ at $\bar{x}$ for w amounts to the sets $\operatorname{epi}\vartheta_t$ converging to $\operatorname{epi}\varphi$ as $t\downarrow 0$ and that the functions ϑ_t and φ are proper, l.s.c. and convex. Appealing to [54, Theorem 11.34] tells us that the former is equivalent to the sets $\operatorname{epi}\vartheta_t^*$ converging to $\operatorname{epi}\varphi^*$ as $t\downarrow 0$. This, in particular, means that for any $v\notin\mathcal{A}(\bar{x},w)$ and any sequence $t_k\downarrow 0$, there exists a sequence $v_k\to v$ such that

$$\varphi^*(v) = \lim_{k\to\infty}\vartheta_{t_k}^*(v_k).$$

If $v \notin \partial\psi(\bar{x})$, then we have

$$\psi(\bar{x}) + \psi^*(v) - \langle v, \bar{x} \rangle > 0.$$

Since ψ^* is l.s.c., we get

$$\liminf_{k\to\infty} \frac{\psi(\bar{x}) + \psi^*(v_k) - \langle v_k, \bar{x} \rangle}{\frac{1}{2}t_k} + \frac{\mathrm{d}\psi(\bar{x})(w) - \langle v_k, w \rangle}{\frac{1}{2}} \geq \infty,$$

which in turn confirms that

$$\varphi^*(v) = \lim_{k\to\infty} \vartheta^*_{t_k}(v_k) = \lim_{k\to\infty} \frac{1}{t_k}\Big(\frac{\psi(\bar{x}) + \psi^*(v_k) - \langle v_k, \bar{x} \rangle}{\frac{1}{2}t_k} + \frac{\mathrm{d}\psi(\bar{x})(w) - \langle v_k, w \rangle}{\frac{1}{2}}\Big) = \infty.$$

If $v \in \partial\psi(\bar{x})$ but $\mathrm{d}\psi(\bar{x})(w) \neq \langle v, w \rangle$, we obtain $\langle v, w \rangle < \mathrm{d}\psi(\bar{x})(w)$. Since we always have

$$\psi(\bar{x}) + \psi^*(v_k) - \langle v_k, \bar{x} \rangle \geq 0 \quad \text{ for all } \ k \in I\!N,$$

we arrive at

$$\varphi^*(v) = \lim_{k\to\infty} \vartheta^*_{t_k}(v_k) \geq \lim_{k\to\infty} \frac{\mathrm{d}\psi(\bar{x})(w) - \langle v_k, w \rangle}{\frac{1}{2}t_k} = \infty,$$

which justifies (3.17) when $v \notin \mathcal{A}(\bar{x}, w)$ and hence finishes the proof. □

Proposition 3.19 was first established using a different method in [49, Proposition 3.5] for convex piecewise linear-quadratic functions. It was extended in [10, Theorem 3.1] for any convex functions under a restrictive assumption. Indeed, this result demands that the second subderivative be the same as the second-order directional derivative. Although this condition holds for convex piecewise linear-quadratic functions, it fails for many important functions occurring in constrained and composite optimization problems including the set indicator functions and eigenvalue functions. As discussed below, however, our assumptions are satisfied for all these examples.

Example 3.20 Suppose that $g : \mathbb{Y} \to \overline{\mathbb{R}}$ is an l.s.c. convex function and $\bar{z} \in \mathbb{Y}$.

(a) If $\mathbb{Y} = \mathbb{R}^m$, g is convex piecewise linear-quadratic (Example 3.4), and $\bar{z} \in \operatorname{dom} g$, then it follows from Example 3.4 and [54, Exercise 13.61] that g is parabolically regular at

$\bar{z}$ for every $y \in \partial g(\bar{z})$ and parabolically epi-differentiable at $\bar{z}$ for every $w \in \operatorname{dom} \mathrm{d}g(\bar{z})$, respectively, and thus all the assumptions of Proposition 3.19 are satisfied for this function.

(b) If $\mathbb{Y} = \mathcal{S}^m$, g is either the maximum eigenvalue function $\lambda_{\max}$ from (3.11) or the function σ_i from (3.12), and $A \in \mathcal{S}^n$, then by Example 3.5 g is parabolically regular at A for every $V \in \partial g(A)$. Moreover, we deduce from [56, Proposition 2.2] that g is parabolically epi-differentiable at A for every $W \in \mathcal{S}^n$ and thus all the assumptions of Proposition 3.19 are satisfied for these functions.

(c) If $g = \delta_\Theta$ and $\bar{z} \in \Theta$, where Θ is a closed convex set in $\mathbb{Y}$ that is parabolically derivable at $\bar{z}$ for every $w \in T_\Theta(\bar{z})$ and parabolically regular at $\bar{z}$ for every $v \in N_\Theta(\bar{z})$, then g satisfies the assumptions imposed in Proposition 3.19. This example of g was recently explored in detail in [39] and encompasses important sets appearing in constrained optimization problems such as polyhedral convex sets, the second-order cone, and the cone of positive semidefinite symmetric matrices.

(d) Assume that g is differentiable at $\bar{z}$ and that there exists a continuous function $h : \mathbb{Y} \to \mathbb{R}$, which is positively homogeneous of degree 2, such that

$$g(z) = g(\bar{z}) + \langle \nabla g(\bar{z}), z - \bar{z} \rangle + \tfrac{1}{2} h(z - \bar{z}) + o(\|z - \bar{z}\|^2).$$

Such a function g is called twice semidifferentiable (cf. [54, Example 13.7]) and often appears in the augmented Lagrangian function associated with (1.1); see [39, Section 8] for more detail. This second-order expansion clearly justifies the parabolic epi-differentiability of g at $\bar{z}$ for every $w \in \mathbb{Y}$. Moreover, one has

$$\mathrm{d}^2 g(\bar{z}, \nabla g(\bar{z}))(w) = h(w) = \mathrm{d}^2 g(\bar{z})(w \,|\, u) - \langle \nabla g(\bar{z}), u \rangle \quad \text{for all} \quad u, w \in \mathbb{Y},$$

which in turn shows that g is parabolically regular at $\bar{z}$ for $\nabla g(\bar{z})$ due to Proposition 3.8.

It is important to mention that the restrictive assumption on the second subderivative, used in [10, Theorem 3.1], does not hold for cases (b)-(d) in Example 3.20.

3.4 Chain Rule for Parabolically Regular Functions

Our main objective in this section is to derive a chain rule for the parabolic regularity of the composite representation (3.3). This opens the door to obtain a chain rule for the second subderivative, and, more importantly, allows us to establish the twice epi-differentiability of the latter composite form.

Taking into account representation (3.3) and picking a subgradient $\bar{v} \in \partial_{-} f(\bar{x})$, we define the set of *Lagrangian multipliers* associated with $(\bar{x}, \bar{v})$ by

$$\Lambda(\bar{x}, \bar{v}) = \big\{ y \in \mathbb{Y} \,|\, \nabla F(\bar{x})^* y = \bar{v},\ y \in \partial g(F(\bar{x})) \big\}.$$

In what follows, we say that a function $f : \mathbb{X} \to \overline{\mathbb{R}}$ with $(\bar{x}, \bar{v}) \in \partial f$ and having the composite representation (3.3) at $\bar{x}$ satisfies the basic assumptions at $(\bar{x}, \bar{v})$ if in addition the following conditions fulfill:

(H1) the metric subregularity constraint qualification holds for the constraint set (3.4) at $\bar{x}$;

(H2) for any $y \in \Lambda(\bar{x}, \bar{v})$, the function g from (3.3) is parabolically epi-differentiable at $F(\bar{x})$ for every $u \in K_g(F(\bar{x}), y)$;

(H3) for any $y \in \Lambda(\bar{x}, \bar{v})$, the function g is parabolically regular at $F(\bar{x})$ for y.

We begin with the following result in which we collect lower and upper estimates for the second subderivative of f taken from (3.3).

Proposition 3.21 (properities of second subderivatives for composite functions) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ have the composite representation* (3.3) *at $\bar{x} \in \operatorname{dom} f$, $\bar{v} \in \partial f(\bar{x})$, and let the basic assumptions* (H1) *and* (H2) *hold for f at $(\bar{x}, \bar{v})$. Then the second subderivative*

$\mathrm{d}^2 f(\bar{x},\bar{v})$ *is a proper l.s.c. function with*

$$\operatorname{dom} \mathrm{d}^2 f(\bar{x},\bar{v}) = K_f(\bar{x},\bar{v}). \tag{3.18}$$

Moreover, for every $w \in \mathbb{X}$ *we have the lower estimate*

$$\mathrm{d}^2 f(\bar{x},\bar{v})(w) \geq \sup_{y\in\Lambda(\bar{x},\bar{v})} \big\{\langle y, \nabla^2 F(\bar{x})(w,w)\rangle + \mathrm{d}^2 g(F(\bar{x}),y)(\nabla F(\bar{x})w)\big\}, \tag{3.19}$$

while for every $w \in K_f(\bar{x},\bar{v})$ *we obtain the upper estimate*

$$\mathrm{d}^2 f(\bar{x},\bar{v})(w) \leq \inf_{z\in\mathbb{X}} \big\{ -\langle z,\bar{v}\rangle + \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid \nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w,w))\big\} < \infty \tag{3.20}$$

Proof.

By Proposition 3.16(ii), we have $\partial^p f(\bar{x}) = \partial f(\bar{x})$. Appealing now to Propositions 3.1(iii) and 3.6 confirms, respectively, that $\mathrm{d}^2 f(\bar{x},\bar{v})$ is a proper l.s.c. function and that (3.18) holds. The lower estimate (3.19) can be justified as [54, Theorem 13.14] in which this estimate was derived under condition (2.4). To obtain (3.20), observe first that the basic assumption (H1) yields

$$w \in K_f(\bar{x},\bar{v}) \iff \nabla F(\bar{x})w \in K_g(F(\bar{x}),y) \tag{3.21}$$

for every $y \in \Lambda(\bar{x},\bar{v})$. Pick $w \in K_f(\bar{x},\bar{v})$. Since g is parabolically epi-differentiable at $F(\bar{x})$ for $\nabla F(\bar{x})w$ due to (H2), Theorem 3.17(iii) implies that f is parabolically epi-differentiable at $\bar{x}$ for w, and so $\operatorname{dom} \mathrm{d}^2 f(\bar{x})(w \mid \cdot) \neq \emptyset$. This combined with (3.14) and (3.10) results in (3.20) and hence completes the proof. □

While looking simple, the above result carries important information by which we can achieve a chain rule for the second subderivative. To do so, we should look for conditions under which the lower and upper estimates (3.19) and (3.20), respectively, coincide. This motivates us to consider the unconstrained optimization problem

$$\min_{z\in\mathbb{X}} \ -\langle z,\bar{v}\rangle + \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \mid \nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w,w)). \tag{3.22}$$

When the basic assumptions (H1)-(H3) are satisfied, (3.22) is a convex optimization problem for any $w \in K_f(\bar{x}, \bar{v})$. Using Proposition 3.19 allows us to obtain the dual problem of (3.22) and then examine whether their optimal values coincide. We pursue this goal in the following result.

Theorem 3.22 (duality relationships) *Let $f : \mathbb{X} \to \overline{\mathbb{R}}$ have the composite representation* (3.3) *at $\bar{x} \in \operatorname{dom} f$, $\bar{v} \in \partial^- f(\bar{x})$, and let the basic assumptions* (H1)-(H3) *hold for f at $(\bar{x}, \bar{v})$. Then for each $w \in K_f(\bar{x}, \bar{v})$, the following assertions are satisfied:*

(i) *the dual problem of* (3.22) *is given by*

$$\max_{y \in \mathbb{Y}} \ \langle y, \nabla^2 F(\bar{x})(w, w)\rangle + \mathrm{d}^2 g(F(\bar{x}), y)(\nabla F(\bar{x}) w) \quad \textit{subject to} \ \ y \in \Lambda(\bar{x}, \bar{v}); \tag{3.23}$$

(ii) *the optimal values of the primal and dual problems* (3.22) *and* (3.23)*, respectively, are finite and coincide; moreover, we have $\Lambda(\bar{x}, \bar{v}, w) \cap (\tau \mathbb{B}) \neq \emptyset$, where $\Lambda(\bar{x}, \bar{v}, w)$ stands for the set of optimal solutions to the dual problem* (3.23) *and where*

$$\tau := \kappa \ell \|\nabla F(\bar{x})\| + \kappa \|\bar{v}\| + \ell \tag{3.24}$$

with ℓ and κ taken from (3.3) *and* (3.5)*, respectively.*

Proof.

Pick $w \in K_f(\bar{x}, \bar{v})$ and observe from (4.13) that $\nabla F(\bar{x}) w \in K_g(F(\bar{x}), y)$ for all $y \in \Lambda(\bar{x}, \bar{v})$. This together with Proposition 3.19 ensures that the parabolic subderivative of g at $F(\bar{x})$ for $\nabla F(\bar{x}) w$ is a proper, l.s.c., and convex function. Using this combined with [54, Example 11.41] and (3.17) tells us that the dual problem of (3.22) is

$$\max_{y \in \mathbb{Y}} \ \langle y, \nabla^2 F(\bar{x})(w, w)\rangle + \mathrm{d}^2 g(F(\bar{x}), y)(\nabla F(\bar{x}) w) \quad \text{subject to} \ \ y \in \Lambda(\bar{x}, \bar{v}) \cap \mathcal{D},$$

where $\mathcal{D} := \{y \in \mathbb{Y} |\, \mathrm{d} g(F(\bar{x}))(\nabla F(\bar{x}) w) = \langle y, \nabla F(\bar{x}) w\rangle\}$. Since $\nabla F(\bar{x}) w \in K_g(F(\bar{x}), y)$ for all $y \in \Lambda(\bar{x},\bar{v})$, we obtain

$$y \in \Lambda(\bar{x},\bar{v}) \cap \mathcal{D} \iff y \in \Lambda(\bar{x},\bar{v}).$$

Combining these confirms that the dual problem of (3.22) is equivalent to (3.23) and thus finishes the proof of (i). To prove (ii), consider the optimal value function $\vartheta : \mathbb{Y} \to [-\infty,\infty]$, defined by

$$\vartheta(p) = \inf_{z\in\mathbb{X}} \big\{ -\langle \bar{v}, z\rangle + \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \,|\, \nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w,w) + p)\big\}, \quad p \in \mathbb{Y}. \tag{3.25}$$

We proceed with the following claim:

Claim. *We have $\partial\vartheta(0) \neq \emptyset$.*

To justify the claim, we first need to show $\vartheta(0) \in \mathbb{R}$. To do so, observe that $\bar{v} \in \partial f(\bar{x}) = \partial^p f(\bar{x})$ due to Proposition 3.16(ii). Thus, it follows from Proposition 3.1(iii) and (3.20) that there is a constant $r \in \mathbb{R}_+$ such that for any $w \in K_f(\bar{x},\bar{v})$ we have

$$-r\|w\|^2 \le \mathrm{d}^2 f(\bar{x},\bar{v})(w) \le \vartheta(0) < \infty,$$

which in turn implies that $\vartheta(0) \in \mathbb{R}$. Next, we are going to show that

$$\vartheta(p) \ge \vartheta(0) - \tau\|p\| \quad \text{for all } \ p \in \mathbb{X}, \tag{3.26}$$

where τ is taken from (3.24). To this end, take $(p,z) \in \mathbb{Y}\times\mathbb{X}$ such that

$$u_p := \nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w,w) + p \in \operatorname{dom} \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \,|\, \cdot).$$

By (3.11), we get $u_p \in T^2_{\operatorname{dom} g}(F(\bar{x}), \nabla F(\bar{x})w)$. Define now the set-valued mapping $S_w : \mathbb{Y} \rightrightarrows \mathbb{X}$ by

$$S_w(p) := \big\{z \in \mathbb{X} |\ \nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w,w) + p \in T^2_{\operatorname{dom} g}(F(\bar{x}), \nabla F(\bar{x})w)\big\}, \ \ p \in \mathbb{Y}.$$

So, we get $z \in S_w(p)$. It was shown that in Theorem (3.14) that the mapping S_w enjoys the uniform outer Lipschitzian property at 0 with constant κ taken from (3.5), namely for every

$p \in \mathbb{Y}$ we have

$$S_w(p) \subset S_w(0) + \kappa \|p\| I\!B.$$

This combined with $z \in S_w(p)$ results in the existence of $z_0 \in S_w(0)$ and $b \in I\!B$ such that $z = z_0 + \kappa\|p\|b$. It follows from (3.11) and $z_0 \in S_w(0)$ that

$$\nabla F(\bar{x})z_0 + \nabla^2 F(\bar{x})(w,w) \in \operatorname{dom} \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \,|\, \cdot).$$

Since we have

$$u_p - \big(\nabla F(\bar{x})z_0 + \nabla^2 F(\bar{x})(w,w)\big) = p + \kappa\|p\|\nabla F(\bar{x})b,$$

and since the parabolic subderivative $\mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \,|\, \cdot)$ is Lipschitz continuous relative to its domain due to Proposition 3.18, we get the relationships

$$\begin{aligned} -\langle \bar{v}, z\rangle + \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \,|\, u_p) \ &\geq\ -\langle \bar{v}, z_0\rangle + \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \,|\, \nabla F(\bar{x})z_0 + \nabla^2 F(\bar{x})(w,w)) \\ &\quad\ -\ell \,\|p + \kappa \,\|p\|\nabla F(\bar{x})b\,\| - \kappa\|p\|\langle \bar{v}, b\rangle \\ &\geq\ \vartheta(0) - \big(\ell\kappa\|\nabla F(\bar{x})\| + \kappa\|\bar{v}\| + \ell\big)\|p\|, \end{aligned}$$

which together with (3.24) justify (3.26). Remember that the parabolic subderivative of g at $F(\bar{x})$ for $\nabla F(\bar{x})w$ is a proper and convex function. This implies that the function

$$(z,p) \mapsto -\langle \bar{v}, z\rangle + \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w \,|\, \nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w,w) + p)$$

is convex on $\mathbb{X} \times \mathbb{Y}$. Using this together with [54, Proposition 2.22] tells us that ϑ is a convex function on $\mathbb{Y}$. Thus, we conclude from (3.26) and [39, Proposition 5.1] that there exists a subgradient $\bar{y}$ of ϑ at 0 such that

$$\bar{y} \in \partial\vartheta(0) \cap (\tau I\!B), \tag{3.27}$$

which completes the proof of the claim.

Employing now (3.27) and [4, Theorem 2.142] confirms that the optimal values of the primal and dual problems (3.22) and (3.23), respectively, coincide and that

$$\Lambda(\bar{x},\bar{v},w)=\partial\vartheta(0).$$

This together with (3.27) justifies (ii) and hence completes the proof. $\square$

The above theorem extends the recent results obtained in [39, Propositions 5.4 & 5.5] for constraint sets, namely when the function g in (3.3) is the indicator function of a closed convex set. We should add here that for constraint sets, the dual problem (3.23) can be obtained via elementary arguments. However, for the composite form (3.3) a similar result requires using rather advanced theory of epi-convergence.

Remark 3.23 (duality relationship under metric regularity) In the framework of Theorem 3.22, we want to show that replacing assumption (H1) with the strictly stronger qualification condition (2.4) allows us not only to drop the imposed Lipschitz continuity of g from (3.3) but also to simplify the proof of Theorem 3.22. To this end, let $w\in K_f(\bar{x},\bar{v})$ and define the function

$$\psi(u):=\mathrm{d}^2g(F(\bar{x}))(\nabla F(\bar{x})w\,|\,u),\quad u\in\mathbb{X}.$$

By Proposition 3.19, ψ is a proper, l.s.c., and convex function. Employing [54, Proposition 13.12] tells us that

$$T_{\operatorname{epi}g}(p)+T^2_{\operatorname{epi}g}\big(p,q\big)\subset T^2_{\operatorname{epi}g}\big(p,q\big)=\operatorname{epi}\psi,\tag{3.28}$$

where $p:=\big(F(\bar{x}),g(F(\bar{x}))\big)$ and $q:=\big(\nabla F(\bar{x})w,\mathrm{d}g(F(\bar{x}))(\nabla F(\bar{x})w)\big)$ and where the equality in the right side comes from [54, Example 13.62(b)]. We are going to show that the validity of (2.4) yields

$$N_{\operatorname{dom}\psi}(u)\cap\ker\nabla F(\bar{x})^*=\{0\}\tag{3.29}$$

for any $u\in\operatorname{dom}\psi$. To this end, pick $u\in\operatorname{dom}\psi$ and conclude from (3.28) and [54, Exercise 6.44] that

$$N_{\operatorname{epi}\psi}\big(u,\psi(u)\big)\subset N_{T_{\operatorname{epi}g}(p)}(0)\cap N_{\operatorname{epi}\psi}\big(u,\psi(u)\big)=N_{\operatorname{epi}g}(p)\cap N_{\operatorname{epi}\psi}\big(u,\psi(u)\big).$$

This together with (2.4) and the relationship $N_{\operatorname{dom}\psi}(u)=\partial^\infty\psi(u)$ stemming from the convexity of ψ confirms the validity of (3.29). Appealing now to [4, Theorem 2.165] gives another proof of Theorem 3.22 when assumption (H1) therein is replaced with the strictly stronger constraint qualification (2.4).

The established duality relationships in Theorem 3.22 open the door to derive a chain rule for parabolically regular functions and to find an exact chain rule for the second subderivative of the composite function (3.3) under our basic assumptions.

Theorem 3.24 (chain rule for parabolic regularity) *Let $f:\mathbb{X}\to\overline{\mathbb{R}}$ have the composite representation* (3.3) *at $\bar{x}\in\operatorname{dom}f$, $\bar{v}\in\partial^- f(\bar{x})$, and let the basic assumptions* (H1)-(H3) *hold for f at $(\bar{x},\bar{v})$. Then f is parabolically regular at $\bar{x}$ for $\bar{v}$. Furthermore, for every $w\in\mathbb{X}$, the second subderivative of f at $\bar{x}$ for $\bar{v}$ is calculated by*

$$\begin{aligned}\mathrm{d}^2 f(\bar{x},\bar{v})(w) &= \max_{y\in\Lambda(\bar{x},\bar{v})}\big\{\langle y,\nabla^2F(\bar{x})(w,w)\rangle+\mathrm{d}^2 g(F(\bar{x}),y)(\nabla F(\bar{x})w)\big\} \qquad (3.30)\\ &= \max_{y\in\Lambda(\bar{x},\bar{v})\cap(\tau\mathbb{B})}\big\{\langle y,\nabla^2F(\bar{x})(w,w)\rangle+\mathrm{d}^2 g(F(\bar{x}),y)(\nabla F(\bar{x})w)\big\},\end{aligned}$$

where τ is taken from (3.24).

Proof.

It was recently observed in [38, Corollary 3.7] that the Lagrange multiplier set $\Lambda(\bar{x},\bar{v})$ enjoys the following property:

$$\Lambda(\bar{x},\bar{v})\,\cap\,(\tau\mathbb{B})\neq\emptyset. \qquad (3.31)$$

Take $w\in K_f(\bar{x},\bar{v})$. By (3.19) and Theorem 3.22(ii), we obtain

$$\max_{y\in\Lambda(\bar{x},\bar{v})}\big\{\langle y,\nabla^2F(\bar{x})(w,w)\rangle+\mathrm{d}^2 g(F(\bar{x}),y)(\nabla F(\bar{x})w)\big\}\le \mathrm{d}^2 f(\bar{x},\bar{v})(w). \qquad (3.32)$$

On the other hand, using (3.14), (3.10), and Theorem 3.22(ii), respectively, gives us the inequalities

$$\begin{aligned}
\mathrm{d}^2 f(\bar{x},\bar{v})(w) &\le \inf_{z\in\mathbb{X}}\big\{\mathrm{d}^2 f(\bar{x})(w\,|\,z) - \langle z,\bar{v}\rangle\big\}\\
&= \inf_{z\in\mathbb{X}}\big\{-\langle z,\bar{v}\rangle + \mathrm{d}^2 g(F(\bar{x}))(\nabla F(\bar{x})w\,|\,\nabla F(\bar{x})z + \nabla^2 F(\bar{x})(w,w))\big\}\\
&= \max_{y\in\Lambda(\bar{x},\bar{v})\cap(\tau I\!B)}\big\{\langle y,\nabla^2 F(\bar{x})(w,w)\rangle + \mathrm{d}^2 g(F(\bar{x}),y)(\nabla F(\bar{x})w)\big\}.
\end{aligned}$$

These combined with (3.32) ensure that the claimed second subderivative formulas for f at $\bar{x}$ for $\bar{v}$ hold for any $w\in K_f(\bar{x},\bar{v})$ and that

$$\mathrm{d}^2 f(\bar{x},\bar{v})(w) = \inf_{z\in\mathbb{X}}\big\{\mathrm{d}^2 f(\bar{x})(w\,|\,z) - \langle z,\bar{v}\rangle\big\}\quad\text{for all}\;\; w\in K_f(\bar{x},\bar{v}).$$

Appealing now to Proposition 3.8, we conclude that f is parabolically regular at $\bar{x}$ for $\bar{v}$.

What remains is to validate the second subderivative formulas for $w\notin K_f(\bar{x},\bar{v})$. It follows from Theorem 3.17(iii) that f is parabolically epi-differentiable at $\bar{x}$ for every $w\in K_f(\bar{x},\bar{v})$ and thus $\operatorname{dom}\mathrm{d}^2 f(\bar{x})(w\,|\,\cdot)\neq\emptyset$ for every $w\in K_f(\bar{x},\bar{v})$. So, by Proposition 3.6 we have $\operatorname{dom}\mathrm{d}^2 f(\bar{x},\bar{v}) = K_f(\bar{x},\bar{v})$. Since the second subderivative $\mathrm{d}^2 f(\bar{x},\bar{v})$ is a proper function, we obtain $\mathrm{d}^2 f(\bar{x},\bar{v})(w)=\infty$ for all $w\notin K_f(\bar{x},\bar{v})$. On the other hand, we understand from (4.13) that $w\notin K_f(\bar{x},\bar{v})$ amounts to $\nabla F(\bar{x})w\notin K_g(F(\bar{x}),y)$ for every $y\in\Lambda(\bar{x},\bar{v})$. Combining the basic assumption (H2) and Proposition 3.6 tells us that for every $y\in\Lambda(\bar{x},\bar{v})$ we have $\mathrm{d}^2 g(F(\bar{x}),y)(\nabla F(\bar{x})w)=\infty$ whenever $w\notin K_f(\bar{x},\bar{v})$. This together with (3.31) confirms that both sides in (3.30) are ∞ for every $w\notin K_f(\bar{x},\bar{v})$ and thus the claimed formulas for the second subderivative of f hold for this case. This completes the proof. $\square$

A chain rule for parabolic regularity of the composite function (3.3), where g is not necessarily locally Lipschitz continuous relative to its domain, was established in [4, Proposition 3.104]. The assumptions utilized in the latter result were stronger than those used in Theorem 3.24. Indeed, [4, Proposition 3.104] assumes that g is second-order regular in the sense of [4, Definition 3.93] and the basic constraint qualification (2.4) is satisfied and uses a different approach to derive this result. When g is a convex piecewise linear-quadratic, parabolic regularity of the composite function (3.3) was established in [54, Theorem 13.67] under the stronger condition (2.4). Theorem 3.24 covers the aforementioned results and shows that we can achieve a similar conclusion under the significantly weaker condition (3.5).

As an immediate consequence of the above theorem, we can easily guarantee the twice epi-differentiability of the composite form (3.3) under our basic assumptions.

Corollary 3.25 (chain rule for twice epi-differentiability) *Let the function f from* (3.3) *satisfy all the assumptions of Theorem* 3.24*. Then f is twice epi-differentiable at $\bar{x}$ for $\bar{v}$.*

Proof.

By Theorem 3.17(iii), f is parabolically epi-differentiable at $\bar{x}$ for every $w \in K_f(\bar{x}, \bar{v})$. Employing now Theorems 3.24 and 3.10 implies that f is twice epi-differentiable at $\bar{x}$ for $\bar{v}$.

□

Remark 3.26 (discussion on twice epi-differentiability) Corollary 3.25 provides a far-going extension of the available results for the twice epi-differentiability of extended-real-valued functions. To elaborate more, suppose that $f : \mathbb{X} \to \overline{\mathbb{R}}$ has a composite form (3.3) at $\bar{x} \in \operatorname{dom} f$. Then the following observations hold:

(a) If $\mathbb{X} = \mathbb{R}^n$, $\mathbb{Y} = \mathbb{R}^m$, and g in (3.3) is convex piecewise linear-quadratic, then Rockafellar proved in [49] that under the fulfillment of the basic constraint qualification (2.4), f is twice epi-differentiable. This result was improved recently in [38, Theorem 5.2], where it was shown that using the strictly weaker condition (3.5) in the Rockafellar's framework [49] suffices to ensure the twice epi-differentiability of f. Taking into account Example 3.20(a) tells us both these results can be derived from Corollary 3.25.

(b) If $\mathbb{X} = \mathbb{R}^n$, $\mathbb{Y} = \mathcal{S}^m$, and g is either the maximum eigenvalue function $\lambda_{\max}$ from (3.11) or the function σ_i from (3.12), then we fall into the framework considered by Turki in [57, Theorems 2.3 & 2.5] in which he justified the twice epi-differentiability of f. Since in this framework we have $\operatorname{dom} g = \mathcal{S}^m$, both conditions (2.4) and (3.5) are automatically satisfied. By Example 3.20(b), the twice epi-differentiability of f can be deduced from Corollary 3.25.

(c) If $\mathbb{X} = \mathbb{R}^n$, $\mathbb{Y} = \mathbb{R}^m$, and $g = \delta_\Theta$ with the closed convex set Θ taken from Example 3.20(c), we fall into the framework considered in [39]. In this case, Corollary 3.25 can cover the twice epi-differentiability of f obtained in [39, Corollary 5.11].

(d) If $\mathbb{X} = \mathbb{R}^n$, $\mathbb{Y} = \mathbb{R}^m$, and g is a proper, convex, l.s.c., and positively homogeneous, then we fall into the framework, considered by Shapiro in [55]. In this case, the composite form (3.3) is called *decomposable*; see [34, 55] for more detail about this class of extended-real-valued functions. It was proven in [34, Lemma 5.3.27] that for this case of g, the composite form (3.3) is twice epi-differentiable if it is convex and if the nondegeneracy condition for this setting holds; see [34, Definition 5.3.1] for the definition of this condition. In this framework, by the positive homogeneity of g and $F(\bar{x}) = 0$, coming from [34, Definition 5.3.1], we can easily show that g is parabolically regular. Moreover the assumed nondegeneracy condition in [34, Lemma 5.3.27] yields the validity of condition (2.4). As pointed out in Remark 3.23, the Lipschitz continuity of g in the composite form (3.3) can be relaxed when condition (2.4) is satisfied. Since the nondegeneracy condition implies that the set of Lagrange multipliers $\Lambda(\bar{x}, \bar{v})$ is a singleton, we can use estimates (3.19) and (3.20) to justify parabolic regularity of the composite form (3.3) in the framework of [34]. This together with Corollary 3.25 allows to recover [34, Lemma 5.3.27]. Furthermore, we can drop the convexity of the composite form (3.3), assumed in [34].

3.5 Second-Order Optimality Conditions for Composite Problems

In this section, we focus mainly on obtaining second-order optimality conditions for the composite problem (1.1), where $\varphi : \mathbb{X} \to \mathbb{R}$ and $F : \mathbb{X} \to \mathbb{Y}$ are twice differentiable and the function $g : \mathbb{Y} \to \overline{\mathbb{R}}$ is an l.s.c. convex function that is locally Lipschitz continuous relative to its domain. The latter means that for any $y \in \operatorname{dom} g$, the function g is Lipschitz continuous around y relative to its domain. Important examples of constrained and composite optimization problems can be achieved when g is one of the functions considered in Example 3.20. For any pair $(x, y) \in \mathbb{X} \times \mathbb{Y}$, the Lagrangian associated with the composite problem (1.1) is defined by

$$L(x, y) = \phi(x) + \langle F(x), y\rangle - g^*(y),$$

where g^* is the Fenchel conjugate of the convex function g. We begin with the following result in which we collect second-order optimality conditions for (1.1) when our basic assumptions are satisfied. Recall that a point $\bar{x} \in \mathbb{X}$ is called a feasible solution to the composite problem (1.1) if we have $F(\bar{x}) \in \operatorname{dom} g$.

Theorem 3.27 (second-order optimality conditions) *Let $\bar{x}$ be a feasible solution to problem* (1.1) *and let $f := g \circ F$ and $\bar{v} := -\nabla\phi(\bar{x}) \in \partial^- f(\bar{x})$ with φ, g, and F taken from* (1.1)*. Assume that the basic assumptions* (H1)-(H3) *hold for f at $(\bar{x}, \bar{v})$. Then the following second-order optimality conditions for the composite problem* (1.1) *are satisfied:*

(i) *if $\bar{x}$ is a local minimum of* (1.1)*, then the second-order necessary condition*

$$\max_{y \in \Lambda(\bar{x}, \bar{v})} \big\{ \langle \nabla^2_{xx} L(\bar{x}, y) w, w\rangle + \mathrm{d}^2 g(F(\bar{x}), y)(\nabla F(\bar{x}) w) \big\} \geq 0$$

holds for all $w \in K_f(\bar{x}, \bar{v})$;

(ii) *the validity of the second-order condition*

$$\max_{y\in\Lambda(\bar{x},\bar{v})}\left\{\langle\nabla^2_{xx}L(\bar{x},y)w,w\rangle+\mathrm{d}^2g(F(\bar{x}),y)(\nabla F(\bar{x})w)\right\}>0\quad \textit{for all}\quad w\in K_f(\bar{x},\bar{v})\setminus\{0\} \tag{3.33}$$

amounts to the existence of constants $\ell>0$ and $\varepsilon>0$ such that the second-order growth condition

$$\psi(x)\geq\psi(\bar{x})+\tfrac{\ell}{2}\|x-\bar{x}\|^2\quad \textit{for all}\ \ x\in I\!B_\varepsilon(\bar{x}) \tag{3.34}$$

holds, where $\psi:=\phi+g\circ F$.

Proof.

To justify (i), note that since $\bar{x}$ is a local minimum of (1.1), it is a local minimum of $\psi=\phi+f$. Moreover, $-\nabla\phi(\bar{x})\in\partial^- f(\bar{x})$ amounts to $0\in\partial\psi(\bar{x})$. Thus, by definition, we arrive at $\mathrm{d}^2\psi(\bar{x},0)(w)\geq 0$ for all $w\in\mathbb{X}$. Since φ is twice differentiable at $\bar{x}$, we obtain the following sum rule for the second subderivatives:

$$\mathrm{d}^2\psi(\bar{x},0)(w)=\langle\nabla^2\phi(\bar{x})w,w\rangle+\mathrm{d}^2f(\bar{x},\bar{v})(w)\quad\text{for all}\ \ w\in\mathbb{X}. \tag{3.35}$$

Combing these with the chain rule (3.30) proves (i).

Turing now to (ii), we infer from [54, Theorem 13.24(c)] that $\mathrm{d}^2\psi(\bar{x},0)(w)>0$ for all $w\in\mathbb{X}\setminus\{0\}$ amounts to the existence of some constants $\ell>0$ and $\varepsilon>0$ for which the second-order growth condition (3.34) holds. Remember from (3.18) and (3.35) that

$$\operatorname{dom}\mathrm{d}^2\psi(\bar{x},0)=\operatorname{dom}\mathrm{d}^2f(\bar{x},\bar{v})=K_f(\bar{x},\bar{v}). \tag{3.36}$$

Using these, the chain rule (3.30), and the sum rule (3.35) proves the claimed equivalence in (ii) and thus finishes the proof. □

Now we consider the constrained optimization problem (1.2)

$$\text{minimiz}\ \ \phi(x)\quad\text{subject to}\ \ F(x)\in\Theta, \tag{3.37}$$

as a particular case of composite optimization when $g = \delta_\Theta$. we assume ϕ and F are countinuouly twice differentiable around $\bar{x}$ and Θ is closed convex set. We set $\Omega := \{x|\ F(x) \in \Theta\}$ then for $\bar{v} \in N_\Omega(\bar{x})$ It is easy to observe that

$$K_\Omega(\bar{x}, \bar{v}) := K_{\delta_\Omega}(\bar{x}, \bar{v}) = T_\Omega(\bar{x}) \cap \{\bar{v}\}^\perp.$$

Note if the set-valued mapping $x \to F(x) - \Theta$ is metrically subregular at $\bar{x}$ then by Proposition 3.13 we have $N_\Omega(\bar{x}) = \nabla F(\bar{x})^* N_\Theta(F(\bar{x}))$ and for all $\lambda \in \Lambda(\bar{x}, \bar{v})$ we have

$$w \in K_\Omega(\bar{x}, \bar{v}) \quad \text{if and only if} \quad \nabla F(\bar{x})w \in K_\Theta(F(\bar{x}), \lambda).$$

In order to facilitate our presentation in this section, we list our basic assumptions that we often impose on the convex set Θ from (4.1) at the point $\bar{x}$ with $F(\bar{x}) \in \Theta$:

(H1) for every $\lambda \in \Lambda(\bar{x})$, the set Θ is parabolically derivable at $F(\bar{x})$ for all vectors of the form $\nabla F(\bar{x})w$ in $K_\Theta(F(\bar{x}), \lambda)$;

(H2) the set Θ is parabolically regular at $F(\bar{x})$ for every $\lambda \in \Lambda(\bar{x}, \bar{v})$.

It was shown in [39] that both assumptions (H1) and (H2) hold for important convex sets appearing in constrained optimization problems including polyhedral convex sets, the second-order cone, and the cone of symmetric and positive semidefinite matrices. Indeed, it was proven in [39, Proposition 5.14] that any $\mathcal{C}^2$-cone reducible is satisfying these assumptions.

Imposing these assumptions on the set Θ opens the door for deriving second-order optimality conditions for the constrained optimization problem (4.1). In the following Corollary we derive the second-order optimality condition for the constrained problem (1.2). The Larangian function associated to the problem is defined by

$$L(x, \lambda) = \phi(x) + \langle \lambda, F(x) \rangle.$$

Corollary 3.28 (second-order optimality conditions of constrained optimization) *Let the basic assumptions* (H1) *and* (H2) *hold for* Θ *from* (4.1) *and* $\bar{v} = -\nabla\varphi(\bar{x})$. *Assume further that the mapping* $x \mapsto F(x) - \Theta$ *is metrically subregular at* $(\bar{x}, 0)$. *Set* $\Omega := \{x \in \mathbb{R}^n |\, F(x) \in \Theta\}$. *Then we have the following second-order optimality conditions for the constrained problem* (4.1)*:*

(i) *If* $\bar{x}$ *is a local minimum of* (4.1)*, then the second-order necessary condition*

$$\max_{\lambda \in \Lambda(\bar{x})} \left\{ \langle \nabla^2_{xx} L(\bar{x}, \lambda) w, w \rangle + \mathrm{d}^2 \delta_\Theta(F(\bar{x}), \lambda)(\nabla F(\bar{x}) w) \right\} \geq 0$$

holds for all $w \in K_\Theta(\bar{x}, \bar{v})$.

(ii) *The validity of the second-order condition for all* $w \in K_\Theta(\bar{x}, \bar{v}) \setminus \{0\}$

$$\max_{\lambda \in \Lambda(\bar{x})} \left\{ \langle \nabla^2_{xx} L(\bar{x}, \lambda) w, w \rangle + \mathrm{d}^2 \delta_\Theta(F(\bar{x}), \lambda)(\nabla F(\bar{x}) w) \right\} > 0$$

amounts to the existence of constants $\ell > 0$ *and* $\varepsilon > 0$ *such that the growth condition*

$$\phi(x) \geq \phi(\bar{x}) + \frac{\ell}{2} \|x - \bar{x}\|^2 \quad \text{for all } x \in I\!B_\varepsilon(\bar{x}) \cap \Omega$$

holds.

Proof.

In Theorem 3.27 let $g = \delta_\Theta$. □

An important class of constrained optimization is nonlinear programming (2.33):

$$\begin{aligned} \text{minimize} \quad & \phi(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, .., r \\ & f_i(x) = 0, \quad i = r+1, ..., m \end{aligned} \tag{3.38}$$

where $\phi : \mathbb{R}^n \to \mathbb{R}$ and each $f_i : \mathbb{R}^n \to \mathbb{R}$ is continuously twice differentiable. Denote by Ω the feasible solution of the above nonlinear programming. $\bar{x} \in \Omega$ is called stationary point of the above problem if there exists a vector $\lambda = (\lambda_1, , , , , \lambda_m)$ - called Lagrangian multiplier

- such that

$$\nabla_x L(\bar{x},\lambda)=0 \quad , \quad \lambda f_i(\bar{x})=0, \quad \lambda_i \geq 0 \quad \text{such that} \quad i=1,...,r \tag{3.39}$$

where $L(x,\lambda) = \phi(x)+\sum_{i=1}^m \lambda_i f_i(x)$. Denote $\Lambda(\bar{x})$ The set of all Lagrangian multipliers associated to the point $\bar{x}\in\Omega$. The critical cone at $\bar{x}$ for this problems reduces to

$$K_\Omega(\bar{x}) := \{w\in\mathbb{R}^n |\, \langle\nabla f_i(\bar{x}),w\rangle \leq 0,\ i\in I(\bar{x}) \quad \langle\nabla f(\bar{x}),w\rangle = 0 \,\text{for}\, i=r{+}1,...,m \quad \langle w,\nabla\phi(\bar{x})\rangle = 0\}$$

where $I(\bar{x})$ is the set of active index associated to $\bar{x}\in\Omega$.

Corollary 3.29 (second-order optimality conditions of nonlinear programming) *Let $\bar{x}$ be satisfying* (2.35) *for the above nonlinear programming. Assume the metric subregularity constraint qualification* (2.34)*:*

(i) *If $\bar{x}$ is a local minimum, then for $w\in K_\Omega(\bar{x})$ we have*

$$\max_{\lambda\in\Lambda(\bar{x})} \left\{\langle\nabla^2_{xx}L(\bar{x},\lambda)w,w\rangle\right\} \geq 0$$

holds.

(ii) *If the above inequality holds strictly for all $w\in K_\Omega(\bar{x})$, then $\bar{x}$ is a local minimizer of the nonlinear programming. Actually the latter case amounts to the existence of constants $\ell>0$ and $\varepsilon>0$ such that the quadratic growth condition*

$$\phi(x)\geq\phi(\bar{x})+\frac{\ell}{2}\|x-\bar{x}\|^2 \quad \textit{for all } x\in I\!B_\varepsilon(\bar{x})\cap\Omega$$

Proof.

In Corollary 3.28 set $\Theta := \mathbb{R}^r_- \times \{0\}^{m-r}$ and $F := (f_1,...,f_2)$. Note that $\bar{x}$ satisfies (2.35) thus it is a staitiory point which means

$$\Lambda(\bar{x}) = \{\lambda\in\mathbb{R}^m |\ \nabla_x L(\bar{x},\lambda)=0, \quad \lambda\in N_\Theta(F(\bar{x}))\} \neq \emptyset.$$

□

If we impose the stronger constraint qualification (LICQ) in Corollary 3.29, we can expect a stronger results. The next corollary addresses this issue. Recall that the point $\bar{x} \in \Omega$ satisfies the *linearly independent constraint qualification* (LICQ) for the above nonlinear programming if the set

$$\{\nabla f_i(\bar{x}) \text{ for } i \in I(\bar{x}), \quad \nabla f_i(\bar{x}) \text{ for } i = r+1, ..., m\}$$

is linearly independent. Also recall that under any constraint qualification the tangent cone at $\bar{x}$ to the feasible solution of the problem nonlinear programming can be written as follow

$$T_\Omega(\bar{x}) = \{w \in \mathbb{R}^n \mid \nabla\langle f_i(\bar{x}), w\rangle \leq 0 \text{ for } i \in I(\bar{x}), \quad \langle f_i(\bar{x}), w\rangle = 0 \text{ for } i = r+1, ..., m\}.$$

Corollary 3.30 (second-order optimality conditions under LICQ) *Let $\bar{x}$ be satisfying (2.35) for the above nonlinear programming with the Lagrangian multiplier $\bar{\lambda} = (\bar{\lambda}_1, ..., \bar{\lambda}_m)$. Assume the linearly independent constraint qualification holds at $\bar{x}$:*

(i) *If $\bar{x}$ is a local minimum, then for $w \in T_\Omega(\bar{x})$ we have*

$$\langle \nabla^2_{xx} L(\bar{x}, \bar{\lambda})w, w\rangle \geq 0$$

holds.

(ii) *If the above inequality holds strictly for all $w \in T_\Omega(\bar{x})$, then $\bar{x}$ is a local minimizer of the nonlinear programming. Actually the latter case amounts to the existence of constants $\ell > 0$ and $\varepsilon > 0$ such that the quadratic growth condition*

$$\phi(x) \geq \phi(\bar{x}) + \frac{\ell}{2}\|x - \bar{x}\|^2 \quad \textit{for all } x \in I\!B_\varepsilon(\bar{x}) \cap \Omega.$$

Proof.

Note that under LICQ the set of Lagrangian multipiler is unique, thus we have $\Lambda(\bar{x}) = \{\bar{\lambda}\}$. Furthermore, it is not difficult to check that $T_\Omega(\bar{x}) \subseteq \{\nabla\phi(\bar{x})\}^\perp$. Therefore the results follows from Corollary 3.29. $\square$

Remark 3.31 (discussion on second-order optimality conditions) The second-order optimality conditions for composite problems were established in [4, Theorems 3.108 & 3.109] for (1.1) by expressing (1.1) equivalently as a constrained problem and then appealing to the theory of second-order optimality conditions for the latter class of problems. While not assuming that g is locally Lipschitz continuous relative to its domain, theses results were established under condition (2.4) and the second-order regularity in the sense of [4, Definition 3.93] that are strictly stronger than condition (3.5) and the parabolic regularity, respectively, we imposed in Theorem 3.27. Another major difference is that we require that g be parabolically epi-differentiable (assumption (H2)), which was not assumed in [4]. This assumption plays an important role in our developments and has two important consequences: 1) it makes the parabolic subderivative be a *convex* function and help us obtain a precise formula for the Fenchel conjugate of the parabolic subderivative in our framework; 2) it allows to establish the equivalence between (4.3) and the growth condition in Theorem 3.27. These facts were not achieved in [4]; indeed, [4, Theorem 3.109] was written in terms of the conjugate of the parabolic subderivative and only states that condition (4.3) implies the growth condition therein.As discussed in Remark 3.23, if we replace condition (3.5) with the stronger condition (2.4), the imposed Lipschitz continuity of g can be relaxed in our developments. It is worth mentioning that the imposed Lipschitz continuity of g relative to its domain, utilized in this paper, does not seem to be restrictive and allows us to provide an umbrella under which second-order variational analysis for composite problems can be carried out under condition (2.4) in the same level of perfection as those for constrained problems. We believe that if we strengthen condition (2.4) to the metric subregularity of the epigraphical mapping $(x, \alpha) \mapsto (F(x), \alpha) - \operatorname{epi} g$, the imposed Lipschitz continuity of g can be relaxed in our developments.

Cominetti [10, Theorem 5.1] established second-order optimality conditions for the composite problem (1.1) similar to Theorem 3.27 without making a connection between (4.3) and the growth condition (3.34). As mentioned in our discussion after Example 3.20, the results in [10] were established under condition (2.4) and a restrictive assumption on the second subderivative, which does not hold for important classes of composite problems. When we are in the framework of Remark 3.26(a), Theorem 3.27 was first achieved by Rockafellar in [51, Theorem 4.2] under condition (2.4) and was improved recently in [38, Theorem 6.2] by replacing the latter condition with (3.5). For the framework of Remark 3.26(b), the second-order optimality conditions from Theorem 3.27 were obtained in [57, Theorem 4.2]. Finally, if we are in the framework of Remark 3.26(c), Theorem 3.27 covers our recent developments in [39].

We end this section by obtaining a characterization of strong metric subregularity of the subgradient mapping of the objective function of the composite problem (1.1).

Theorem 3.32 (strong metric subregularity of the subgradient mappings) *Let $\bar{x}$ be a feasible solution to problem* (1.1) *and let $f := g \circ F$ and $\bar{v} := -\nabla\varphi(\bar{x}) \in \partial^- f(\bar{x})$ with φ, g, and F taken from* (1.1). *Assume that the basic assumptions* (H1)-(H3) *hold for f at $(\bar{x}, \bar{v})$ and that both φ and F are $\mathcal{C}^2$-smooth around $\bar{x}$. Then the following conditions are equivalent:*

- **(i)** *the point $\bar{x}$ is a local minimizer for $\psi = \varphi + f$ and the subgradient mapping $\partial\psi$ is strongly metrically subregular at $(\bar{x}, 0)$;*
- **(ii)** *the second-order sufficient condition* (4.3) *holds.*

Proof.

We conclude from (3.35) and (3.36) that (4.3) amounts to the fulfillment of the condition

$$\mathrm{d}^2\psi(\bar{x}, 0)(w) > 0 \quad \text{for all } \ w \in \mathbb{X} \setminus \{0\}. \tag{3.40}$$

If (i) holds, we conclude from the local optimality of $\bar{x}$ that $\mathrm{d}^2\psi(\bar{x}, 0)(w) \geq 0$ for all $w \in \mathbb{X}$.

Since (ii) is equivalent to (3.40), it suffices to show that there is no $w \in \mathbb{X} \setminus \{0\}$ such that $\mathrm{d}^2\psi(\bar{x},0)(w) = 0$. Suppose on the contrary that there exists $\bar{w} \in \mathbb{X} \setminus \{0\}$ satisfying the latter condition. This means that $\bar{w}$ is a minimizer for the problem

$$\text{minimize } \tfrac{1}{2}\mathrm{d}^2\psi(\bar{x},0)(w) \quad \text{subject to} \quad w \in \mathbb{X}.$$

Since both φ and F are $\mathcal{C}^2$-smooth around $\bar{x}$, we can show using similar arguments as Proposition 2.24 that ψ is prox-regular and subdifferentially continuousat $\bar{x}$ for 0, and we have $\partial^- \phi = \partial^{L-}\phi = \partial^p \phi$. These together with the Fermat stationary principle and (4.2) results in

$$0 \in \partial^{L-}\big(\tfrac{1}{2}\mathrm{d}^2\psi(\bar{x},0)\big)(\bar{w}) = D(\partial^-\psi)(\bar{x},0)(\bar{w}). \tag{3.41}$$

Since $\partial\psi$ is strongly metrically subregular at $(\bar{x},0)$, we deduce from (3.3) that $\bar{w} = 0$, a contradiction. This proves (ii).

To justify the opposite implication, assume that (ii) holds. According to Theorem 3.27(ii), $\bar{x}$ is a local minimizer for ψ. Pick now $w \in \mathbb{X}$ such that $0 \in D(\partial^-\psi)(\bar{x},0)(w)$. To obtain (i), we require by (3.3) to show that $w = 0$. Employing now (3.41) yields $0 \in \partial^{L-}\big(\tfrac{1}{2}\mathrm{d}^2\psi(\bar{x},0)\big)(w)$. This combined with [9, Lemma 3.7] confirms that $\mathrm{d}^2\psi(\bar{x},0)(w) = \langle 0, w\rangle = 0$. Remember that (ii) is equivalent to (3.40). Combining these results in $w = 0$ and thus proves (i). $\square$

The above result was first observed in [11, Theorem 4G.1] for a subclass of nonlinear programming problems and was extended in [9, Theorem 4.6] for $\mathcal{C}^2$-cone reducible constrained optimization problems and in [39, Theorem 9.2] for parabolically regular constrained optimization problems. The theory of the twice epi-differentiability, obtained in this paper, provides an easy path to achieve a similar result for the composite problem (1.1).

It is worth mentioning that similar characterizations as [39, Theorem 4.2] can be achieved for the KKT system of (1.1). Furthermore, Corollary 3.11 provides a systematic method to calculate proto-derivatives of subgradient mappings of functions enjoying the composite form (3.3), a path we will pursue in our future research.

CHAPTER 4 APPLICATIONS IN THE SQP METHOD

The main attention of this chapter is given to study of the strong metric subregularity of the KKT system of the constrained optimization problem

$$\min_{x\in\mathbb{R}^n} \quad \varphi(x) \quad \text{subject to} \quad F(x)\in\Theta, \tag{4.1}$$

where $\varphi\colon\mathbb{R}^n\to\mathbb{R}$ and $F\colon\mathbb{R}^n\to\mathbb{R}^m$ are twice continuously differentiable functions around reference points, where Θ is a closed and convex set in $\mathbb{R}^m$. Following Chapter 3 the second-order analysis of optimization problems requires certain second-order regularity condition on the constraint set Θ. We will assume in this chapter that the set Θ is parabolically regular; see the next section for the definition of such a set.

Our first goal in this chapter is to provide a systematic study of strong metric subregularity of the KKT system of (4.1) using the second-order analysis conducted for parabolically regular sets in chapter 3. It is well-known [12] that for nonlinear programs the latter property amounts to the uniqueness of Lagrange multipliers as well as the second-order sufficient condition. Later it was observed in [13] that a similar result for $\mathcal{C}^2$-cone reducible constrained problems, namely (4.1) with Θ being $\mathcal{C}^2$-cone reducible in the sense of [4, Definition 3.135], can be established if the uniqueness of Lagrange multipliers is replaced by the strong Robinson constraint qualification (see condition (4.9)). We show that the well-known result for NLPs can be retrieved for parabolically regular constrained problems if we assume further the multiplier mapping is calm. Moreover, our results reveal that the combination of uniqueness of Lagrange multipliers and the calmness of the multiplier mapping amounts to the strong Robinson constraint qualification. These illustrate that the calmness of multiplier mapping, being automatically satisfied for NLPs, is a property that is required in order to achieve a similar result as those in NLPs for the constrained problem (4.1) in general. Such a calmness property was recently used in [42] in order to characterize noncriticality of Lagrange multipliers for generalized KKT systems that encompass the KKT system for the constrained optimization problem (4.1).

Our second goal is to provide an important application of the established characterizations of the strong metric subregularity of the KKT system of (4.1) in the basic sequential programming method (SQP) for this problem. For the NLPs framework, the sharpest result was achieved by Bonnans [1] in which he showed that the combination of the uniqueness of Lagrange multipliers and the second-order sufficient condition ensures that the basic SQP method can generate a sequence that is convergent and the rate of convergence is superlinear. We will show that Bonnans' result can be extended for the parabolically regular constrained optimization problems if we further assume that the multiplier mapping is calm. The main source of this chapter is [40].

A mapping $S\colon X \rightrightarrows \mathbb{Y}$ is called calm at $(\bar{x}, \bar{y}) \in S$ if there are $\ell \geq 0$ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$S(x) \cap V \subset S(\bar{x}) + \ell\|x - \bar{x}\|I\!B \quad \text{for all } \ x \in U$$

holds. The mapping S is called isolated calm at $(\bar{x}, \bar{y})$ if there are $\ell \geq 0$ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ for which we have

$$S(x) \cap V \subset \big\{\bar{y}\big\} + \ell\|x - \bar{x}\|I\!B \quad \text{for all } \ x \in U.$$

It is well-known that the calmness and isolated calmness of a set-valued mapping S at $(\bar{x}, \bar{y})$ amount to the metric subregularity and strong metric subregularity, respectively, of S^{-1} at $(\bar{y}, \bar{x})$.

We recall below an important consequence of parabolic regularity that was recently observed in chapter 3. Recall that a set-valued mapping $S : \mathbb{X} \rightrightarrows \mathbb{Y}$ is called proto-differentiable at $\bar{x}$ for $\bar{y}$ with $(\bar{x}, \bar{y}) \in S$ if the set S is geometrically derivable at $(\bar{x}, \bar{y})$. When this condition holds for the set-valued mapping S at $\bar{x}$ for $\bar{y}$, we refer to $DS(\bar{x}, \bar{y})$ as the proto-derivative of F at $\bar{x}$ for $\bar{y}$.

Proposition 4.1 (proto-differentiability of normal cone mappings) *Let Θ be a closed convex set in $\mathbb{R}^n$ and let $\bar{x} \in \Theta$ and $\bar{v} \in N_\Theta(\bar{x})$. Assume further that Θ is parabolically derivable at $\bar{x}$ for every vector $w \in K_\Theta(\bar{x}, \bar{v})$, and that Θ is parabolically regular at $\bar{x}$ for $\bar{v}$. Then the following equivalent conditions hold:*

(i) *the indicator function δ_Θ is twice epi-differentiable at $\bar{x}$ for $\bar{v}$;*

(ii) *the normal cone mapping N_Θ is proto-differentiable at $\bar{x}$ for $\bar{v}$ and*

$$DN_\Theta(\bar{x}, \bar{v})(w) = \partial\big[\frac{1}{2}\mathrm{d}^2\delta_\Theta(\bar{x}, \bar{v})\big](w) \quad \textit{for all } w \in \mathbb{R}^n. \tag{4.2}$$

Furthermore, we have $\operatorname{dom} \mathrm{d}^2\delta_\Theta(\bar{x}, \bar{v}) = K_\Theta(\bar{x}, \bar{v})$ *and the second subderivative* $\mathrm{d}^2\delta_\Theta(\bar{x}, \bar{v})$ *is a proper convex function.*

Proof.

The validity and equivalence of (i) and (ii) follows from Corollary 3.11 by setting $f = \delta_\Theta$. Finally, it follows from [54, Proposition 13.20] that the second subderivative $\mathrm{d}^2\delta_\Theta(\bar{x}, \bar{v})$ is proper convex. $\square$

We provide a simple but useful consequence of the above result in which the proto-derivative of the normal cone mapping will be calculated at the origin.

Corollary 4.2 (proto-derivatives of parabolically regular sets) *Let $\Theta \subset \mathbb{R}^n$ satisfy the assumptions of Proposition 4.1. Then we have*

$$DN_\Theta(\bar{x}, \bar{v})(0) = N_{K_\Theta(\bar{x},\bar{v})}(0) = \big(K_\Theta(\bar{x}, \bar{v})\big)^*. \tag{4.3}$$

Proof.

It follows from (4.2) that

$$DN_\Theta(\bar{x}, \bar{v})(0) = \partial\big[\mathrm{d}^2\delta_\Theta(\bar{x}, \bar{v})\big](0).$$

Let $u \in DN_\Theta(\bar{x},\bar{v})(0)$ and thus get $u \in \partial\big[\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})\big](0)$. Since the second subderivative $\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})$ is a convex function and $\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(0) = 0$, we obtain

$$\langle u, w - 0\rangle \le \frac{1}{2}\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(w) - \frac{1}{2}\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(0) = \frac{1}{2}\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(w) \quad \text{for all } \ w \in \mathbb{R}^n.$$

By Proposition 4.1, we have $\operatorname{dom} \mathrm{d}^2\delta_\Theta(\bar{x},\bar{v}) = K_\Theta(\bar{x},\bar{v})$. Let $t > 0$ and $w \in K_\Theta(\bar{x},\bar{v})$. Since the second subderivative $\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})$ is positive homogeneous of degree 2, we obtain $\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(tw) = t^2\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(w)$, which results in

$$t\langle u, w\rangle \le \frac{1}{2}\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(tw) = \frac{t^2}{2}\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(w).$$

This along with $w \in \operatorname{dom} \mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})$ confirms that $\langle u, w\rangle \le 0$ and hence $u \in \big(K_\Theta(\bar{x},\bar{v})\big)^*$. This gives us the inclusion '$\subset$' in (4.3). Assume now $u \in \big(K_\Theta(\bar{x},\bar{v})\big)^*$. So we have $\langle u, w\rangle \le 0$ for all $w \in K_\Theta(\bar{x},\bar{v})$. This leads us to

$$\langle u, w\rangle \le 0 \le \frac{1}{2}\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(w) = \frac{1}{2}\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(w) - \frac{1}{2}\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})(0) \quad \text{for all } \ w \in K_\Theta(\bar{x},\bar{v}).$$

If $w \notin K_\Theta(\bar{x},\bar{v})$, then by Proposition 4.1, we have $\mathrm{d}^2\delta_\Theta(\bar{x} \,|\, \bar{v})(w) = \infty$, which clearly implies that the above inequality holds for such w. Consequently we arrive at $u \in \partial\big[\frac{1}{2}\mathrm{d}^2\delta_\Theta(\bar{x},\bar{v})\big](0)$ and therefore $u \in DN_\Theta(\bar{x},\bar{v})(0)$. This gives the inclusion '$\supset$' in (4.3) and hence completes the proof. $\square$

Considering the constrained problem (4.1), we define its KKT system by the equations

$$\nabla_x L(x,\lambda) = \nabla\varphi(x) + \nabla F(x)^*\lambda = 0, \qquad \lambda \in N_\Theta\big(F(x)\big), \tag{4.4}$$

where $L(x,\lambda) := \varphi(x) + \langle F(x), \lambda\rangle$ is the Lagrangian associated with problem (4.1) and $(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m$. Given a point $\bar{x} \in \mathbb{R}^n$, we define the set of *Lagrange multipliers* associated with $\bar{x}$ by

$$\Lambda(\bar{x}) := \big\{\lambda \in \mathbb{R}^m \;\big|\; \nabla_x L(\bar{x},\lambda) = 0,\ \lambda \in N_\Theta\big(F(\bar{x})\big)\big\}. \tag{4.5}$$

Having $(\bar{x},\bar{\lambda})$ as a solution to the KKT system (4.4) always yields $\bar{\lambda} \in \Lambda(\bar{x})$. It is not hard to see that if $\bar{\lambda} \in \Lambda(\bar{x})$, then $\bar{x}$ is a stationary point of the KKT system (4.4) in the sense that it satisfies the condition

$$0 \in \partial^{-}\big(\varphi + \delta_{\Theta} \circ F\big)(\bar{x}).$$

In order to facilitate our presentation in this section, we list our basic assumptions that we often impose on the convex set Θ from (4.1) at the point $\bar{x}$ with $F(\bar{x}) \in \Theta$:

(H1) for every $\lambda \in \Lambda(\bar{x})$, the set Θ is parabolically derivable at $F(\bar{x})$ for all vectors of the form $\nabla F(\bar{x})w$ in $K_{\Theta}(F(\bar{x}), \lambda)$;

(H2) the set Θ is parabolically regular at $F(\bar{x})$ for every $\lambda \in \Lambda(\bar{x})$.

It was shown in [39] that both assumptions (H1) and (H2) hold for important convex sets appearing in constrained optimization problems including polyhedral convex sets, the second-order cone, and the cone of symmetric and positive semidefinite matrices. Indeed, it was proven in [39, Proposition 5.14] that any $\mathcal{C}^2$-cone reducible is satisfying these assumptions. Imposing these assumptions on the set Θ opens the door for deriving second-order optimality conditions for the constrained optimization problem (4.1). In the following we recall the second-order optimality condition for the constrained optimization problem established in Chapter 3, Theorem 3.28.

Theorem 4.3 (second-order optimality conditions) *Let the basic assumptions* (H1) *and* (H2) *hold for* Θ *from* (4.1) *and* $\bar{v} = -\nabla\varphi(\bar{x})$. *Assume further that the mapping* $x \mapsto F(x) - \Theta$ *is metrically subregular at* $(\bar{x}, 0)$. *Set* $\Omega := \{x \in \mathbb{R}^n \,|\, F(x) \in \Theta\}$. *Then we have the following second-order optimality conditions for the constrained problem* (4.1)*:*

(i) *If* $\bar{x}$ *is a local minimum of* (4.1), *then the second-order necessary condition*

$$\max_{\lambda \in \Lambda(\bar{x})} \big\{\langle \nabla^2_{xx} L(\bar{x}, \lambda)w, w\rangle + \mathrm{d}^2\delta_{\Theta}(F(\bar{x}), \lambda)(\nabla F(\bar{x})w)\big\} \geq 0$$

holds for all $w \in K_{\Omega}(\bar{x}, \bar{v})$.

(ii) *The validity of the second-order condition*

$$\max_{\lambda\in\Lambda(\bar{x})}\big\{\langle\nabla^2_{xx}L(\bar{x},\lambda)w,w\rangle+\mathrm{d}^2\delta_\Theta(F(\bar{x}),\lambda)(\nabla F(\bar{x})w)\big\}>0\quad\text{for all}\quad w\in K_\Omega(\bar{x},\bar{v})\setminus\{0\}$$

amounts to the existence of constants $\ell>0$ *and* $\varepsilon>0$ *such that the growth condition*

$$f(x)\ge f(\bar{x})+\frac{\ell}{2}\|x-\bar{x}\|^2\quad\text{for all}\;\; x\in I\!B_\varepsilon(\bar{x})$$

holds, where $f:=\varphi+\delta_\Theta\circ F$.

As argued in [39], the above second-order necessary and sufficient conditions for parabolically regular problem (4.1) boils down the well-know second-order conditions for $\mathcal{C}^2$-cone reducible constrained problems obtained in [4].

Given a point $\bar{x}\in\mathbb{R}^n$, we define the multiplier mapping $M_{\bar{x}}:\mathbb{R}^n\times\mathbb{R}^m\rightrightarrows\mathbb{R}^m$ associated with $\bar{x}$ by

$$M_{\bar{x}}(v,w):=\big\{\lambda\in\mathbb{R}^m\big|\;(v,w)\in G(\bar{x},\lambda)\big\}\quad\text{with}\;\;(v,w)\in\mathbb{R}^n\times\mathbb{R}^m,\tag{4.6}$$

where the mapping $G\colon\mathbb{R}^n\times\mathbb{R}^m\rightrightarrows\mathbb{R}^n\times\mathbb{R}^m$ is defined by

$$G(x,\lambda):=\begin{bmatrix}\nabla_x L(x,\lambda)\\ \\ -F(x)\end{bmatrix}+\begin{bmatrix}0\\ \\ N_\Theta^{-1}(\lambda)\end{bmatrix}.\tag{4.7}$$

It is easy to observe that $M_{\bar{x}}(0,0)=\Lambda(\bar{x})$, where $\Lambda(\bar{x})$ the set of Lagrange multipliers at $\bar{x}$ from (4.5). We collect some important properties of the multiplier mapping in the following result.

Theorem 4.4 (properties of multipliers mappings) *Let* $(\bar{x},\bar{\lambda})$ *be a solution to the KKT system* (4.4). *Then following properties of the multiplier mapping* $M_{\bar{x}}$ *are equivalent:*

(i) *the multiplier mapping* $M_{\bar{x}}$ *is calm at* $((0,0),\bar{\lambda})$ *and* $\Lambda(\bar{x})=\{\bar{\lambda}\}$*;*

(ii) *the multiplier mapping* $M_{\bar{x}}$ *is isolated calm at* $((0,0),\bar{\lambda})$*;*

(iii) *the dual qualification condition*

$$DN_\Theta(F(\bar{x}), \bar{\lambda})(0) \cap \ker \nabla F(\bar{x})^* = \{0\} \tag{4.8}$$

holds.

In addition, if the convex set Θ *from* (4.1) *satisfy the basic assumptions* (H1) *and* (H2)*, then the above conditions are equivalent to the following one:*

(iv) *The strong Robinson constraint qualification holds:*

$$\nabla F(\bar{x})\mathbb{R}^n + \left[T_\Theta(F(\bar{x})) \cap \{\bar{\lambda}\}^\perp\right] = \mathbb{R}^m. \tag{4.9}$$

Proof.

The equivalence between (i), (ii), and (iii) was already achieved in [42, Theorem 3.1]. Moreover, it was shown in [42, Proposition 4.3] that conditions (iii) and (iv) equivalent for any $\mathcal{C}^2$-cone reducible set. The same argument together with $DN_\Theta(F(\bar{x}), \bar{\lambda})(0) = N_{K_\Theta(F(\bar{x}),\bar{\lambda})}(0)$, resulted from Corollary 4.2, can be utilized to get the same equivalence for sets satisfying the basic assumptions (H1) and (H2). $\square$

We end this section by comparing the dual condition (4.8) with the well-known Robinson constraint qualification for the constrained optimization problem (4.1).

Proposition 4.5 *Let* $(\bar{x}, \bar{\lambda})$ *be a solution to the KKT system* (4.4)*, and let* Θ *from* (4.1) *satisfy the basic assumptions* (H1) *and* (H2)*. If the dual condition* (4.8) *is satisfied, then the following conditions hold:*

(i) *the basic constraint qualification*

$$N_\Theta(F(\bar{x})) \cap \ker \nabla F(\bar{x})^* = \{0\} \tag{4.10}$$

holds;

(ii) *for any vector* $w \in \mathbb{R}^n$ *with* $\nabla F(\bar{x})w \in K_\Theta(F(\bar{x}), \bar{\lambda})$*, we have*

$$N_{K_\Theta(F(\bar{x}),\bar{\lambda})}(\nabla F(\bar{x})w) \cap \ker \nabla F(\bar{x})^* = \{0\}.$$

Proof.

It follows from the basic assumptions (H1) and (H2) together with Proposition 4.1 that δ_Θ is twice epi-differentiable at $F(\bar{x})$ for $\bar{\lambda}$. Using this and (4.3) gives us $DN_\Theta(F(\bar{x}) \,|\, \bar{\lambda})(0) = N_{K_\Theta(F(\bar{x}),\bar{\lambda})}(0)$. Since we always have the inclusion

$$N_\Theta(F(\bar{x})) \subset \mathrm{cl}\,(N_\Theta(F(\bar{x})) + \mathbb{R}\bar{\lambda}) = \big(K_\Theta(F(\bar{x}), \bar{\lambda})\big)^* = N_{K_\Theta(F(\bar{x}),\bar{\lambda})}(0),$$

we conclude from the dual condition (4.8) that the basic constraint qualification (4.10) holds, which proves (i).

Turning to (ii), pick $w \in \mathbb{R}^n$ such that $\nabla F(\bar{x})w \in K_\Theta(F(\bar{x}), \bar{\lambda})$. It is easy to see that

$$N_{K_\Theta(F(\bar{x}),\bar{\lambda})}(\nabla F(\bar{x})w) \subset N_{K_\Theta(F(\bar{x}),\bar{\lambda})}(0) = DN_\Theta(F(\bar{x}), \bar{\lambda})(0).$$

Appealing now to (4.8) proves (ii) and thus completes the proof. □

Since our analysis in this paper will be conducted under the basic assumptions (H1) and (H2) and since the primal and dual conditions (4.8) and (4.9), respectively, are equivalent under these conditions, we will refer to both conditions as the *strong Robinson constraint qualification.*

4.1 Critical Multipliers for Composite Optimization Problems

This section aims to investigate the strong metric subregularity of the KKT system (4.4). It has been well-known that such a property play a prominent role in the study of the rate of convergence of numerical algorithms for optimization problems. The study of the aforementioned property of the KKT system (4.4) is pertinent to analyzing the solution mapping $S\colon \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$, defined by

$$S(v, w) := \big\{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \;|\; (v, w) \in G(x, \lambda)\big\} \quad \text{with} \;\; (v, w) \in \mathbb{R}^n \times \mathbb{R}^m \tag{4.11}$$

via the set-valued mapping G from (4.7). This, indeed, is the solution map to the *canonical perturbation* of (4.4) and so can be seen as the KKT system of the canonically perturbed problem

$$\min_{x\in\mathbb{R}^n} \quad \varphi_0(x) - \langle v, x\rangle \quad \text{subject to} \quad F(x) + w \in \Theta. \tag{4.12}$$

Below we first reveal the relationship between isolated calmness of S and noncriticality of Lagrange multipliers being important for the subsequent results. Recall the concept of critical and noncritical Lagrange multipliers for the KKT system (4.4) from [41, Definition 3.1]: Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system (4.4). The Lagrange multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is said critical for (4.4) if there is a nonzero $w \in \mathbb{R}^n$ satisfying the inclusion

$$0 \in \nabla^2_{xx} L(\bar{x}, \bar{\lambda})w + \nabla F(\bar{x})^* DN_\Theta(F(\bar{x})), \bar{\lambda})(\nabla F(\bar{x})w).$$

The Lagrange multiplier $\bar{\lambda} \in \Lambda(\bar{x})$ is noncritical for (4.4) when the above generalized equation admits only the trivial solution $\xi = 0$.

The above definition is a generalization of the same concept introduced by Izmailov in [26] for nonlinear programs with equality constraints. A characterization of this concepts can be found in [28, Proposition 1.43] for NLPs and in [42, Theorem 4.1] for $\mathcal{C}^2$-cone reducible constrained problems.

Proposition 4.6 (characterization of isolated calmness via noncriticality) *Let* $(\bar{x}, \bar{\lambda})$ *be a solution to the KKT system* (4.4). *Consider the following statements:*

- **(i)** *the Lagrange multiplier* $\bar{\lambda}$ *is noncritical for* (4.4)*;*
- **(ii)** *the mapping* S *from* (4.11) *is isolated calm at* $((0,0), (\bar{x}, \bar{\lambda}))$.

Then the following hold:

- **(a)** *the implication* (ii) $\Longrightarrow$ (i) *always holds;*

(b) *if* $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ *and the multiplier mapping* $M_{\bar{x}}$ *is calm at* $((0,0),\bar{\lambda})$*, then we have* (i) $\Longrightarrow$ (ii).

Proof.

It is not hard to see from the definition of noncriticality that the Lagrange multiplier $\bar{\lambda}$ is noncritical if and only if the following implication holds:

$$\begin{cases} \nabla^2_{xx} L(\bar{x},\bar{\lambda})w + \nabla F(\bar{x})^* u = 0, \\ u \in DN_\Theta(F(\bar{x}),\bar{\lambda})(\nabla F(\bar{x})w) \end{cases} \Longrightarrow w = 0. \tag{4.13}$$

However, from the well-known criterion for the characterization of the isolated calmness of S (see, e.g., [11, Theorem 4G.1]), we know that the solution mapping S is isolated calm at $((0,0),(\bar{x},\bar{\lambda}))$ if and only if $DS((0,0),(\bar{x},\bar{\lambda}))(0,0) = \{0,0)\}$. This can be equivalently expressed as

$$\begin{cases} \nabla^2_{xx} L(\bar{x},\bar{\lambda})w + \nabla F(\bar{x})^* u = 0, \\ u \in DN_\Theta(F(\bar{x}),\bar{\lambda})(\nabla F(\bar{x})w) \end{cases} \Longrightarrow w = 0, u = 0. \tag{4.14}$$

Let (ii) be satisfied. Thus (4.14) holds and so does implication (4.13), which proves (a). To prove (b), assume that (i) is satisfied. This results in the validity of implication (4.13). Moreover, Theorem 4.4 ensures that the condition (4.8) is satisfied. Pick (w,u) satisfying the left side equations of implication (4.14). By (4.13) we obtain $w = 0$. This implies that

$$u \in DN_\Theta(F(\bar{x}),\bar{\lambda})(0) \cap \ker \nabla F(\bar{x})^*.$$

Appealing next to (4.8) confirms that $u = 0$, which tells us that (4.14) holds and hence completes the proof. $\square$

It is important to notice that calmness of the multiplier mapping is essential for implication (i) $\Longrightarrow$ (ii) in Proposition 4.6(b). Indeed, [42, Example 4.8] provides a simple semidefinite program in which the calmness of the Lagrange multiplier mapping $M_{\bar{x}}$ fails while the unique Lagrange multiplier therein is noncritical but the solution mapping S is not isolated calm. However, when Θ is a polyhedral convex set, noncriticality along with uniqueness of Lagrange multipliers does result in the isolated calmness of S. This falls out of Proposition 4.6 because in this case the multiplier mapping $M_{\bar{x}}$ is automatically calm as a direct consequence of the Hoffman lemma.

We proceed with deriving a characterization of the isolated calmness of the solution map S from (4.11) via the second-order sufficient condition.

Theorem 4.7 (characterization of isolated calmness for KKT systems) *Let* $(\bar{x}, \bar{\lambda})$ *be a solution to the KKT system* (4.4) *and let the basic assumptions* (H1) *and* (H2) *hold. Then the following conditions are equivalent:*

- **(i)** *the mapping* S *from* (4.11) *is isolated calm at* $((0,0),(\bar{x},\bar{\lambda}))$ *and* $\bar{x}$ *is local minimum of* (4.1).
- **(ii)** *the second-order sufficient condition*

$$\begin{cases} \langle \nabla^2_{xx} L(\bar{x},\bar{\lambda})w, w\rangle + \mathrm{d}^2\delta_\Theta(F(\bar{x}),\bar{\lambda})(\nabla F(\bar{x})w) > 0 \\ \textit{for all } w \in \mathbb{R}^n \setminus \{0\} \ \ \textit{with} \ \ \nabla F(\bar{x})w \in K_\Theta(F(\bar{x}),\bar{\lambda}). \end{cases} \tag{4.15}$$

  holds, the multiplier mapping $M_{\bar{x}}$ *is calm at* $((0,0),\bar{\lambda})$*, and* $\Lambda(\bar{x}) = \{\bar{\lambda}\}$.
- **(iii)** *the second-order sufficient condition* (4.15) *and the strong Robinson constraint qualification* (4.8) *are satisfied.*
- **(iv)** *the Lagrange multiplier* $\bar{\lambda}$ *from* (4.5) *is noncritical for* (4.4)*,* $\bar{x}$ *is local minimum of* (4.1)*, the multiplier mapping* $M_{\bar{x}}$ *is calm at* $((0,0),\bar{\lambda})$*, and* $\Lambda(\bar{x}) = \{\bar{\lambda}\}$.

Proof.

We begin with proving implication (i) $\Longrightarrow$ (iv). Since S is isolated calm at $((0,0),(\bar{x},\bar{\lambda}))$ and $\Lambda(\bar{x})$ is convex, we obtain $\Lambda(\bar{x}) = \{\bar{\lambda}\}$. The calmness of the Lagrange multiplier mapping

$M_{\bar{x}}$ is calm at $((0,0),\bar{\lambda})$ is a direct consequence of the isolated calmness of S at $((0,0),(\bar{x},\bar{\lambda}))$. Finally, by Proposition 4.6(a), $\bar{\lambda}$ is noncritical and so we get (iv).

Turning now to implication (iv) $\Longrightarrow$ (ii), assume that (iv) holds and hence by Theorem 4.4 deduce that (4.8) is satisfied. This along with Proposition 4.5(i) implies that the basic constraint qualification (4.10) holds. Employing now Theorem 3.27(i) and using $\Lambda(\bar{x})=\{\bar{\lambda}\}$ tell us that the second-order necessary condition

$$\big\langle \nabla^2_{xx}L(\bar{x},\bar{\lambda})w,w\big\rangle + \mathrm{d}^2\delta_\Theta(F(\bar{x}),\bar{\lambda})(\nabla F(\bar{x})w)\geq 0 \quad \text{for all } \; w\in K_\Omega(\bar{x},-\nabla\varphi(\bar{x})) \tag{4.16}$$

fulfills, where $\Omega:=\{x\in\mathbb{R}^n | F(x)\in\Theta\}$. Moreover, we deduce from (4.10) that

$$w\in K_\Omega(\bar{x},-\nabla\varphi(\bar{x})) \iff \nabla F(\bar{x})w\in K_\Theta(F(\bar{x}),\bar{\lambda}).$$

To prove the second-order sufficient condition (4.15), we need to show that the above inequality is strict for any $w\neq 0$. Suppose by contradiction that there is a $\bar{w}\neq 0$ such that

$$\big\langle \nabla^2_{xx}L(\bar{x},\bar{\lambda})\bar{w},\bar{w}\big\rangle + \mathrm{d}^2\delta_\Theta(F(\bar{x}),\bar{\lambda})(\nabla F(\bar{x})\bar{w})=0 \;\text{ with }\; \nabla F(\bar{x})\bar{w}\in K_\Theta(F(\bar{x}),\bar{\lambda}).$$

Consider now the optimization problem

$$\min_{w\in\mathbb{R}^n} \;\; \frac{1}{2}\Big[\big\langle \nabla^2_{xx}L(\bar{x},\bar{\lambda})w,w\big\rangle + \mathrm{d}^2\delta_\Theta(F(\bar{x}),\bar{\lambda})(\nabla F(\bar{x})w)\Big]. \tag{4.17}$$

It follows from (4.16) that $\bar{w}$ is an optimal solution to (4.17). Since $\nabla F(\bar{x})\bar{w}\in K_\Theta(F(\bar{x}),\bar{\lambda})$, by Proposition 4.5(ii), we have the following chain rule:

$$\partial_w\big(\frac{1}{2}\mathrm{d}^2\delta_\Theta(F(\bar{x}),\bar{\lambda})(\nabla F(\bar{x})\cdot)\big)(\bar{w}) = \nabla F(\bar{x})^*\partial\big[\frac{1}{2}\mathrm{d}^2\delta_\Theta(F(\bar{x}),\bar{\lambda})\big](\nabla F(\bar{x})\bar{w}).$$

Applying the subdifferential Fermat principle to the latter problem and then using the above chain rule yield

$$\begin{aligned} 0 \;\;&\in\;\; \nabla^2_{xx}L(\bar{x},\bar{\lambda})\bar{w}+\nabla F(\bar{x})^*\partial\big[\frac{1}{2}\mathrm{d}^2\delta_\Theta(F(\bar{x}),\bar{\lambda})\big](\nabla F(\bar{x})\bar{w}) \\ &=\;\; \nabla^2_{xx}L(\bar{x},\bar{\lambda})\bar{w}+\nabla F(\bar{x})^*DN_\Theta(F(\bar{x}),\bar{\lambda})(\nabla F(\bar{x})\bar{w}), \end{aligned}$$

where the last equality comes from (4.2). Since $\bar{w} \neq 0$, this inclusion justifies that $\bar{\lambda}$ is critical, a contradiction. Thus the inequality in (4.16) is strict, which yields the validity of the second-order sufficient condition (4.15) and hence proves (ii).

To justify (ii) $\Longrightarrow$ (i), suppose that (ii) is satisfied. We are going to prove that $\bar{\lambda}$ is noncritical. To do so, it suffices to show that implication (4.13) is satisfied. Pick (w, u) satisfying on the left side equations of the latter implication. This brings us to

$$\langle \nabla^2_{xx} L(\bar{x}, \bar{\lambda})w, w\rangle + \langle u, \nabla F(\bar{x})w\rangle = 0 \quad \text{and} \quad u \in DN_\Theta(F(\bar{x}), \bar{\lambda})(\nabla F(\bar{x})w). \tag{4.18}$$

The second relationship together with (4.2) yields $u \in \partial\big[\frac{1}{2}\mathrm{d}^2\delta_\Theta(F(\bar{x}), \bar{\lambda})\big](\nabla F(\bar{x})w)$. Thus we have $\nabla F(\bar{x})w \in \operatorname{dom} \mathrm{d}^2\delta_\Theta(F(\bar{x}), \bar{\lambda}) = K_\Theta(F(\bar{x}), \bar{\lambda})$. Moreover, since $\mathrm{d}^2\delta_\Theta(F(\bar{x}), \bar{\lambda})$ is a convex function due to Proposition 4.1, by the definition of the subdifferential from convex analysis we obtain for all $v \in \mathbb{R}^m$ that

$$\langle u, v - \nabla F(\bar{x})w\rangle \leq \frac{1}{2}\mathrm{d}^2\delta_\Theta(F(\bar{x}), \bar{\lambda})(v) - \frac{1}{2}\mathrm{d}^2\delta_\Theta(F(\bar{x}), \bar{\lambda})(\nabla Frchet(\bar{x})w).$$

Let $\varepsilon \in (0, 1)$ and set $v := (1 \pm \varepsilon)\nabla F(\bar{x})w$. Since the second subderivative is positive homogeneous of degree 2, we arrive at

$$\pm\langle u, \nabla F(\bar{x})w\rangle \leq \frac{\varepsilon \pm 2}{2}\mathrm{d}^2\delta_\Theta(F(\bar{x}), \bar{\lambda})(\nabla F(\bar{x})w).$$

Letting $\varepsilon \downarrow 0$ clearly gives us the equality $\mathrm{d}^2\delta_\Theta(F(\bar{x}), \bar{\lambda})(\nabla F(\bar{x})w) = \langle u, \nabla F(\bar{x})w\rangle$. Combining this and (4.18) ensures that

$$\langle \nabla^2_{xx} L(\bar{x}, \bar{\lambda})w, w\rangle + \mathrm{d}^2\delta_\Theta(F(\bar{x}), \bar{\lambda})(\nabla F(\bar{x})w) = 0, \;\; \nabla F(\bar{x})w \in K_\Theta(F(\bar{x}), \bar{\lambda}),$$

which results in $w = 0$ due to the second-order condition (4.3) and thus $\bar{\lambda}$ is noncritical. Appealing now to Proposition 4.6(b) tells that the mapping S is isolated calm at $((0, 0), (\bar{x}, \bar{\lambda}))$. We also conclude from (4.15), $\Lambda(\bar{x}) = \{\bar{\lambda}\}$, and Theorem 3.28(ii) that $\bar{x}$ is a local minimum of (4.1), which proves that (i) holds.

Finally, the equivalence between (ii) and (iii) is a direct consequence of Theorem 4.4.

This completes the proof. □

The equivalence between (i) and (iii) Theorem 4.7 was obtained in [13, Theorem 24] for a $\mathcal{C}^2$-cone reducible set Θ using a different approach. It is important to notice that our proof is deeply rooted into our developments in chapter 3 regarding parabolic regularity and its important consequences in second-order variational analysis. Parts (ii) and (iv) in Theorem 4.7 did not appear in [13, Theorem 24] and are new to the best our knowledge. These parts highlight a significant difference for dealing with the KKT system (4.4) when Θ is a nonpolyhedral set. In the case that Θ is a polyhedral convex set in (4.4), (4.1) boils down to a nonlinear programming problem. It is well-known that for the latter framework (cf. [12, Theorem 2.6]) the uniqueness of Lagrange multipliers together with (4.15) amounts to the isolated calmness of the solution mapping S from (4.11). It was argued in [13] that a similar result for constrained optimization problems in general can not be expected and then shown that if we replace the uniqueness of Lagrange multipliers with the strong Robinson constraint qualification (4.9), then a similar result can be justified (cf. [13, Theorem 24]). However, the authors in [13] did not address the question that why such a result may fail for a constrained optimization problem with a nonpolyhedral set Θ. Theorem 4.7(ii) answers the latter question by revealing that, indeed, the calmness of the multiplier mapping $M_{\bar{x}}$ is essential in order to derive such an equivalence. In the framework of nonlinear programming, the latter calmness property is automatically satisfied by the Hoffman Lemma. However, it may fail for constrained optimization problems as shown in [42, Example 3.4] for a semidefinite programming problem. Moreover, [42, Example 4.8] shows that for the latter example the isolated calmness property of S fails while the set of Lagrange multipliers is a singleton. This says that if the calmness of the multiplier mapping $M_{\bar{x}}$ is not satisfied, the equivalences in Theorem 4.7 may fail in general.

Finally, we want to add to this discussion that [13, Theorem 24] assume the basic constraint qualification (4.10). However, our proof shows that the latter condition can be dropped without any harm.

4.2 Superlinear Convergence of The SQP Method

This section is devoted to establish the primal-dual superlinear convergence of the basic SQP method for the constrained optimization problem (4.1). The SQP has been among the most effective methods for solving *nonlinear* constrained optimization problems. The principal idea of the SQP is to solve a sequence of quadratic approximations, called subproblems, whose optimal solutions converge to that of the original problem under some appropriate assumptions.

Given the current iterate x_k, the generic SQP subproblems for the constrained optimization problem (4.1) are formulated as

$$\begin{cases} \min\limits_{x\in\mathbb{R}^n} & \varphi(x_k)+\langle\nabla\varphi(x_k),x-x_k\rangle+\dfrac{1}{2}\langle H_k(x-x_k),x-x_k\rangle \\ \text{subject to} & F(x_k)+\nabla F(x_k)(x-x_k)\in\Theta, \end{cases} \tag{4.19}$$

where $H_k\in\mathbb{R}^{n\times n}$ for all $k\in I\!N$. The KKT system of this subproblem can be formulated as the generalized equation

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla_x L(x_k,\lambda_k) \\ -F(x_k) \end{bmatrix} + \begin{bmatrix} H_k & \nabla F(x_k)^* \\ -\nabla F(x_k) & 0 \end{bmatrix} \begin{bmatrix} x-x_k \\ \lambda-\lambda_k \end{bmatrix} + \begin{bmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{bmatrix}. \tag{4.20}$$

The generic SQP method for problem (4.1) is therefore as follows:

Algorithm 4.8 (SQP method) Choose $(x_k,\lambda_k)\in\mathbb{R}^n\times\mathbb{R}^m$ and set $k=0$.

(1) If (x_k,λ_k) satisfies the KKT system (4.4), then stop.

(2) Choose a $H_k\in\mathbb{R}^{n\times n}$ and compute (x_{k+1},λ_{k+1}) as a solution to the generalized equation (4.20).

(3) Increase k by 1 and then go back to Step (1).

In *quasi-Newton SQP* methods, H_k is chosen as a quasi-Newton approximation of $\nabla^2_{xx}L(x_k,\lambda_k)$. An efficient method to construct H_k is the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method; see [28, page 212] for more detail about this method. Some interesting discussion and suggestions on how to construct a quasi-Newton SQP approximation can be found in [3, Chapter 18]. In what follows, as [28, page 229], we refer to the *basic* SQP method when we let

$$H_k = \nabla^2_{xx}L(x_k,\lambda_k). \tag{4.21}$$

As it is well-known the basic SQP method can be viewed as a natural extension of the Newton method that is implemented for generalized equations rather than equations. Indeed, we can equivalently express this system as the generalized equation

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla_x L(x,\lambda) \\ -F(x) \end{bmatrix} + \begin{bmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{bmatrix}. \tag{4.22}$$

Employing the Newton method for this generalized equation brings us to the aforementioned basic SQP method for the constrained problem (4.1); see [28, Section 3.1] for more detail.

The Newton method for generalized equations has been investigated extensively since the late 1970s; see [11, 28] and references therein. Josephy's observation in [29] was significantly advanced the topic by showing that strong regularity can ensure the superlinear convergence of the Newton method for variational inequalities. Considering nonlinear programming problems, Robinson [44, Theorem 4.1] showed that strong regularity can be guaranteed under a stronger form of the second-order sufficient condition (4.15) together with the linear independence constraint qualification. The next improvement in the latter framework was achieved by Bonnans in [1], where he showed that for nonlinear programming problems (NLPs) the basic SQP method converges superlinearly if the second-order sufficient condition (4.15) with Θ being a polyhedral set holds and the set of Lagrange multipliers is a singleton. In this section, we are going to extend this result for the constrained problem (4.1). As our results in pervious section reveals, such an extension requires one extra assumption, namely the calmness of the multiplier mapping. To achieve our goal, we utilize [28, Theorem 3.2] in which superlinear convergence of the Newton method for the generalized equation was established under two assumptions. For the generalized equation (4.22), the assumptions utilized in [28, Theorem 3.2] reads as the solution mapping S from (4.11) being isolated calm (this property was called semistability in [28]) and another property called hemistability. Recall from [28, Definition 3.1] that a solution $\bar{x}$ to the generalized equation

$$0 \in f(x) + F(x)$$

with $f : \mathbb{R}^n \to \mathbb{R}^m$ and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is called hemistable if for any x close to $\bar{x}$ the generalized equation

$$0 \in f(x) + \nabla f(x)\eta + F(x + \eta)$$

has a solution η_x such that $\eta_x \to 0$ as $x \to \bar{x}$. Below, are going to show that if the solution mapping S from (4.11) is isolated calm at $((0, 0), (\bar{x}, \bar{\lambda}))$, then the solution $(\bar{x}, \bar{\lambda})$ to (4.22) is hemistable.

Assume that P is a finite dimensional space. For the constrained optimization problem (4.1), we consider the perturbed problem

$$\min_{x \in \mathbb{R}^n} \quad \widetilde{\varphi}(x, p) \quad \text{subject to} \quad \widetilde{F}(x, p) \in \Theta, \tag{4.23}$$

where the functions $\widetilde{\varphi} : \mathbb{R}^n \times P \to \mathbb{R}$ and $\widetilde{F} : \mathbb{R}^n \times P \to \mathbb{R}^m$ are twice continuously differentiable with respect to both x and p. The first part of the following result is an extension of [28, Theorem 1.21], which was proved for a nonlinear programming problem. While the proof exploits a similar argument as the latter result, we provide a short argument for the readers' convenient. In what follows, we refer to a local minimizer that is *isolated* as a strict local minimizer.

Proposition 4.9 (stability properties of perturbed problems) *Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system* (4.4) *and let the basic assumptions* (H1) *and* (H2) *hold. If for the parameter $\bar{p} \in P$ we have*

$$\widetilde{\varphi}(\bar{x}, \bar{p}) = \varphi(\bar{x}), \quad \nabla_x \widetilde{\varphi}(\bar{x}, \bar{p}) = \nabla \varphi(\bar{x}), \quad \widetilde{F}(\bar{x}, \bar{p}) = F(\bar{x}), \quad \nabla_x \widetilde{F}(\bar{x}, \bar{p}) = \nabla F(\bar{x}), \tag{4.24}$$

then the following conditions hold:

(i) *if $\bar{x}$ is a local strict minimum of* (4.23) *for $p = \bar{p}$ and if the basic constraint qualification* (4.10) *holds, then for any $p \in P$ sufficiently close to $\bar{p}$, problem* (4.23) *attains a local minimum x_p converging to $\bar{x}$ as $p \to \bar{p}$;*

(ii) *if the second-order sufficient condition* (4.15) *and the strong Robinson constraint qualification* (4.8) *are satisfied, and*

$$\nabla^2_{xx} \widetilde{\varphi}(\bar{x}, \bar{p}) + \nabla^2_{xx} \langle \bar{\lambda}, \widetilde{F} \rangle (\bar{x}, \bar{p}) = \nabla^2_{xx} L(\bar{x}, \bar{\lambda}), \tag{4.25}$$

then there exist constants $\varepsilon > 0$ and $\ell \geq 0$ such that $\Upsilon(p) \cap \mathbb{B}_\varepsilon(\bar{x}, \bar{\lambda}) \neq \emptyset$ for all $p \in \mathbb{B}_\varepsilon(\bar{p})$ and that

$$\Upsilon(p) \cap \mathbb{B}_\varepsilon(\bar{x}, \bar{\lambda}) \subset \{(\bar{x}, \bar{\lambda})\} + \ell \|p - \bar{p}\| \quad \textit{for all } p \in \mathbb{B}_\varepsilon(\bar{p}), \tag{4.26}$$

where $\Upsilon : Painlev \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$ is the solution mapping to the KKT system of (4.23) *defined by*

$$\Upsilon(p) := \left\{ (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \,\middle|\, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla_x \widetilde{\varphi}(x, p) + \nabla_x \widetilde{F}(x, p)^* \lambda \\ -\widetilde{F}(x, p) \end{bmatrix} + \begin{bmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{bmatrix} \right\}.$$

Proof.

Since $\bar{x}$ is a strict local minimum of (4.23) for $p = \bar{p}$, we obtain an $\varepsilon > 0$ such that $\bar{x}$ is the unique minimizer for the problem

$$\min_{x \in \mathbb{R}^n} \quad \widetilde{\varphi}(x, \bar{p}) \quad \text{subject to} \quad \widetilde{F}(x, \bar{p}) \in \Theta, \; x \in \mathbb{B}_\varepsilon(\bar{x}). \tag{4.27}$$

Consider now the problem

$$\min_{x\in\mathbb{R}^n} \quad \widetilde{\varphi}(x,p) \quad \text{subject to} \quad \widetilde{F}(x,p)\in\Theta,\ x\in I\!B_\varepsilon(\bar{x}). \tag{4.28}$$

It follows from (4.10) and (4.24) that

$$N_\Theta(\widetilde{F}(\bar{x},\bar{p}))\cap\ker\nabla_x\widetilde{F}(\bar{x},\bar{p})^*=\{0\} \tag{4.29}$$

holds. This, indeed, is equivalent to the Robinson constraint qualification for constraint system $\widetilde{\Phi}(x,p)\in C$ at $(\bar{x},\bar{p})$, namely the condition

$$0\in\operatorname{int}(\widetilde{F}(\bar{x},\bar{p})+\nabla_x\widetilde{F}(\bar{x},\bar{p})\mathbb{R}^n-\Theta).$$

Thus we conclude from [4, Theorem 2.86] that for all (x,p) sufficiently close to $(\bar{x},\bar{p})$ the estimate

$$d(x,\Gamma(p))=O(d(\widetilde{F}(x,p)\ ,\ \Theta)) \tag{4.30}$$

holds, where $\Gamma(p):=\{x\in\mathbb{R}^n|\ \widetilde{F}(x,p)\in\Theta\}$. This guarantees that for p sufficiently close to $\bar{p}$, the feasible region of (4.28), namely $\Gamma(p)$, is nonempty. Appealing now to the Weierstrass Theorem, we deduce that (4.28) attains a minimum, denoted $\bar{x}_p$, for p sufficiently close to $\bar{p}$. Now we claim that $\bar{x}_p\to\bar{x}$ as $p\to\bar{p}$. Indeed, if this fails, we find a sequence $p_k\to\bar{p}$ for which $\bar{x}_{p_k}$, an optimal solution of (4.28) for $p=p_k$, does not converge to $\bar{x}$ as $k\to\infty$. Since $\bar{x}_{p_k}\in I\!B_\varepsilon(\bar{x})$, we can find a subsequence of $\bar{x}_{p_k}$ that converges to some $\widetilde{x}\in I\!B_\varepsilon(\bar{x})$ with $\widetilde{x}\neq\bar{x}$. Without relabeling, we assume without loss of generality that $\bar{x}_{p_k}\to\widetilde{x}$. By (4.30), we find $\widetilde{x}_{p_k}\in\Gamma(p_k)$ such that

$$\|\widetilde{x}_{p_k}-\bar{x}\|=O(d(\widetilde{F}(\bar{x},p_k)\ ,\ \Theta))=O(\|p_k-\bar{p}\|).$$

By $p_k\to\bar{p}$, we can assume with no harm that $\widetilde{x}_{p_k}\in I\!B_\varepsilon(\bar{x})$. Since $\bar{x}_{p_k}$ is an optimal solution of (4.28) for $p=p_k$, we get $\widetilde{F}(\bar{x}_{p_k},p_k)\le\phi(\widetilde{x}_{p_k},p_k)$. Passing to the limit brings us to $\widetilde{\varphi}(\widetilde{x},\bar{p})\le\widetilde{\varphi}(\bar{x},\bar{p})$, a contradiction to $\bar{x}$ being the unique minimum of (4.27).

Turing now to (ii), pick $\varepsilon > 0$ such that (i) holds for any $p \in I\!B_\varepsilon(\bar{p})$. We claim now that the set of Lagrange multipliers for (4.23) at $(\bar{x}, \bar{p})$ is $\{\bar{\lambda}\}$. This, in fact, can be deduced from condition (4.8), Theorem 4.4, and (4.24). We conclude from this and (4.25) and from the second-order sufficient condition (4.15) and Theorem 3.27(ii) that $\bar{x}$ is a strict local minimum of problem (4.23) for $p = \bar{p}$. Thus, according to (i), for any such a p, problem (4.23) admits a local minimum x_p with $x_p \to \bar{x}$ as $p \to \bar{p}$. It follows from condition (4.8), Proposition 4.5(i), and (4.24) that the basic constraint qualification (4.29) is satisfied. Shrinking the closed ball $I\!B_\varepsilon(\bar{p})$ if necessary, we assume without loss of generality that the basic constraint qualification holds for (4.29) holds at (x_p, p) for all $p \in I\!B_\varepsilon(\bar{p})$. This results in the existence of a Lagrange multiplier λ_p associated with the local minimizer x_p for (4.23) for all $p \in I\!B_\varepsilon(\bar{p})$. Again by shrinking the ball $I\!B_\varepsilon(\bar{p})$, we can assume that the Lagrange multipliers λ_p are uniformly bounded for all $p \in I\!B_\varepsilon(\bar{p})$.

As argued above, the set of Lagrange multipliers for (4.23) at $(\bar{x}, \bar{p})$ is $\{\bar{\lambda}\}$. Using this and the boundedness of λ_p for all $p \in I\!B_\varepsilon(\bar{p})$, we arrive at $\lambda_p \to \bar{\lambda}$ as $p \to \bar{p}$. This justifies that (x_p, λ_p) is a solution to the KKT system of (4.23), namely $(x_p, \lambda_p) \in \Upsilon(p)$. Since $(x_p, \lambda_p) \to (\bar{x}, \bar{\lambda})$ as $p \to \bar{p}$, by shrinking ε if necessary we get $\Upsilon(p) \cap I\!B_\varepsilon(\bar{x}, \bar{\lambda}) \neq \emptyset$ for all $p \in I\!B_\varepsilon(\bar{p})$.

Finally, to justify (4.26), observe first that by (4.24) we have $(\bar{x}, \bar{\lambda}) \in \Upsilon(\bar{p})$. Employing [11, Corollary 4E.3], (4.24), and (4.25) tells us that the mapping Υ is isolated calm at $(\bar{p}, (\bar{x}, \bar{\lambda}))$ if the implication

$$\begin{cases} \nabla^2_{xx} L(\bar{x}, \bar{\lambda})w + \nabla F(\bar{x})^* u = 0, \\ \\ u \in DN_\Theta(F(\bar{x}), \bar{\lambda})(\nabla F(\bar{x})w) \end{cases} \quad \Longrightarrow w = 0, u = 0$$

holds. To prove this implication, we can use a similar argument as implication (ii) $\Longrightarrow$ (i) in Theorem 4.7 to show that the second-order sufficient condition (4.15) together with condition

(4.8) ensures the validity of the aforementioned implication. This completes the proof. $\square$

The properties established in Proposition 4.9(ii) were investigated before for nonlinear programming problems in [12, Theorem 2.6]. It was extended for $\mathcal{C}^2$-cone reducible constrained problems in [13, Theorem 24] when the perturbed problem (4.23) has the canonically perturbed form (4.12), meaning that $p = (v, w) \in \mathbb{R}^n \times \mathbb{R}^m$ and

$$\widetilde{\varphi}(x, p) = \varphi(x) - \langle v, x \rangle \quad \text{and} \quad \widetilde{F}(x, p) = F(x) + w.$$

The properties in Proposition 4.9(ii) were called in [13] the *robust isolated calmness.* According to Theorem 4.7, the imposed assumptions in Proposition 4.9(ii) amounts to the solution mapping from (4.11) being isolated calm at $((0, 0), (\bar{x}, \bar{\lambda}))$. Combining this with Proposition 4.9(ii) tells us that for canonically perturbed form (4.12) the isolated calmness of S at $((0, 0), (\bar{x}, \bar{\lambda}))$ yields the robust isolated calmness of S at this point. Since the opposite statement obviously holds, we conclude that for such a particular perturbation, the isolated calmness and the robust isolated calmness of S are equivalent. The main reason to consider such a general perturbation as in (4.23) and justify the robust isolated calmness for its KKT system as we did in Proposition 4.9(ii) is that it allows us to verify the solvability of subproblems in the basic SQP method. That does not seem to be achieved via the canonically perturbed form (4.12).

Finally, it is worth mentioning that assumptions (4.24) and (4.25) were imposed in Proposition 4.9 since they are automatically satisfying for our main purpose in this paper, addressed in Proposition 4.10, which is to justify the solvability of subproblems in the basic SQP method. The basic constraint qualification condition (4.10), the second-order condition (4.15), and condition (4.8) in Proposition 4.9 are in terms of the initial data of (4.1). Adjusting them for the perturbed problem (4.23) allows us to drop assumptions (4.24) and (4.25).

Next we are going to show that the basic SQP subproblem (4.31) has always an optimal solution under certain assumptions. The proof follows similar arguments as the case of NLPs that can be found in [28, Proposition 3.37].

Proposition 4.10 (solvability of subproblems in the basic SQP method) *Let $(\bar{x}, \bar{\lambda})$ be a solution to the KKT system* (4.4) *and let the basic assumptions* (H1) *and* (H2) *hold. Then if the second-order sufficient condition* (4.15) *and the strong Robinson constraint qualification* (4.8) *are satisfied, then there is an $\varepsilon > 0$ such that for all $(\widetilde{x}, \widetilde{\lambda}) \in I\!B_\varepsilon(\bar{x}, \bar{\lambda})$ the following conditions hold:*

(i) *the optimization problem*

$$\begin{cases} \min\limits_{x \in \mathbb{R}^n} & \varphi(\widetilde{x}) + \langle \nabla\varphi(\widetilde{x}), x - \widetilde{x}\rangle + \dfrac{1}{2}\langle \nabla^2_{xx} L(\widetilde{x}, \widetilde{\lambda})(x - \widetilde{x}), x - \widetilde{x}\rangle \\ \text{subject to} & F(\widetilde{x}) + \nabla F(\widetilde{x})(x - \widetilde{x}) \in \Theta, \end{cases} \tag{4.31}$$

admits a local minimum.

(ii) *The KKT system of* (4.31)*, which can be formulated as*

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla_x L(\widetilde{x}, \widetilde{\lambda}) \\ -F(\widetilde{x}) \end{bmatrix} + \begin{bmatrix} \nabla^2_{xx} L(\widetilde{x}, \widetilde{\lambda}) & \nabla\Phi(\widetilde{x})^* \\ -\nabla F(\widetilde{x}) & 0 \end{bmatrix} \begin{bmatrix} x - \widetilde{x} \\ \lambda - \widetilde{\lambda} \end{bmatrix} + \begin{bmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{bmatrix}, \tag{4.32}$$

has a solution (x, λ) that converges to $(\bar{x}, \bar{\lambda})$ as $(\widetilde{x}, \widetilde{\lambda}) \to (\bar{x}, \bar{\lambda})$.

Proof.

Set $p := (\widetilde{x}, \widetilde{\lambda})$ and $\bar{p} := (\bar{x}, \bar{\lambda})$. Then (4.31) can be equivalently written as the parametric optimization problem

$$\min_{x \in \mathbb{R}^n} \quad \widetilde{\varphi}(x, p) \quad \text{subject to} \quad \widetilde{F}(x, p) \in \Theta, \tag{4.33}$$

where

$$\widetilde{\varphi}(x, p) := \varphi(\widetilde{x}) + \langle \nabla\varphi(\widetilde{x}), x - \widetilde{x}\rangle + \frac{1}{2}\langle \nabla^2_{xx} L(\widetilde{x}, \widetilde{\lambda})(x - \widetilde{x}), x - \widetilde{x}\rangle$$

and $\widetilde{F}(x,p) := F(\widetilde{x}) + \nabla F(\widetilde{x})(x-\widetilde{x})$ for any $(x,p) \in \mathbb{R}^n \times Painlev$. We claim now that $x := \bar{x}$ is a strict local minimizer for (4.33) associated with $\bar{p}$. To this end, observe first that the KKT system of (4.33) associated with $\bar{p}$ has a representation of the form

$$\begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla\varphi(\bar{x}) + \nabla^2_{xx}L(\bar{x},\bar{\lambda})(x-\bar{x}) + \nabla F(\bar{x})^*\lambda \\ \\ -F(\bar{x}) - \nabla F(\bar{x})(x-\bar{x}) \end{bmatrix} + \begin{bmatrix} 0 \\ \\ N_\Theta^{-1}(\lambda) \end{bmatrix}.$$

Clearly, $(\bar{x},\bar{\lambda})$ is a solution to this KKT system, which therefore implies that $\bar{x}$ is a stationary point for (4.33) associated with $\bar{p}$ and that $\bar{\lambda}$ is a Lagrange multiplier associated with $\bar{x}$ for the latter problem. Moreover, it is not hard to observe from this KKT system for (4.33) associated with $p = \bar{p}$ that the set of Lagrange multipliers for the latter problem at $(\bar{x},\bar{p})$ coincides with that of problem (4.1) at $\bar{x}$. Combining these results in the set of Lagrange multipliers for (4.33) at $(\bar{x},\bar{p})$ being $\{\bar{\lambda}\}$. Since we have

$$\nabla^2_{xx}\widetilde{\varphi}(\bar{x},\bar{p}) + \nabla^2_{xx}\langle\bar{\lambda},\widetilde{F}\rangle(\bar{x},\bar{p}) = \nabla^2_{xx}L(\bar{x},\bar{\lambda}) \quad \text{and} \quad \nabla_x\widetilde{F}(\bar{x},\bar{p}) = \nabla F(\bar{x}),$$

and since the second-order sufficient condition (4.15) holds for (4.1) at $(\bar{x},\bar{\lambda})$, we have

$$\begin{aligned} & \big\langle\big[\nabla^2_{xx}\widetilde{\varphi}(\bar{x},\bar{p}) + \nabla^2_{xx}\langle\bar{\lambda},\widetilde{F}\rangle(\bar{x},\bar{p})\big]w,w\big\rangle + \mathrm{d}^2\delta_\Theta(\widetilde{F}(\bar{x},\bar{p}),\bar{\lambda})(\nabla_x\widetilde{F}(\bar{x},\bar{p})w) \\ = \;& \big\langle\nabla^2_{xx}L(\bar{x},\bar{\lambda})w,w\big\rangle + \mathrm{d}^2\delta_\Theta(F(\bar{x}),\bar{\lambda})(\nabla F(\bar{x})w) > 0, \\ & \quad \text{for all } w \in \mathbb{R}^n \setminus \{0\} \;\text{ with }\; \nabla_x\widetilde{F}(\bar{x},\bar{p})w = \nabla F(\bar{x})w \in K_\Theta(F(\bar{x}),\bar{\lambda}). \end{aligned}$$

This tells us that the second-order sufficient condition for (4.33) at $\big((\bar{x},\bar{p}),\bar{\lambda}\big)$ fulfills. Employing now Theorem 3.27(ii) results in $\bar{x}$ being a strict local minimizer for (4.33) for $p = \bar{p}$. We conclude from Proposition 4.9(i) that there is an $\varepsilon > 0$ such that for all $(\widetilde{x},\widetilde{\lambda}) = p \in I\!B_\varepsilon(\bar{p})$ the parametric problem (4.33) admits a local minimum x_p such that $x_p \to \bar{x}$ as $p \to \bar{p}$, and so justifies (i).

Turing now to (ii), observed first that the KKT system of (4.31) is the same is the KKT system of the parametric problem (4.33). By Proposition 4.9(ii), for any $p = (\widetilde{x},\widetilde{\lambda})$ close to $\bar{p} := (\bar{x},\bar{\lambda})$ the generalized equation (4.32) admits a solution (x_p,λ_p). Moreover, (4.26) ensures that $(x_p, \lambda_p) \to (\bar{x}, \bar{\lambda})$ as $p \to \bar{p}$, which completes the proof of (ii). $\square$

The solvability of subproblems in the basic SQP for problems of nonlinear programming was established first by Robinson [43, Theorem 3.1] when in addition to the second-order sufficient condition (4.15) the linear independence constraint qualification and the strict complementarity condition hold. This was improved by Bonnans [1, Proposition 6.3] for this class of problems when the second-order sufficient condition (4.15) is satisfied and the set of Lagrange multipliers is a singleton. Proposition 4.10 extends Bonnans's result for any parabolically regular constrained optimization problems.

We are now in a position to derive the superlinear convergence of the basic SQP method. Assume a sequence x_k converges to $\bar{x}$. We say that convergence is *Q-superlinear* if $\|x_{k+1} - \bar{x}\| = o(\|x_k - \bar{x}\|)$ as $k \to \infty$. In what follows, we drop the letter Q and simply talk about superlinear convergence of a sequence.

Theorem 4.11 (superlinear convergence of the basic SQP method) *Let $(\bar{x}, \bar{\lambda})$ be a solution for the KKT system* (4.4) *and let the basic assumptions* (H1) *and* (H2) *hold. If the second-order sufficient condition* (4.15) *is satisfied at $(\bar{x}, \bar{\lambda})$, the multiplier mapping $M_{\bar{x}}$ is calm at $((0,0), \bar{\lambda})$, and $\Lambda(\bar{x}) = \{\bar{\lambda}\}$, then for any starting point $(x_0, \lambda_0) \in \mathbb{R}^n \times \mathbb{R}^m$ sufficiently close to $(\bar{x}, \bar{\lambda})$, Algorithm* 4.8 *with H_k from* (4.21) *generates an iterative sequence $(x_k, \lambda_k) \in \mathbb{R}^n \times \mathbb{R}^m$ that converges to $(\bar{x}, \bar{\lambda})$, and the rate of convergence is superlinear.*

Proof.

The claimed results come as a direct consequence of [28, Theorem 3.2] for the KKT system (4.4) that can equivalently written as the generalized equation (4.22). So the semistability and hemistability of $(\bar{x}, \bar{\lambda})$ utilized in [28, Theorem 3.2] are the isolated calmness of the solution map S and the property in Proposition 4.10(ii), respectively, which both fulfill here due to Theorem 4.7 and Proposition 4.10, respectively. This completes the proof. $\square$

The above result extends the sharpest currently known local convergence result, obtained in [1, Theorem 5.1], for NLPs to any parabolically regular constrained optimization problems. A similar result was established for second-order cone programming problems in [**?**,30] under the strong regularity of the KKT system (4.4) that is strictly stranger than the imposed assumptions in Theorem 4.11.

The calmness of the multiplier mapping $M_{\bar{x}}$ holds automatically for NLPs due to the Hoffman lemma. Moreover, we showed in [42, Theorem 5.10] that the latter property is satisfied when the convex set Θ in (4.1) is the second-order cone or the cone of positive semidefinite symmetric matrices if the strict complementarity condition is satisfied for the KKT system (4.4). Recall from [4, Definition 4.74] that the *strict complementarity* condition holds for (4.4) at $\bar{x}$ if there is a $\lambda \in \Lambda(\bar{x})$ such that $\lambda \in \operatorname{ri} N_\Theta(F(\bar{x}))$.

The superlinear convergence of a generated sequence (x_k, λ_k) (but not the existence of such a sequence) by the basic SQP method in Theorem 4.11 can be derived from [7, Theorem 6.4] in which the superlinear convergence of the Newton method for the generalized equation

$$0 \in f(x) + F(x)$$

with $f : \mathbb{R}^n \to \mathbb{R}^m$ and $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ was obtained when the mapping $f + F$ is strongly metrically subregular. Adopting the latter result to our framework – the generalized equation (4.22) – the assumed strong metric subregularity is equivalent to the isolated calmness of the solution mapping S from (4.11). Furthermore, by Theorem 4.7 we can equivalently translate the imposed assumptions in Theorem 4.11 as the solution mapping S satisfying isolated calmness and the point $\bar{x}$ being a local minimizer for (4.1). This tells that in contrast with [7] we impose one more assumption, namely the local optimality of $\bar{x}$, and so one may think it would be possible to drop this condition with no harm. However, that can not be happened since [7, Theorem 6.4] assumes the existence of iterations (x_k, λ_k) – solvability of subproblems – which always stay in a neighborhood of $(\bar{x}, \bar{\lambda})$. As the following example, taken from [28, Example 3.3], reveals, the isolated calmness of the solution mapping S alone is not enough to guarantee the solvability of subproblems in our setting and the local optimality of $\bar{x}$ is an essential assumption.

Example 4.12 (failure of solvability of subproblems in the basic SQP method) Consider the nonlinear program

$$\min -\frac{1}{2}x + \frac{1}{6}x^3 \quad \text{subject to} \quad x \geq 0. \tag{4.34}$$

Taking the definition of the solution map S from (4.11) for this problem, we have

$$S(v, w) = \big\{(x, \lambda)|\, v = -x + \frac{1}{2}x^2 + \lambda \quad \text{and} \quad x + w \in N_{\mathbb{R}_-}(\lambda)\big\},$$

where $(v, w) \in \mathbb{R} \times \mathbb{R}$. Set $(\bar{x}, \bar{\lambda}) := (0, 0)$ and observe that $(\bar{x}, \bar{\lambda}) \in S(0, 0)$. It is not hard to see that $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ and that $\bar{x}$ is *not* a local minimizer for problem (4.34) – indeed, $x = 1$ is the unique minimizer for this problem. We show, however, the solution map S is isolated calm at $\big((0, 0), (\bar{x}, \bar{\lambda})\big)$. To this end, we prove that the mapping G from (4.7), adopted for problem (4.34), is strongly metrically subregular at $\big((\bar{x}, \bar{\lambda}), (0, 0)\big)$ using (4.14). Since the mapping G for this framework can be formulated as

$$G(x, \lambda) = \begin{bmatrix} -x + \frac{1}{2}x^2 + \lambda \\ -x \end{bmatrix} + \begin{bmatrix} 0 \\ N_{\mathbb{R}_-}(\lambda) \end{bmatrix},$$

we obtain

$$DG((\bar{x}, \bar{\lambda}), (0, 0))(\xi, \eta) = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} + \begin{bmatrix} 0 \\ DN_{\mathbb{R}_-}(\bar{\lambda}, \bar{x})(\eta) \end{bmatrix}.$$

Since $DN_{\mathbb{R}_-}(\bar{\lambda}, \bar{x})(\eta) = N_{\mathbb{R}_-}(\eta)$, we get

$$(0, 0) \in DG((\bar{x}, \bar{\lambda}), (0, 0))(\xi, \eta) \iff \xi = \eta \quad \text{and} \quad \xi \in N_{\mathbb{R}_-}(\eta),$$

from which we clearly obtain $\xi = \eta = 0$. By (4.14), the mapping G is strongly metrically subregular at $\big((\bar{x},\bar{\lambda}),(0,0)\big)$. We are going to show the generalized equation (4.32) has no solution for all $(\widetilde{x},\widetilde{\lambda})\in I\!B_{1/2}(\bar{x},\bar{\lambda})$ with $\widetilde{x}\neq\bar{x}$. To furnish this, pick $(\widetilde{x},\widetilde{\lambda})\in I\!B_{1/2}(\bar{x},\bar{\lambda})$ with $\widetilde{x}\neq\bar{x}$ and observe that the generalized equation (4.32) for problem (4.34) has a representation of the form

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} -\widetilde{x}+\frac{1}{2}\widetilde{x}^2+\widetilde{\lambda} \\ -\widetilde{x} \end{bmatrix} + \begin{bmatrix} -1+\widetilde{x} & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} x-\widetilde{x} \\ \lambda-\widetilde{\lambda} \end{bmatrix} + \begin{bmatrix} 0 \\ N_{\mathbb{R}_-}(\lambda) \end{bmatrix}.$$

This brings us to the following relationships:

$$\lambda=\frac{1}{2}\widetilde{x}^2-x(\widetilde{x}-1),\ x\lambda=0,\ \ x\geq 0,\ \ \lambda\leq 0.$$

From the second relation, we conclude that either $x=0$ or $\lambda=0$. If the former holds, we deduce from the first equation that $\lambda=\frac{1}{2}\widetilde{x}^2>0$, a contradiction. If the latter holds, we get $x=\dfrac{\frac{1}{2}\widetilde{x}^2}{\widetilde{x}-1}<0$, a contradiction. This justifies that the KKT system associated with subproblems of the basic SQP for problem (4.34) has no solution for such a pair $(\widetilde{x},\widetilde{\lambda})$.

Since we have $\Lambda(\bar{x})=\{\bar{\lambda}\}$, the Robinson constraint qualification (4.10) fulfills. So if the subproblems (4.31), adopted for problem (4.34), has a local optimal solution associated with $(\widetilde{x},\widetilde{\lambda})$, then we will end up having a solution for the generalized equation (4.32), which it is not possible.

We would like to add here that if we assume the metric regularity of the mapping G from (4.7) (remember that $G=S^{-1}$), then the solvability of the SQP method can be ensured by [11, Theorem 6D.2]. Such a result as well as a stronger version of that under the strong metric regularity can be found in [11, Chapter 6]. However, we will show below that in our framework the metric regularity yields the strong metric subregularity of the mapping G, which is, indeed, equivalent to isolated calmness of solution mapping S from (4.11). As Example 4.12 shows, the isolated calmness of S alone does not provide assurance for the existence of the SQP iterations. In order to proceed, we recall that the limiting (Dini-

Hadamard / Mordukhovich) normal cone to $\Omega \subset \mathbb{R}^n$ at $\bar{x}$ is defined by

$$N_\Omega^{L-}(\bar{x}) = \big\{v \in \mathbb{R}^n \mid \exists\, x_k \to \bar{x},\ \ v_k \to v \ \text{ with } \ v_k \in N_\Omega^-(x_k)\big\},$$

and that the coderivative of a set-valued mapping $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is defined by

$$D^*F(\bar{x},\bar{y})(u) := \big\{w \in \mathbb{R}^n \mid (w,-u) \in N_{\mathrm{gph}\,F}(\bar{x},\bar{y})\big\}, \quad u \in \mathbb{R}^m.$$

Proposition 4.13 *Let* $(\bar{x},\bar{\lambda})$ *be a solution for the KKT system* (4.4) *and let the basic assumptions* (H1) *and* (H2) *hold. If the mapping* G *from* (4.7) *is metrically regular around* $((\bar{x},\bar{\lambda}),(0,0))$, *then it is strongly metrically subregular at this point.*

Proof.

According the coderivative criterion for the characterization of metric regularity (see, e.g., [33, Theorem 3.3]), the mapping G enjoys the latter property around $((\bar{x},\bar{\lambda}),(0,0))$ if and only if

$$(0,0) \in D^*G((\bar{x},\bar{\lambda}),(0,0))(w,u) \implies w = 0,\ u = 0. \tag{4.35}$$

It is not hard to see that for any $(w,u) \in \mathbb{R}^n \times \mathbb{R}^m$ the coderivative of G can be calculated by

$$D^*G((\bar{x},\bar{\lambda}),(0,0))(w,u) = \begin{bmatrix} \nabla^2_{xx}L(\bar{x},\bar{\lambda})w - \nabla F(\bar{x})^*u \\ \nabla F(\bar{x})w + D^*N_\Theta^{-1}(\bar{\lambda},F(\bar{x}))(u) \end{bmatrix}.$$

Similarly, it is well-known that (see, e.g., [11, Theorem 4G.1]) G is strongly metrically subregular at $((\bar{x},\bar{\lambda}),(0,0))$ if and only if

$$(0,0) \in DG((\bar{x},\bar{\lambda}),(0,0))(w,u) \implies w = 0,\ u = 0. \tag{4.36}$$

Moreover we have

$$DG((\bar{x},\bar{\lambda}),(0,0))(w,u) = \begin{bmatrix} \nabla^2_{xx}L(\bar{x},\bar{\lambda})w + \nabla F(\bar{x})^* u \\ \\ -\nabla F(\bar{x})w + DN_\Theta^{-1}(\bar{\lambda},F(\bar{x}))(u) \end{bmatrix},$$

where $(w,u) \in \mathbb{R}^n \times \mathbb{R}^m$. By the the basic assumptions (H1) and (H2) and Proposition 4.1(ii), the normal cone mapping N_C is proto-differentiable at $\Phi(\bar{x})$ for $\bar{\lambda}$. The latter together with [54, Theorem 13.57] gives us the derivative-coderivative inclusion

$$DN_\Theta(F(\bar{x}),\bar{\lambda})(\eta) \subset D^*N_\Theta(F(\bar{x}),\bar{\lambda})(\eta) \quad \text{for all } \ \eta \in \mathbb{R}^m.$$

Pick $(w,u) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying the left side of implication (4.36). This and the aforementioned the derivative-coderivative inclusion tell us that $(w,-u)$ satisfies in the equations in the left side of implication (4.35), which results in $w = 0$ and $u = 0$. This confirms (4.36) and thus G is strongly metrically subregular at $((\bar{x},\bar{\lambda}),(0,0))$. $\square$

A similar result as the above result was obtained for NLPs in [11, Lemma 4F.8] by using the characterization of the isolated calmness via the graphical derivative. It was extended later in [13, Corollary 25] for $\mathcal{C}^2$-cone reducible constrained optimization problems using a result obtained by Fusek [14]. The above proposition extends the aforementioned results for any parabolically regular constrained problems using a different proof based on a derivative-coderivative inclusion.

REFERENCES

[1] J. F. Bonnans, *Local analysis of Newton-type methods for variational inequalities and nonlinear programming*, Appl. Math. Optim. 29 (1994), 161–186.

[2] A. Ben-Tal and J. Zowe, *A unified theory of first- and second-order conditions for extremum problems in topological vector spaces*, Math. Program. 19 (1982), 39–76.

[3] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, and C. Sagastizabal, *Numerical Optimization: Theoretical and Practical Aspects*, 2nd ed., Springer-Verlag, Berlin, (2006).

[4] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, (2000).

[5] J. F. Bonnans, R. Cominetti and A. Shapiro, *Second-order optimality conditions based on parabolic second-order tangent sets*, SIAM J. Optim. 9 (1998), 466–492.

[6] J. V. Burke, A. S. Lewis, and M. L. Overton, *Optimal stability and eigenvalue multiplicity,* Found. Comput. Math. 1 (2001), 205–225.

[7] R. Cibulka, A. L. Dontchev and A. Kruger, *Strong metric subregularity of mappings in variational analysis and optimization*, J. Math. Anal. Appl, (2017) .

[8] R. W. Chaney, *On second derivatives for nonsmooth functions*, Nonlinear Anal. 9 (1985), 1189–1209.

[9] N. H. Chieu, L. V. Hien, T. T. A. Nghia, and H. A. Tuan, *Second-order optimality conditions for strong local minimizers via subgradient graphical derivative*, (2019), arXiv:1903.05746.

[10] R. Cominetti, *On pseudo-differentiability*, Trans. Amer. Math. Soc. 324 (1991), 843–847.

[11] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings: A View from Variational Analysis*, 2nd edition, Springer, Dordrecht, (2014).

[12] A. L. Dontchev and R. T. Rockafellar, *Characterizations of Lipschitzian stability in nonlinear programming*, Mathematical Programming With Data Perturbations (A. V. Fiacco, ed.), (1997), 65–82.

[13] C. Ding, D. Sun and L. Zhang, *Characterization of the robust isolated calmness for a class of conic programming problems*, SIAM J. Optim. 27 (2017), 67–90.

[14] P. Fusek, *Isolated zeros of Lipschitzian metrically regular $\mathbb{R}^n$-functions*, Optimization. 49 (2001), 425–446.

[15] H. Gfrerer, *First-order and second-order characterizations of metric subregularity and calmness of constraint set mappings*, SIAM J. Optim. 21 (2011) 1439-1474.

[16] H. Gfrerer *Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints*, SIAM J. Optim. 24 (2014) 898–931.

[17] H. Gfrerer and B. S. Mordukhovich, *Second-order variational analysis of parametric constraint and variational systems*, SIAM J. Optim. 29 (2019), 423–453.

[18] H. Gfrerer and J. V. Outrata, *On computation of generalized derivatives of the normal-cone mapping and their applications* Math. Oper. Res. 41 (2016), 1535–1556.

[19] J.-B. Hiriart-Urruty, *Approximating a second-order directional derivative for nonsmooth convex functions*, SIAM J. Control. Optim. 20 (1982), 381–404.

[20] J.-B. Hiriart-Urruty, *Limiting behavior of the approximate first- and second-order directional derivatives for a convex function*, Nonlinear Anal. 6 (1982), 1309–1326.

[21] Hoffman AJ , *On approximate solutions of systems of linear inequalities*, J. Res. Nat. Bureau Stand. 49(1952), 263–265.

[22] N. T. V. Hang, B. S. Mordukhovich and M. E. Sarabi, *Second-order variational analysis in second-order cone programming*, Math. Program. 180 (2020), 75–116.

[23] A.D. Ioffe, *Variational analysis of a composite function: A formula for the lower second-order epi-derivative*, J. Math. Anal. Appl. 160 (1991) 379–405.

[24] A.D. Ioffe, J.V. Outrata *On metric and calmness qualification conditions in subdifferential calculus*, Set-ValuedAnal. 16 (2008), 199–227.

[25] A.D. Ioffe, *Variational Analysis of Regular Mappings Theory and Applications*, Springer, Cham, Switzerland. (2017).

[26] A. F. Izmailov, *On the analytical and numerical stability of critical Lagrange multipliers*, Comput. Math. Math. Phys. 45 (2005), 930–946.

[27] A. F. Izmailov and M. V. Solodov, *Stabilized SQP revisited*, Math. Program. 133 (2012), 93–120.

[28] A. F. Izmailov and M. V. Solodov, *Newton-Type Methods for Optimization and Variational Problems*, Springer, New York, (2014).

[29] N .H. Josephy, *Newton's method for generalized equations and the PIES energy model*, Ph.D. Dissertation, Department of Industrial Engineering, University of Wisconsin-Madison (1979).

[30] H. Kato and M. Fukushima, *An SQP-type algorithm for nonlinear second-order cone programs*, Optim Lett. 1 (2007) 129–144.

[31] A. Khan, C. Tammer, C. Zălinescu, *Set-valued optimization: An Introduction with Applications*, 2nd edition, Springer, Berlin Heidelberg, (2016).

[32] A. B. Levy, *Second-order epi-derivatives of composite functionals*, Ann. Oper. Res. 101 (2001), 267–281.

[33] B. S. Mordukhovich, *Variational Analysis and Applications*, Springer Monographs in Mathematics, Springer, Cham, Switzerland, (2018).

[34] A. Milzarek, *Numerical Methods and Second Order Theory for Nonsmooth Problems*, Ph.D. Dissertation, University of Munich, (2016).

[35] A. Mohammadi, *First-order chain rule for nonsmooth functions with applications to composite optimization*, Preprint (2020)

[36] A. Mohammadi and B. Mordukhovich, *Variational analysis in normed spaces with applications to constrained optimization*, Submitted (2020) arXiv:2006.00462

[37] A. Mohammadi and M. E. Sarabi, *Twice epi-differentiability of extended-real-valued functions with applications in composite optimization*, SIAM J. Optim. To appear (2020), arXiv:1911.05236.

[38] A. Mohammadi, B. Mordukhovich and M. E. Sarabi, *Variational analysis of composite models with applications to continuous optimization*, Math. Oper. Res. To appear (2019), arXiv:1905.08837.

[39] A. Mohammadi, B. Mordukhovich and M. E. Sarabi, *Parabolic regularity via geometric variational analysis*, Trans. Amer. Math. Soc. To appear (2019), arXiv:1909.00241.

[40] A. Mohammadi, B. Mordukhovich and M. E. Sarabi, *Superlinear convergence of the sequential quadratic method in constrained optimization*, J. Optim. Theory. Appl (2020), arXiv:1910.06894.

[41] B. S. Mordukhovich and M. E. Sarabi, *Critical Multipliers in Variational Systems via Second-Order Generalized Differentiation*, math. program., 169 (2017) 605–648.

[42] B. S. Mordukhovich and M. E. Sarabi, *Criticality of Lagrange Multipliers in Variational Systems*, SIAM J. Optim., 29 (2019) 1524–1557.

[43] S.M. Robinson, *Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear-programming algorithms*, Math. Program. 7 (1974), 1–16.

[44] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), 43–62.

[45] R. A. Poliquin and R. T. Rockafellar, *Prox-regular functions in variational analysis* Trans. Amer. Math. Soc. 348 (1996), 1805–1838.

[46] S. M. Robinson, *Generalized equations and their solutions, I: Basic theory*, Math. Program. Study (1979), 128–141.

[47] T. Rockafellar, *Extended nonlinear programming.* Di Pillo G, Giannessi F, eds. Nonlinear Optimization and Related Topics (Kluwer, Dordrecht, The Netherlands), (2000) 381–399.

[48] R. T. Rockafellar, *Maximal monotone relations and the second derivatives of non-smooth functions*, Ann. Inst. H. Poincaré Analyse Non Linéaire 2 (1985), 167–184.

[49] R. T. Rockafellar, *first- and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc. 307 (1988), 75–108.

[50] R. T. Rockafellar and R. J-B. Wets, *Linear-quadratic programming and optimal control*, SIAM J. Control Opt. 25 (1987) 781—814.

[51] R. T. Rockafellar, *Second-order optimality conditions in nonlinear programming obtained by way of epi-derivatives*, Math. Oper. Res. 14 (1989), 462–484.

[52] R. T. Rockafellar, *Generalized second derivatives of convex functions and saddle functions*, Trans. Amer. Math. Soc. 322 (1990), 51–77.

[53] R. T. Rockafellar and R. J-B. Wets, *A Lagrangian finite-generation technique for solving linear-quadratic problems in stochastic programming*, Math Programming Studies 28 (1986), 63—93.

[54] R. T. Rockafellar and R. J-B. Wets, *Variational Analysis*, Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol. 317, Springer, Berlin, 2006.

[55] A. Shapiro, *On a class of nonsmooth composite functions*, Math. Oper. Res. 28 (2003), 677–692.

[56] M. Torki, *Second-order directional derivatives of all eigenvalues of a symmetric matrix*, Nonlinear Anal. 46 (2001), 1133–1150.

[57] M. Torki, *First- and second-order epi-differentiability in eigenvalue optimization*, J. Math. Anal. Appl. 234 (1999), 391–416.

ABSTRACT

VARIATIONAL ANALYSIS OF COMPOSITE OPTIMIZATION

by

ASHKAN MOHAMMADI

August 2020

Advisor: Dr. Boris. S. Mordukhovich

Major: Mathematics (Applied)

Degree: Doctor of Philosophy

The dissertation is devoted to the study of the first- and second-order variational analysis of the composite functions with applications to composite optimization. By considering a fairly general composite optimization problem, our analysis covers numerous classes of optimization problems such as constrained optimization; in particular, nonlinear programming, second-order cone programming and semidefinite programming(SDP). Beside constrained optimization problems our framework covers many important composite optimization problems such as the extended nonlinear programming and eigenvalue optimization problem. In first-order analysis we develop the exact first-order calculus via both subderivative and subdifferential. For the second-order part we develop calculus rules via second subderivatives (which was a long standing open problem). Furthermore, we establish twice epi-differentiability of composite functions. Then we apply our results to composite optimization problem to obtain first- and second-order order optimality conditions under the weakest constraint qualification, the metric subregularity constraint qualification. Finally we apply our results to verify the super linear convergence in SQP methods for constrained optimization.

AUTOBIOGRAPHICAL STATEMENT

Ashkan Mohammadi

Education

- *Ph.D. in Applied Mathematics*, Wayne State University, August (2020)
- *M.S. in Statistics*, Wayne State University, September (2019)
- *M.S. in Applied Mathematics*, University of Tehran, Tehran, Iran (2014)
- *B.S. in Applied Mathematics*, University of Tehran, Tehran, Iran (2012)

Publications and Preprints

- *Twice epi-differentiability of extended real-valued functions with applications in composite optimization*, *SIAM journal of Optimization. To appear* (2019) arXiv:1911.05236 (with E. Sarabi)
- *Variational analysis in normed spaces with applications to constrained optimization*, Submitted (2020) arXiv:2006.00462 (with B. Mordukhovich)
- *Superlinear convergence of the sequential quadratic method in constrained optimization*, J. Optim. Theory. Appl (2020) arXiv:1910.06894 (with B. Mordukhovich, E. Sarabi)
- *Parabolic regularity in geometric variational analysis*, Trans. Amer. Math. Soc. To appear, (2019) arXiv:1909.00241 (with B. Mordukhovich, E. Sarabi)
- *Variational analysis of composite models with applications to continuous optimization*, Math. Oper. Res. To appear, (2019) arXiv:1905.08837 (with B. Mordukhovich, E. Sarabi)