

# STABILITY OF KKT SYSTEMS AND SUPERLINEAR CONVERGENCE OF THE SQP METHOD UNDER PARABOLIC REGULARITY

ASHKAN MOHAMMADI<sup>1</sup> BORIS S. MORDUKHOVICH<sup>2</sup> and M. EBRAHIM SARABI<sup>3</sup>

**Abstract.** This paper pursues a two-fold goal. Firstly, we aim to derive novel second-order characterizations of important robust stability properties of perturbed Karush-Kuhn-Tucker systems for a broad class of constrained optimization problems generated by parabolically regular sets. Secondly, the obtained characterizations are applied to establish well-posedness and superlinear convergence of the basic sequential quadratic programming method to solve parabolically regular constrained optimization problems.

**Key words.** Variational analysis, constrained optimization, KKT systems, metric subregularity and calmness, critical and noncritical multipliers, SQP methods, superlinear convergence

**Mathematics Subject Classification (2000)** 90C31, 65K99, 49J52, 49J53

## 1 Introduction

This paper is devoted to the second-order variational analysis and numerical applications for a general class of constrained optimization problems formulated in Section 2. It has been well recognized in optimization theory and its applications that second-order analysis concerning both qualitative/theoretical and quantitative/numerical aspects of constrained optimization requires certain second-order regularity conditions. In this paper we explore a novel one, which is the *parabolic regularity* of the constraint set at the point in question.

Actually the notion of parabolic regularity for extended-real-valued functions was first formulated by Rockafellar and Wets [29, Definition 13.65], but was not studied there except for fully amenable compositions with the immediate application to the unconstrained format of second-order optimality; see also [3, Subsection 3.3.5]. We are not familiar with any other publications where parabolic regularity was either further investigated, or applied to structural problems of constrained optimization.

Very recently [19], a systematic study of parabolically regular sets (both convex and non-convex) has been conducted in our paper, where we reveal a fundamental role of this concept in second-order variational analysis and its important applications to optimization theory. In particular, it is shown in [19] that parabolic regularity is *more general* than known second-order regularity notions, is *preserved* under various operations on sets, ensures—among other significant results—*twice epi-differentiability* of set indicator functions, *precise/equality type calculi* of second subderivatives and second-order tangents, etc. Furthermore, parabolic regularity married to the *metric subregularity constraint qualification* allows us to efficiently deal with *constraint*

---

<sup>1</sup>Department of Mathematics, Wayne State University, Detroit, MI 48202, USA (ashkan.mohammadi@wayne.edu). Research of this author was partly supported by the US National Science Foundation under grant DMS-1808978 and by the US Air Force Office of Scientific Research under grant #15RT0462.

<sup>2</sup>Department of Mathematics, Wayne State University, Detroit, MI 48202, USA (boris@math.wayne.edu). Research of this author was partly supported by the US National Science Foundation under grants DMS-1512846 and DMS-1808978, by the US Air Force Office of Scientific Research under grant #15RT0462, and by the Australian Research Council Discovery Project DP-190100555.

<sup>3</sup>Department of Mathematics, Miami University, Oxford, OH 45065, USA (sarabim@miamioh.edu).

*systems* in optimization with obtaining *computation formulas* for major second-order generalized derivatives, which are instrumental, in particular, for the development and justification of primal-dual numerical algorithms of optimization. Let us also mention in the list of achievements of [19] that parabolic regularity leads us to deriving *no-gap second-order optimality conditions* in constrained optimization and establishing a *quadratic growth* of *augmented Lagrangians* that are highly important for subsequent numerical applications.

In this paper we continue the study and applications of parabolic regularity in the new directions that were not explored in [19]. Our *first goal* is to obtain complete *second-order characterizations* of some major stability properties and their *robust* versions for perturbed *Karush-Kuhn-Tucker* (KKT) systems associated with optimization problems whose constraints are described by parabolically regular sets. We mainly concentrate here on two equivalent (up to taking the inverse mapping) properties known as *strong metric subregularity* and *isolated calmness* together with their robust counterparts. Using parabolic regularity, we establish novel characterizations of these properties in the general KKT setting under consideration that cover known results for nonlinear programs (NLPs) while contain essentially new ingredients for non-polyhedral constraint systems. Note that the approach developed in this paper allows us to significantly simplify the proofs of the known results for NLPs. Observe also that the recent characterizations of *noncritical Lagrange multipliers* obtained in [25] for nonpolyhedral systems play a significant role in deriving the aforementioned stability results for perturbed KKT systems and their numerical applications discussed below.

The *second goal* of this paper is to develop applications of the obtained second-order characterizations of robust stability for KKT systems to the *sequential quadratic programming* (SQP) method for solving optimization problems with parabolically regular constraints. In the case of NLPs, the sharpest result for the method was achieved by Bonnans [1] who established the *superlinear convergence* of the basic SQP method under the uniqueness of Lagrange multipliers and the fulfillment of the classical second-order sufficient condition. We now significantly extend this result to the general class of parabolically regular constrained optimization problems under an appropriate version of the second-order sufficiency with adding a new condition, which automatically holds for NLPs.

The rest of the paper is organized as follows. Section 2 presents and discusses the *basic concepts* of first-order and second-order variational analysis and generalized differentiation that are systematically used in the subsequent material. In Section 3 we define the crucial notion of *set parabolic regularity*, formulate important results for it taken from [19] and needed in this paper, and then derive some new assertions, which are required for our main achievements below.

In Section 4 we first reveal close relationships between *noncriticality* of Lagrange multipliers and *strong metric subregularity* of the KKT mapping associated with the constrained optimization problem under consideration. It turns out to be instrumental for deriving new second-order characterizations of the isolated calmness property for the solution map to the canonically perturbed KKT system. The latter result is used further in this section to establish *robust stability* of the extended KKT system associated with *general perturbations* of the original constrained optimization problem. In the case of *canonical perturbations*, it gives us the property of solution maps to KKT systems that is labeled as *robust isolated calmness* in [6].

Section 5 is devoted to the application of the obtained stability results to establishing local *superlinear convergence* of the *basic SQP method* to numerically solve optimization problems with *parabolically regular constraints*. The stability result under general perturbations developed in Section 4 plays a crucial role in justifying the *solvability* of subproblems in the basic SQP method without which the SQP iterations are not well posed. Achieving it and imposing the obtained second-order sufficient condition together with the uniqueness of Lagrange multipliers and the calmness of the multiplier mappings (the latter is automatic for NLPs), we prove superlinear

convergence of the basic SQP method and thus properly extend Bonnans' remarkable result to a significantly more general class of nonpolyhedral problems in constrained optimization.

The concluding Section 6 summarizes the main results obtained in the paper and discusses some directions of the future research.

## 2 Basic Concepts and Discussions

The notation and terminology in this paper are standard in variational analysis and generalized differentiation; see, e.g., [22, 29]. For the reader's convenience and notational unification we use as a rule small Greek letters to signify scalar and extended-real-valued (i.e., with values in the extended real line  $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ ) functions, small Latin letters for vectors and single-valued mappings, and capital letters for sets and set-valued mappings. For a nonempty subset  $\Omega$  of  $\mathbb{R}^n$ , the symbols  $\text{ri } \Omega$ ,  $\text{cl } \Omega$ , and  $\Omega^*$  stand for the relative interior, closure, and polar of  $\Omega$ , respectively. We write  $x \xrightarrow{\Omega} \bar{x}$  to indicate that  $x \rightarrow \bar{x}$  with  $x \in \Omega$ . The distance between a point  $x \in \mathbb{R}^n$  and a set  $\Omega \subset \mathbb{R}^n$  is denoted by  $\text{dist}(x; \Omega)$ . The indicator function of  $\Omega$  is defined by  $\delta_\Omega(x) := 0$  for  $x \in \Omega$  and  $\delta_\Omega(x) := \infty$  otherwise. The notation  $\mathbb{B}$  stands for the closed unit ball in the space in question, while  $\mathbb{B}_r(x) := x + r\mathbb{B}$  mean the closed ball centered at  $x$  with radius  $r > 0$ . The symbol  $x = o(t)$  with  $x \in \mathbb{R}^n$  and  $t > 0$  tells us that  $\|x\|/t \rightarrow 0$  as  $t \downarrow 0$ . Let us also mention that the notation  $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$  indicates the possibility of set values  $F(x) \subset \mathbb{R}^m$  (including the empty set  $\emptyset$ ) of  $F$  for some  $x \in \mathbb{R}^n$ , in contrast to the standard notation  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  for single-valued mappings as well as extended-real-valued functions.

Now we recall those constructions of generalized differentiation for sets, set-valued mappings, and extended-real-valued functions that are broadly used in what follows. By generalized differential constructions for sets we understand tangent and normal cones, which are closely related to (in fact induce) the corresponding generalized derivatives for mappings and functions.

Given a set  $\Omega \subset \mathbb{R}^n$ , the (Bouligand-Severi) *tangent cone*  $T_\Omega(\bar{x})$  to  $\Omega$  at  $\bar{x} \in \Omega$  is defined by

$$T_\Omega(\bar{x}) := \{w \in \mathbb{R}^n \mid \exists t_k \downarrow 0, \quad w_k \rightarrow w \quad \text{as } k \rightarrow \infty \text{ with } \bar{x} + t_k w_k \in \Omega\}. \quad (2.1)$$

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

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

The set  $\Omega$  is *geometrically derivable* at  $\bar{x}$  if every tangent vector  $w$  to  $\Omega$  at  $\bar{x}$  is derivable.

The (Mordukhovich limiting) *normal cone* to  $\Omega$  at  $\bar{x} \in \Omega$  is given by

$$N_\Omega(\bar{x}) = \{v \in \mathbb{R}^n \mid \exists x_k \rightarrow \bar{x}, \quad v_k \rightarrow v \quad \text{with } v_k \in T_\Omega^*(x_k)\}, \quad (2.2)$$

where  $T_\Omega^*(x) := \{v \in \mathbb{R}^n \mid \langle v, w \rangle \leq 0 \text{ for all } w \in T_\Omega(x)\}$  is the dual cone of (2.1). If  $\Omega$  is convex, (2.1) and (2.2) reduce, respectively, to the tangent and normal cones of convex analysis. Due to the intrinsic nonconvexity of the normal cone even in simple settings (i.e., for the graph of the function  $|x|$  at  $(0, 0) \in \mathbb{R}^2$ ), and thus the normal cone (2.1) is not tangentially generated.

Consider next a set-valued mapping/multifunction  $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$  with its domain and graph given, respectively, by

$$\text{dom } F := \{x \in \mathbb{R}^n \mid F(x) \neq \emptyset\} \quad \text{and} \quad \text{gph } F := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y \in F(x)\}.$$

Then we define the two generalized derivatives generated by the tangent cone (2.1) and the normal cone (2.2) to the graph of  $F$  in the following way, respectively. The *graphical derivative*

to  $F$  at  $(\bar{x}, \bar{y}) \in \text{gph } F$  is

$$DF(\bar{x}, \bar{y})(u) := \{v \in \mathbb{R}^m \mid (u, v) \in T_{\text{gph } F}(\bar{x}, \bar{y})\}, \quad u \in \mathbb{R}^n, \quad (2.3)$$

while the *coderivative* of  $F$  at this point is

$$D^*F(\bar{x}, \bar{y})(v) := \{u \in \mathbb{R}^n \mid (u, -v) \in N_{\text{gph } F}(\bar{x}, \bar{y})\}, \quad v \in \mathbb{R}^m. \quad (2.4)$$

Note that the coderivative (2.4) is not dual to the graphical derivative (2.3), since the values of (2.4) are often nonconvex. Despite (actually due to) its nonconvexity, the coderivative enjoys *full calculus* based on the variational/extremal principles of variational analysis; see [21, 22] for more details. Furthermore, in its terms we get complete pointwise characterizations of fundamental *well-posedness* properties of set-valued mappings related to metric regularity, linear openness/covering, and robust Lipschitzian stability known as the Mordukhovich coderivative criteria; see [20, 21, 29]. We present below the one for metric regularity used in this paper. Recall that a mapping  $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$  is *metrically regular* around  $(\bar{x}, \bar{y}) \in \text{gph } F$  if there exist neighborhoods  $U$  of  $\bar{x}$  and  $V$  of  $\bar{y}$  together with a number  $\kappa \geq 0$  ensuring the estimate

$$\text{dist}(x; F^{-1}(y)) \leq \kappa \text{dist}(y; F(x)) \quad \text{for all } (x, y) \in U \times V. \quad (2.5)$$

The *coderivative criterion* for metric regularity says that a closed-graph mapping  $F$  is metrically around the point  $(\bar{x}, \bar{y})$  if and only if we have the kernel condition

$$\ker D^*F(\bar{x}, \bar{y}) := \{v \in \mathbb{R}^m \mid 0 \in D^*F(\bar{x}, \bar{y})(v)\} = \{0\}; \quad (2.6)$$

see [22, Theorem 3.3] for a simplified proof and discussions in finite dimensions.

Along with the usage of (2.5) and (2.6), we largely employ the following significant modifications and relaxations of (2.5). The mapping  $F$  is said to be *metrically subregular* at  $(\bar{x}, \bar{y})$  if the validity of (2.5) is required only when  $y = \bar{y}$ , i.e., if there exist  $\kappa \geq 0$  and a neighborhood  $U$  of  $\bar{x}$  ensuring the validity of the distance estimate

$$\text{dist}(x; F^{-1}(\bar{y})) \leq \kappa \text{dist}(\bar{y}; F(x)) \quad \text{for all } x \in U. \quad (2.7)$$

Furthermore,  $F$  is *strongly metrically subregular* at  $(\bar{x}, \bar{y})$  if (2.7) holds with  $F^{-1}(\bar{y}) = \{\bar{x}\}$ , i.e.,

$$\|x - \bar{x}\| \leq \kappa \text{dist}(\bar{y}; F(x)) \quad \text{for all } x \in U. \quad (2.8)$$

It has been well realized in variational analysis that the subregularity properties (2.7) and (2.8) are equivalent to the following calmness and isolated calmness counterparts for inverse mappings; see, e.g., [8, 22]. Recall that  $F: \mathbb{R}^n \rightrightarrows \mathbb{R}^m$  is *calm* at  $(\bar{x}, \bar{y}) \in \text{gph } F$  if there exist  $\ell \geq 0$  and neighborhoods  $U$  of  $\bar{x}$  and  $V$  of  $\bar{y}$  such that

$$F(x) \cap V \subset F(\bar{x}) + \ell\|x - \bar{x}\|\mathbb{B} \quad \text{for all } x \in U. \quad (2.9)$$

We say that the mapping  $F$  has the *isolated calmness* property at  $(\bar{x}, \bar{y})$  if  $F(\bar{x}) = \{\bar{y}\}$  in (2.9), i.e., there exist  $\ell \geq 0$  and neighborhoods  $U$  of  $\bar{x}$  and  $V$  of  $\bar{y}$  for which

$$F(x) \cap V \subset \{\bar{y}\} + \ell\|x - \bar{x}\|\mathbb{B} \quad \text{for all } x \in U. \quad (2.10)$$

Although properties (2.7)–(2.10) are much less investigated and applied in the literature in comparison with metric regularity and related two-point properties of multifunctions, there is an increasing number of publications dealing with such one-point formulations; see, e.g., [4–8, 11–13, 15, 18, 19, 22, 28, 31, 33] and the references therein. In this paper we pay the main

attention to the properties of strong metric subregularity and isolated calmness for KKT systems with the subsequent applications to the SQP method.

Given next an extended-real-valued function  $\varphi: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  with the domain  $\text{dom } \varphi := \{x \in \mathbb{R}^n \mid \varphi(x) < \infty\}$  and the epigraph  $\text{epi } \varphi := \{(x, \mu) \in \mathbb{R}^{n+1} \mid \mu \geq \varphi(x)\}$ , suppose in what follows that  $\varphi$  is lower semicontinuous (l.s.c.) around the reference point. The *subdifferential* (collection of subgradients) of  $\varphi$  at  $\bar{x}$  is defined by

$$\partial\varphi(\bar{x}) := \{v \in \mathbb{R}^n \mid (v, -1) \in N_{\text{epi } \varphi}(\bar{x}, \varphi(\bar{x}))\} \quad (2.11)$$

via the normal cone (2.2) to the epigraph of  $\varphi$  at  $(\bar{x}, \varphi(\bar{x}))$ . If  $\varphi$  is convex, the subdifferential (2.11) reduces to the classical subgradient set of convex analysis.

We proceed further with the second-order constructions of generalized differentiation needed in this paper. Given  $\varphi: \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , consider the parametric family of *second-order difference quotients* for  $\varphi$  at  $\bar{x} \in \text{dom } \varphi$  relative to  $\bar{v} \in \mathbb{R}^n$  by

$$\Delta_t^2\varphi(\bar{x}, \bar{v})(w) := \frac{\varphi(\bar{x} + tw) - \varphi(\bar{x}) - t\langle \bar{v}, w \rangle}{\frac{1}{2}t^2} \quad \text{with } w \in \mathbb{R}^n \text{ and } t > 0.$$

The *second subderivative* of  $\varphi$  at  $\bar{x}$  for  $\bar{v}$  and  $w$  is defined by

$$d^2\varphi(\bar{x}, \bar{v})(w) := \liminf_{\substack{t \downarrow 0 \\ w' \rightarrow w}} \Delta_t^2\varphi(\bar{x}, \bar{v})(w'). \quad (2.12)$$

Then  $\varphi$  is said to be *twice epi-differentiable* at  $\bar{x}$  for  $\bar{v}$  if the sets  $\text{epi } \Delta_t^2\varphi(\bar{x}, \bar{v})$  converge (in the sense of the convergence of the corresponding distance functions) to  $\text{epi } d^2\varphi(\bar{x}, \bar{v})$  as  $t \downarrow 0$ . If in addition the second subderivative is a proper function (i.e., does not take the value  $-\infty$  and is finite at some point), we say that  $\varphi$  is *properly twice epi-differentiable* at  $\bar{x}$  for  $\bar{v}$ . By [29, Proposition 7.2], the twice epi-differentiability of  $f$  at  $\bar{x}$  for  $\bar{v}$  can be understood equivalently as for every  $w \in \mathbb{R}^n$  and every sequence  $t_k \downarrow 0$ , there exists a sequence  $w_k \rightarrow w$  such that

$$\Delta_{t_k}^2\varphi(\bar{x}, \bar{v})(w_k) \rightarrow d^2\varphi(\bar{x}, \bar{v})(w).$$

### 3 Parabolic Regularity and Related Issues

We begin with the formulation of the fundamental parabolic regularity property for sets.

**Definition 3.1 (parabolic regularity of sets).** A set  $\Omega \subset \mathbb{R}^n$  is called **PARABOLICALLY REGULAR** at  $\bar{x} \in \Omega$  for a vector  $\bar{v} \in \mathbb{R}^n$  if for each  $w$  with  $d^2\delta_\Omega(\bar{x}, \bar{v})(w) < \infty$  there exist, among sequences  $t_k \downarrow 0$  and  $w_k \rightarrow w$  satisfying

$$\Delta_{t_k}^2\delta_\Omega(\bar{x}, \bar{v})(w_k) \rightarrow d^2\delta_\Omega(\bar{x}, \bar{v})(w) \quad \text{as } k \rightarrow \infty,$$

those with the additional property that

$$\limsup_{k \rightarrow \infty} \frac{\|w_k - w\|}{t_k} < \infty.$$

It is shown in [19] that this property holds for sets appearing frequently in constrained optimization such as polyhedral convex sets, the second-order/Lorentz/ice-cream cone, and the cone of symmetric positive semidefinite matrices. More broadly, any  $\mathcal{C}^2$ -cone *reducible set* in the sense of [3, Definition 3.135] is always parabolically regular; see [19, Proposition 5.14]. However, the converse statement fails even for simple examples as demonstrated in [19, Example 5.16]. It

is worth mentioning that, contrary to the  $\mathcal{C}^2$ -cone reducibility, the parabolic regularity enjoys good calculus including intersection rule under reasonable assumptions.

The *second-order tangent set* to  $\Omega$  at  $\bar{x} \in \Omega$  for a tangent vector  $w \in T_\Omega(\bar{x})$  from (2.1) is

$$T_\Omega^2(\bar{x}, w) = \{u \in \mathbb{R}^n \mid \exists t_k \downarrow 0, u_k \rightarrow u \text{ as } k \rightarrow \infty \text{ with } \bar{x} + t_k w + \frac{1}{2} t_k^2 u_k \in \Omega\}. \quad (3.1)$$

A set  $\Omega$  is said to be *parabolically derivable* at  $\bar{x}$  for  $w$  if  $T_\Omega^2(\bar{x}, w) \neq \emptyset$  and for each  $u \in T_\Omega^2(\bar{x}, w)$  there exist  $\varepsilon > 0$  and an arc  $\xi : [0, \varepsilon] \rightarrow \Omega$  with  $\xi(0) = \bar{x}$ ,  $\xi'_+(0) = w$ , and  $\xi''_+(0) = u$ , where

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

The latter property is satisfied in rather general settings; see [29] and [19]. Recall further that the *critical cone* for a closed set  $\Omega$  at  $x$  for  $v$  with  $(x, v) \in \text{gph } N_\Omega$  is defined by

$$K_\Omega(x, v) = T_\Omega(x) \cap \{v\}^\perp. \quad (3.2)$$

Now we are ready to present an important result, which actually summarizes several achievements from [19]. To proceed, recall that a set-valued mapping  $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$  is *proto-differentiable* at  $\bar{x}$  for  $\bar{y}$  with  $(\bar{x}, \bar{y}) \in \text{gph } F$  if the set  $\text{gph } F$  is geometrically derivable at  $(\bar{x}, \bar{y})$ . When this condition holds for  $F$ , we refer to  $DF(\bar{x}, \bar{y})$  as the *proto-derivative* of  $F$  at  $\bar{x}$  for  $\bar{y}$ .

**Theorem 3.2 (proto-differentiability of normal cone mappings).** *Let  $\Omega \subset \mathbb{R}^n$  be closed and convex set, and let  $\bar{v} \in N_\Omega(\bar{x})$  with  $\bar{x} \in \Omega$ . Assume that the set  $\Omega$  is parabolically derivable at  $\bar{x}$  for every vector  $w \in K_\Omega(\bar{x}, \bar{v})$ , and that  $\Omega$  is parabolically regular at  $\bar{x}$  for  $\bar{v}$ . Then the following equivalent conditions are satisfied:*

- (i) *The indicator function  $\delta_\Omega$  is twice epi-differentiable at  $\bar{x}$  for  $\bar{v}$ .*
- (ii) *The normal cone mapping  $N_\Omega$  is proto-differentiable at  $\bar{x}$  for  $\bar{v}$  with the proto-derivative*

$$DN_\Omega(\bar{x}, \bar{v})(w) = \partial \left[ \frac{1}{2} d^2 \delta_\Omega(\bar{x}, \bar{v}) \right](w) \quad \text{for all } w \in \mathbb{R}^n. \quad (3.3)$$

Furthermore, we have that  $\text{dom } d^2 \delta_\Omega(\bar{x}, \bar{v}) = K_\Omega(\bar{x}, \bar{v})$ , and that the second subderivative  $d^2 \delta_\Omega(\bar{x}, \bar{v})$  is a proper convex function on  $\mathbb{R}^n$ .

**Proof.** The fulfillment of both conditions (i) and (ii) and their equivalence are established in [19, Theorem 3.5] and [19, Corollary 3.6]. The claimed formula for the domain of the second epi-derivative is taken from [19, Theorem 3.3]. Finally, it follows from [29, Proposition 13.20] that the second subderivative  $d^2 \delta_\Omega(\bar{x}, \bar{v})$  is a proper convex function.  $\square$

Exploiting the imposed convexity of the set  $\Omega$  allows us to derive the following enhanced representations of the proto-derivative  $DN_\Omega(\bar{x}, \bar{v})$  from Theorem 3.2 at the origin.

**Corollary 3.3 (proto-derivatives of normal cones at the origin).** *Let  $\Omega \subset \mathbb{R}^n$  satisfy all the assumptions of Theorem 3.2. Then we have the formulas*

$$DN_\Omega(\bar{x}, \bar{v})(0) = K_\Omega(\bar{x}, \bar{v})^* = N_{K_\Omega(\bar{x}, \bar{v})}(0). \quad (3.4)$$

**Proof.** The convexity of  $\Omega$ , and hence of  $K_\Omega(\bar{x}, \bar{v})$ , allows us to deduce the second equality claimed in (3.4). We proceed now with justifying the first equality therein. It follows from the representation in (3.3) that

$$DN_\Omega(\bar{x}, \bar{v})(0) = \partial \left[ \frac{1}{2} d^2 \delta_\Omega(\bar{x}, \bar{v}) \right](0).$$

Pick any  $u \in DN_{\Omega}(\bar{x}, \bar{v})(0)$  and thus get  $u \in \partial[\frac{1}{2}d^2\delta_{\Omega}(\bar{x}, \bar{v})](0)$ . Since the second subderivative  $d^2\delta_{\Omega}(\bar{x}, \bar{v})$  is a convex function with  $d^2\delta_{\Omega}(\bar{x}, \bar{v})(0) = 0$ , we obtain

$$\langle u, w - 0 \rangle \leq \frac{1}{2}d^2\delta_{\Omega}(\bar{x}, \bar{v})(w) - \frac{1}{2}d^2\delta_{\Omega}(\bar{x}, \bar{v})(0) = \frac{1}{2}d^2\delta_{\Omega}(\bar{x}, \bar{v})(w) \quad \text{for all } w \in \mathbb{R}^n.$$

Theorem 3.2 tells us that  $\text{dom } d^2\delta_{\Omega}(\bar{x}, \bar{v}) = K_{\Omega}(\bar{x}, \bar{v})$ . To proceed further, pick  $t > 0$  and  $w \in K_{\Omega}(\bar{x}, \bar{v})$  and, with taking into account that the second subderivative  $d^2\delta_{\Omega}(\bar{x}, \bar{v})$  is positive homogeneous of degree 2, arrive at  $d^2\delta_{\Omega}(\bar{x}, \bar{v})(tw) = t^2d^2\delta_{\Omega}(\bar{x}, \bar{v})(w)$ , which results in

$$t\langle u, w \rangle \leq \frac{1}{2}d^2\delta_{\Omega}(\bar{x}, \bar{v})(tw) = \frac{1}{2}t^2d^2\delta_{\Omega}(\bar{x}, \bar{v})(w).$$

This along with the choice of  $w \in \text{dom } d^2\delta_{\Omega}(\bar{x}, \bar{v})$  implies that  $\langle u, w \rangle \leq 0$  and hence gives us  $u \in K_{\Omega}(\bar{x}, \bar{v})^*$ . It verifies the inclusion “ $\subset$ ” for the first equality in (3.4).

Assuming now that  $u \in K_{\Omega}(\bar{x}, \bar{v})^*$ , we get  $\langle u, w \rangle \leq 0$  for all  $w \in K_{\Omega}(\bar{x}, \bar{v})$ . This leads us to

$$\langle u, w \rangle \leq 0 \leq \frac{1}{2}d^2\delta_{\Omega}(\bar{x}, \bar{v})(w) = \frac{1}{2}d^2\delta_{\Omega}(\bar{x}, \bar{v})(w) - \frac{1}{2}d^2\delta_{\Omega}(\bar{x}, \bar{v})(0) \quad \text{for all } w \in K_{\Omega}(\bar{x}, \bar{v}).$$

If  $w \notin K_{\Omega}(\bar{x}, \bar{v})$ , then it follows from Theorem 3.2 that  $d^2\delta_{\Omega}(\bar{x}, \bar{v})(w) = \infty$ , which clearly ensures that the above inequality holds for such  $w$ . Then using the convexity of second subderivative  $d^2\delta_{\Omega}(\bar{x}, \bar{v})$  and the subdifferential construction of convex analysis, we arrive at  $u \in \partial[\frac{1}{2}d^2\delta_{\Omega}(\bar{x}, \bar{v})](0)$  and thus deduce from (3.3) that  $u \in DN_{\Omega}(\bar{x}, \bar{v})(0)$ . This verifies the inclusion “ $\supset$ ” for the first equality in (3.4) and hence completes the proof of the corollary.  $\square$

Next we start considering the *constrained optimization* problem given by

$$\min_{x \in \mathbb{R}^n} \varphi_0(x) \quad \text{subject to} \quad f(x) \in \Theta, \quad (3.5)$$

where  $\varphi_0: \mathbb{R}^n \rightarrow \mathbb{R}$  and  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$  are  $\mathcal{C}^2$ -smooth mappings around the reference points, and where  $\Theta$  is a closed and convex set in  $\mathbb{R}^m$ . Define its *KKT system* associated with (3.5) by

$$\nabla_x L(x, \lambda) = \nabla \varphi_0(x) + \nabla f(x)^* \lambda = 0, \quad \lambda \in N_{\Theta}(f(x)), \quad (3.6)$$

where  $L(x, \lambda) := \varphi_0(x) + \langle f(x), \lambda \rangle$  is the *Lagrangian* associated with (3.5) and  $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ . Given a point  $\bar{x} \in \mathbb{R}^n$ , denote the set of *Lagrange multipliers* at  $\bar{x}$  by

$$\Lambda(\bar{x}) := \{\lambda \in \mathbb{R}^m \mid \nabla_x L(\bar{x}, \lambda) = 0, \lambda \in N_{\Theta}(f(\bar{x}))\}. \quad (3.7)$$

Having  $(\bar{x}, \bar{\lambda})$  as a solution to the KKT system (3.6) 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 (3.6) in the sense that it satisfies the condition

$$0 \in \partial(\varphi_0 + (\delta_{\Theta} \circ g))(\bar{x}), \quad (3.8)$$

where the subdifferential of the generally nonconvex composition in (3.8) is taken from (2.11).

Now it is time to formulate the *standing assumptions* on the closed and convex set in (3.5) that are used in the rest of the paper unless otherwise stated.

**(H1)** For every  $\lambda \in \Lambda(\bar{x})$  the set  $\Theta$  is *parabolically derivable* at  $f(\bar{x})$  for all the vectors in the form  $\nabla f(\bar{x})w \subset K_{\Theta}(f(\bar{x}), \lambda)$  with some  $w \in \mathbb{R}^n$ .

**(H2)** The set  $\Theta$  is *parabolically regular* at  $f(\bar{x})$  for every  $\lambda \in \Lambda(\bar{x})$ .

As mentioned above, both assumptions (H1) and (H2) hold for important classes of convex sets that naturally and frequently appear in constrained optimization; see, in particular, [19, Proposition 5.14]. Imposing these assumptions on the set  $\Theta$  opens the door for deriving enhanced second-order optimality conditions for general problems of constrained optimization. The next result is obtained in [19, Theorem 6.1] as a consequence of the comprehensive calculus rules achieved therein for the second subderivatives.

**Theorem 3.4 (no-gap second-order optimality conditions).** *Let  $\bar{x}$  be a feasible solution to problem (3.5) under the validity of the standing assumptions (H1) and (H2) on  $\Theta$  with  $\bar{v} := -\nabla\varphi_0(\bar{x})$ . Assume further that the mapping  $x \mapsto f(x) - \Theta$  is metrically subregular at  $(\bar{x}, 0)$  and consider the set of feasible solutions to (3.5) given by*

$$\Omega := \{x \in \mathbb{R}^n \mid f(x) \in \Theta\}. \quad (3.9)$$

*The following second-order optimality conditions for problem (3.5) hold:*

**(i)** *If  $\bar{x}$  is a local minimum of (3.5), then the second-order necessary condition*

$$\max_{\lambda \in \Lambda(\bar{x})} \{\langle \nabla_{xx}^2 L(\bar{x}, \lambda)w, w \rangle + d^2 \delta_\Theta(f(\bar{x}), \lambda) (\nabla f(\bar{x})w) \} \geq 0$$

*is satisfied for all the critical vectors  $w \in K_\Omega(\bar{x}, \bar{v})$ .*

**(ii)** *The fulfillment of the no-gap second-order condition*

$$\max_{\lambda \in \Lambda(\bar{x})} \{\langle \nabla_{xx}^2 L(\bar{x}, \lambda)w, w \rangle + d^2 \delta_\Theta(f(\bar{x}), \lambda) (\nabla f(\bar{x})w) \} > 0 \quad \text{for all } w \in K_\Omega(\bar{x}, \bar{v}) \setminus \{0\}$$

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

$$\varphi(x) \geq \varphi(\bar{x}) + \frac{\ell}{2} \|x - \bar{x}\|^2 \quad \text{for all } x \in \mathbb{B}_\varepsilon(\bar{x}) \quad \text{with } \varphi := \varphi_0 + (\delta_\Theta \circ f),$$

*which implies that  $\bar{x}$  is a strict local minimizer for (3.5).*

Fix  $\bar{x} \in \mathbb{R}^n$  and define the *multiplier mapping*  $M_{\bar{x}}: \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^m$  associated with  $\bar{x}$  by

$$M_{\bar{x}}(v, w) := \{\lambda \in \mathbb{R}^m \mid (v, w) \in G(\bar{x}, \lambda)\} \quad \text{for } (v, w) \in \mathbb{R}^n \times \mathbb{R}^m, \quad (3.10)$$

which can be viewed as the canonical perturbation  $(v, w)$  of the unparameterized *KKT mapping*  $G: \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$  given by

$$G(x, \lambda) := \begin{bmatrix} \nabla_x L(x, \lambda) \\ -f(x) \end{bmatrix} + \begin{bmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{bmatrix}. \quad (3.11)$$

It is easy to see that  $M_{\bar{x}}(0, 0) = \Lambda(\bar{x})$ , where  $\Lambda(\bar{x})$  the set of Lagrange multipliers at  $\bar{x}$  taken from (3.7). We collect some important properties of the multiplier mapping in the next theorem.

**Theorem 3.5 (properties of multipliers and qualification conditions).** *Let  $(\bar{x}, \bar{\lambda})$  be a solution to the KKT system (3.6). The following properties 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  $M_{\bar{x}}$  enjoys the isolated calmness property at  $((0, 0), \bar{\lambda})$ .*
- (iii)** *We have the dual qualification condition*

$$DN_\Theta(f(\bar{x}), \bar{\lambda})(0) \cap \ker \nabla f(\bar{x})^* = \{0\}. \quad (3.12)$$

*If in addition the convex set  $\Theta$  from (3.5) satisfies the standing assumptions (H1) and (H2), then the above conditions are equivalent to:*

- (iv)** *The strong Robinson constraint qualification*

$$\nabla f(\bar{x})\mathbb{R}^n + [T_\Theta(f(\bar{x})) \cap \{\bar{\lambda}\}^\perp] = \mathbb{R}^m. \quad (3.13)$$

**Proof.** The equivalence between (i), (ii), and (iii) follows from [25, Theorem 3.1]. Furthermore, it is shown in [25, Proposition 4.3] that conditions (iii) and (iv) are equivalent for any  $\mathcal{C}^2$ -cone reducible set. The same arguments together with the representation  $DN_\Theta(f(\bar{x}), \bar{\lambda})(0) = K_\Theta(f(\bar{x}), \bar{\lambda})^*$  obtained in Corollary 3.3 under the weaker assumptions (H1) and (H2) can be utilized to verify the claimed equivalence between (iii) and (iv).  $\square$

We end this section by comparing the dual qualification condition (3.12) with the *basic constraint qualification* for (3.5), which is also known as the *Robinson constraint qualification* for problems of conic programming.

**Proposition 3.6 (comparing constraint qualifications).** *Let  $(\bar{x}, \bar{\lambda})$  solve the KKT system (3.6), and let  $\Theta$  in (3.5) satisfy the standing assumptions (H1) and (H2). If the dual qualification condition (3.12) holds, then the following constraint qualifications are also fulfilled:*

(i) *We have the basic constraint qualification*

$$N_\Theta(f(\bar{x})) \cap \ker \nabla f(\bar{x})^* = \{0\}. \quad (3.14)$$

(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 Theorem 3.2 under the standing assumptions (H1) and (H2) that  $\delta_C$  is twice epi-differentiable at  $f(\bar{x})$  for  $\bar{\lambda}$ . Employing this and (3.4) 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

$$N_\Theta(f(\bar{x})) \subset \text{cl}[N_\Theta(f(\bar{x})) + \mathbb{R}\bar{\lambda}] = K_\Theta(f(\bar{x}), \bar{\lambda})^* = N_{K_\Theta(f(\bar{x}), \bar{\lambda})}(0),$$

the dual qualification condition (3.12) ensures that the basic constraint qualification (3.14) holds, which therefore verifies assertion (i).

Turning to (ii), pick  $w \in \mathbb{R}^n$  with  $\nabla f(\bar{x})w \in K_\Theta(f(\bar{x}), \bar{\lambda})$  and get from Corollary 3.3 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 (3.12) justifies (ii) and thus completes the proof of the proposition.  $\square$

It is important to mention that, by the coderivative criterion (2.6), the basic constraint qualification condition (3.14) amounts to saying that the mapping  $x \mapsto f(x) - \Theta$  is *metrically regular* around  $(\bar{x}, 0)$ . This assumption, which has been conventionally involved in the study of constrained optimization problems of type (3.5), is significantly more restrictive than the *metric subregularity* of the mapping  $x \mapsto f(x) - \Theta$  at  $(\bar{x}, 0)$  imposed in Theorem 3.4 as well as in other results of [19] and of this paper.

Our subsequent analysis and applications in this paper are conducted under the standing assumptions (H1) and (H2). It follows from Theorem 3.5 that in this setting the primal and dual conditions (3.12) and (3.13), respectively, are equivalent. Having it in mind, we refer to both of these conditions as the *strong Robinson constraint qualification*.

## 4 Robust Stability of Perturbed KKT Systems

This section is devoted to deriving second-order characterizations of *robust stability* of KKT systems by which we precisely mean the equivalent robust versions of strong metric regularity at the reference point of the KKT mapping  $G(x, \lambda)$  from (3.11) and of isolated calmness of its inverse mapping  $S: \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$  defined by

$$S(v, w) := \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \mid (v, w) \in G(x, \lambda)\} \quad \text{with } (v, w) \in \mathbb{R}^n \times \mathbb{R}^m, \quad (4.1)$$

The latter mapping can be seen as the *solution map* to the *canonical perturbation* of the original KKT system (3.6), i.e., as the KKT system associated with the canonically perturbed constrained optimization problem

$$\min_{x \in \mathbb{R}^n} \varphi_0(x) - \langle v, x \rangle \quad \text{subject to } f(x) + w \in \Theta. \quad (4.2)$$

It follows from [7] that for NLPs the isolated calmness property of (4.1) is characterized by the simultaneous validity of the classical second-order sufficient optimality condition and the uniqueness of Lagrange multipliers. As further shown in [6], a similar result holds for the case of conic programs in (3.5) with a  $\mathcal{C}^2$ -cone reducible set  $\Theta$ , provided that the uniqueness of Lagrange multipliers is replaced by the strong Robinson constraint qualification (3.13); see Remark 4.3 for more discussions. In this section we prove that the result of [7] can be extended to a general class of constrained optimization problems of type (3.5) with *parabolically regular* sets  $\Theta$  under the additional assumption that the *multiplier mapping* (3.10) is *calm*. Moreover, our results reveal that the combination of the uniqueness of Lagrange multipliers and the calmness of the multiplier mapping amounts to the strong Robinson constraint qualification. It shows that the calmness of the multiplier mapping, being automatically satisfied for NLPs, cannot be dismissed in general. The usage of this property for the study of the robust stability of KKT systems and for the subsequent SQP applications is new not only for parabolically regular problems but also for less general classes of nonpolyhedral programs, while this condition has been recently employed in [25] to characterize noncriticality of Lagrange multipliers.

We refer the reader to [25] for more discussions on and sufficient conditions for the *calmness* of the multiplier mapping (3.10). Besides the NLP setting that corresponds to the *polyhedrality* of the set  $\Theta$  in (3.5), this calmness holds when the  $\Theta$  is either second-order/Lorentz cone or the cone of semidefinite symmetric matrices under the validity of the *strict complementarity* condition, i.e., if there exists  $\lambda \in \Lambda(\bar{x})$  with  $\lambda \in \text{ri } N_\Theta(f(\bar{x}))$ ; see [25, Theorem 5.10].

To achieve our goals, we first establish relationships between the isolated calmness of the solution map  $S$  from (4.1) and the concept of noncriticality of Lagrange multipliers of the KKT systems under consideration. Recall the notions of critical and noncritical multipliers taken from [24, Definition 3.1]. Let  $(\bar{x}, \bar{\lambda})$  be a solution to the KKT system (3.6). Then the multiplier  $\bar{\lambda} \in \Lambda(\bar{x})$  is said to be *critical* for (3.6) if there is a nonzero vector  $w \in \mathbb{R}^n$  satisfying the inclusion

$$0 \in \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w + \nabla f(\bar{x})^* D N_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w). \quad (4.3)$$

The multiplier  $\bar{\lambda} \in \Lambda(\bar{x})$  is *noncritical* for (3.6) if (4.3) admits only the trivial solution  $w = 0$ . The above definitions are far-going extensions of those introduced by Izmailov in [14] for nonlinear programs with equality constraints, where  $\Theta = \{0\} \subset \mathbb{R}^m$ , and thus the appeal to generalized differentiation is not required. Characterizations of noncriticality are given in [15, Proposition 1.43] for NLPs and in [25, Theorem 4.1] for  $\mathcal{C}^2$ -cone reducible constrained problems. Now we come up to the following relationships between noncriticality and isolated calmness.

**Proposition 4.1 (characterizations of isolated calmness via noncriticality).** *Let  $(\bar{x}, \bar{\lambda})$  be a solution to the KKT system (3.6). Consider the following statements:*

- (i) *The Lagrange multiplier  $\bar{\lambda}$  is noncritical for (3.6).*
- (ii) *The solution map  $S$  from (4.1) enjoys the isolated calmness property at  $((0, 0), (\bar{x}, \bar{\lambda}))$ .*

*Then we have the following relationships:*

- (a) *Implication (ii)  $\implies$  (i) is always fulfilled.*
- (b) *The converse implication (i)  $\implies$  (ii) also holds provided that  $\Lambda(\bar{x}) = \{\bar{\lambda}\}$  and that the multiplier mapping  $M_{\bar{x}}$  from (3.10) is calm at  $((0, 0), \bar{\lambda})$ .*

**Proof.** It is not hard to deduce from the definition of noncriticality that the Lagrange multiplier  $\bar{\lambda}$  is noncritical if and only if we have

$$\begin{cases} \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w + \nabla f(\bar{x})^*w = 0, \\ u \in DN_{\Theta}(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w) \end{cases} \implies w = 0. \quad (4.4)$$

On the other hand, it follows from the Levy-Rockafellar criterion (see, e.g., [8, Theorem 4G.1]) that  $S$  enjoys the isolated calmness property at  $((0, 0), (\bar{x}, \bar{\lambda}))$  if and only if  $DS((0, 0), (\bar{x}, \bar{\lambda}))(0, 0) = \{(0, 0)\}$ , which amounts to saying that

$$\begin{cases} \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w + \nabla f(\bar{x})^*w = 0, \\ u \in DN_{\Theta}(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w) \end{cases} \implies w = 0, u = 0. \quad (4.5)$$

It is clear therefore that assertion (a) is satisfied. Suppose further that the multiplier  $\bar{\lambda}$  is noncritical, and so the implication in (4.4) holds. Then Theorem 3.5 ensures the validity of the dual qualification condition (3.12) under the assumptions made in (b). Picking a pair  $(w, u)$  satisfying the left-hand side relations in (4.5), we obtain from (4.4) that  $w = 0$ . This yields

$$u \in DN_{\Theta}(f(\bar{x}), \bar{\lambda})(0) \cap \ker \nabla f(\bar{x})^*.$$

Appealing now to (3.12) confirms that  $u = 0$ , which tells us in turn that (4.5) holds, and hence we complete the proof of assertion (b) and of the whole proposition.  $\square$

It is important to notice that the imposed calmness of the multiplier mapping (3.10) is essential for the validity of (i)  $\implies$  (ii) in Proposition 4.1. Indeed, [25, Example 4.8] provides a simple semidefinite program for which the calmness of the multiplier mapping  $M_{\bar{x}}$  fails and the unique Lagrange multiplier therein is noncritical, but the solution map  $S$  does not satisfy the isolated calmness property. On the other hand, the polyhedrality of the convex set  $\Theta$ , along with the noncriticality of the unique Lagrange multiplier, ensures the isolated calmness of  $S$ . This follows from Proposition 4.1 due to the automatic fulfillment of the calmness property of  $M_{\bar{x}}$  under the imposed polyhedrality as a direct consequence of the classical Hoffman lemma.

Now we are ready to derive a major result of this section giving us complete second-order characterizations of the isolated calmness property of the KKT solution map (4.1) under parabolic regularity of the underlying set  $\Theta$  in (3.5).

**Theorem 4.2 (second-order characterizations of isolated calmness for KKT systems).** *Let  $(\bar{x}, \bar{\lambda})$  be a solution to the KKT system (3.6) under the standing assumptions (H1) and (H2). Then the following assertions are equivalent:*

(i) *The solution map  $S$  from (4.1) enjoys the isolated calmness property at  $((0, 0), (\bar{x}, \bar{\lambda}))$ , and  $\bar{x}$  is a local minimizer of (3.5).*

(ii) *The second-order sufficient optimality condition*

$$\begin{cases} \langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w, w \rangle + d^2 \delta_{\Theta}(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w) > 0 \\ \text{for all } w \in \mathbb{R}^n \setminus \{0\} \text{ with } \nabla f(\bar{x})w \in K_{\Theta}(f(\bar{x}), \bar{\lambda}) \end{cases} \quad (4.6)$$

*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 optimality condition (4.6) and the strong Robinson constraint qualification (3.12) are satisfied simultaneously.*

(iv) *The Lagrange multiplier  $\bar{\lambda}$  from (3.7) is noncritical for (3.6),  $\bar{x}$  is a local minimizer of (3.5), the multiplier mapping  $M_{\bar{x}}$  is calm at  $((0, 0), \bar{\lambda})$ , and  $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ .*

**Proof.** We begin with verifying implication (i)  $\implies$  (iv). Since  $S$  enjoys the isolated calmness property at  $((0, 0), (\bar{x}, \bar{\lambda}))$  and  $\Lambda(\bar{x})$  is convex, we get  $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ . The calmness of the Lagrange multiplier mapping  $M_{\bar{x}}$  at  $((0, 0), \bar{\lambda})$  is a consequence of the isolated calmness of  $S$  at  $((0, 0), (\bar{x}, \bar{\lambda}))$ . Finally, Proposition 4.1(a) tells us that  $\bar{\lambda}$  is noncritical, and so we arrive at (iv).

Next we prove implication (iv)  $\implies$  (ii). Assume that (iv) holds and deduce from Theorem 3.5 that (3.12) is satisfied. This along with Proposition 3.6(i) implies that the basic constraint qualification (3.14) is also satisfied. As mentioned before, the qualification condition (3.14) guarantees the metric subregularity of the mapping  $x \mapsto f(x) - \Theta$  at  $(\bar{x}, 0)$  required in Theorem 3.4. Employing Theorem 3.4(i) with  $\Lambda(\bar{x}) = \{\bar{\lambda}\}$  tells us that the second-order necessary optimality condition

$$\langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w, w \rangle + d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w) \geq 0 \quad \text{for all } w \in K_\Omega(\bar{x}, -\nabla \varphi_0(\bar{x})) \quad (4.7)$$

fulfills, where  $\Omega$  is taken from (3.9). Furthermore, it follows from (3.14) that

$$w \in K_\Omega(\bar{x}, -\nabla \varphi_0(\bar{x})) \iff \nabla f(\bar{x})w \in K_\Theta(f(\bar{x}), \bar{\lambda}).$$

To verify the second-order sufficient condition (4.6), we need to show that the above inequality is strict for any  $w \neq 0$ . Arguing by contradiction, suppose that there exists a nonzero vector  $\bar{w} \in \mathbb{R}^n$  satisfying the equality

$$\langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})\bar{w}, \bar{w} \rangle + d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})\bar{w}) = 0 \quad \text{with } \nabla f(\bar{x})\bar{w} \in K_\Theta(f(\bar{x}), \bar{\lambda}).$$

Consider now the auxiliary optimization problem

$$\min_{w \in \mathbb{R}^n} \frac{1}{2} \left( \langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w, w \rangle + d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w) \right) \quad (4.8)$$

and deduce from (4.7) that  $\bar{w}$  is an optimal solution to (4.8). It follows from Theorem 3.2 that  $\text{dom } d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda}) = K_\Theta(f(\bar{x}), \bar{\lambda})$ . Combining this with Proposition 3.6(ii) confirms that

$$N_{\text{dom } d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda})}(\nabla f(\bar{x})w) \cap \ker \nabla f(\bar{x})^* = \{0\}.$$

Since the second subderivative  $d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda})$  is a convex function, the latter condition allows us to obtain the subdifferential chain rule

$$\partial_w \left( \frac{1}{2} d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x}) \cdot) \right) (\bar{w}) = \nabla f(\bar{x})^* \partial \left[ \frac{1}{2} d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda}) \right] (\nabla f(\bar{x})\bar{w}).$$

Remembering that  $\bar{w}$  is a minimizer for (4.8) and employing the subdifferential Fermat rule together the chain rule above and formula (3.3) yield

$$\begin{aligned} 0 &\in \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})\bar{w} + \nabla f(\bar{x})^* \partial \left[ \frac{1}{2} d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda}) \right] (\nabla f(\bar{x})\bar{w}) \\ &= \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})\bar{w} + \nabla f(\bar{x})^* D N_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})\bar{w}). \end{aligned}$$

Thus we arrive at the inclusion of type (4.3). Since  $\bar{w} \neq 0$ , it tells us that  $\bar{\lambda}$  is a critical multiplier, a contradiction. This shows that the inequality in (4.7) is strict, which proves the validity of the second-order sufficient condition (4.6) and hence verifies (ii).

The next step is to check the fulfillment of (ii)  $\implies$  (i). Assuming (ii), we are going to prove that  $\bar{\lambda}$  is noncritical, which amounts to show that implication (4.4) holds. To proceed, pick any pair  $(w, u)$  satisfying the left-hand side relations in (4.4). This brings us to

$$\langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w, w \rangle + \langle u, \nabla f(\bar{x})w \rangle = 0 \quad \text{and} \quad u \in D N_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w). \quad (4.9)$$

The second condition in (4.9) together with (3.3) ensures that  $u \in \partial[\frac{1}{2}d^2\delta_\Theta(f(\bar{x}), \bar{\lambda})](\nabla f(\bar{x})w)$ . Thus we obtain the relationships

$$\nabla f(\bar{x})w \in \text{dom } d^2\delta_\Theta(f(\bar{x}), \bar{\lambda}) = K_\Theta(f(\bar{x}), \bar{\lambda}).$$

Recalling again that  $d^2\delta_\Theta(f(\bar{x}), \bar{\lambda})$  is a convex function, it follows from the definition of the subdifferential in convex analysis that

$$\langle u, v - \nabla f(\bar{x})w \rangle \leq \frac{1}{2}d^2\delta_\Theta(f(\bar{x}), \bar{\lambda})(v) - \frac{1}{2}d^2\delta_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w)$$

whenever  $v \in \mathbb{R}^m$ . Pick further any  $\varepsilon \in (0, 1)$  and set  $v := (1 \pm \varepsilon)\nabla f(\bar{x})w$ . Since the second subderivative is positive homogeneous of degree 2, we get

$$\pm \langle u, \nabla f(\bar{x})w \rangle \leq \frac{\varepsilon \pm 2}{2}d^2\delta_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w).$$

Letting  $\varepsilon \downarrow 0$  therein clearly gives us the equality  $d^2\delta_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w) = \langle u, \nabla f(\bar{x})w \rangle$ . Combining the latter with (4.9) implies that

$$\langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w, w \rangle + d^2\delta_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w) = 0, \quad \nabla f(\bar{x})w \in K_\Theta(f(\bar{x}), \bar{\lambda}),$$

which results in  $w = 0$  due to the second-order condition (3.4). This verifies that the multiplier  $\bar{\lambda}$  is noncritical. Appealing now to Proposition 4.1(b) tells us that the solution map  $S$  enjoys the isolated calmness property at  $((0, 0), (\bar{x}, \bar{\lambda}))$ . We also conclude from (4.6),  $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ , and Theorem 3.4(ii) that  $\bar{x}$  is a local minimizer of (3.5), which proves that (i) holds.

To complete the proof of theorem, it remains to observe that the equivalence between (ii) and (iii) is a direct consequence of Theorem 3.5.  $\square$

Let us now discuss relationships of the obtained characterizations of isolated calmness for KKT systems with known results on the subject.

**Remark 4.3 (comparison with known results on the KKT isolated calmness).** The equivalence between assertion (i) and (iii) of Theorem 4.2 was obtained in [6, Theorem 24] for the case where  $\mathcal{C}^2$ -cone reducible sets  $\Theta$  by using a different approach. Our new proof for the general case is deeply rooted into the recent developments on parabolic regularity and its important consequences in second-order variational analysis developed in [19]. Characterizations (ii) and (iv) of Theorem 4.2 did not appear either in [6, Theorem 24] or any other publication on nonpolyhedral problems. These characterizations highlight significant differences in dealing with KKT systems of type (3.6) where  $\Theta$  is a nonpolyhedral set. If  $\Theta$  is polyhedron in (3.6), the optimization problem (3.5) can be reduced to a nonlinear program. It is known from [7, Theorem 2.6] that in the NLP case the uniqueness of Lagrange multipliers together with (4.6) amounts to the isolated calmness of the solution mapping  $S$  from (4.1). As argued in [6], a similar result is not expected for nonpolyhedral constrained optimization while it holds if the uniqueness of multipliers is replaced by the strong Robinson constraint qualification (3.13); see [6, Theorem 24]. However, the authors of [6] did not address the question to understand why the result with the multiplier uniqueness may fail for nonpolyhedral problems. Theorem 4.2(ii) answers this question by revealing that the calmness of the multiplier mapping  $M_{\bar{x}}$  (which is automatic for NLPs) is essential to derive such an equivalence. As shown in [25, Example 3.4], the calmness of  $M_{\bar{x}}$  may fail for a two-dimensional semidefinite program (and hence for a three-dimensional ice-cream program) even in the case where the set of Lagrange multipliers is a singleton. In that example, the equivalences of Theorem 4.2 also fail.

Finally in this remark, observe that the assumptions of [6, Theorem 24] include the basic constraint qualification (3.14). However, our proof shows that the latter condition can be dropped without any harm even in the more general case under consideration in Theorem 4.2.

Now we consider a parameterized version of problem (3.5) together with the corresponding parameterized KKT system and establish crucial *well-posedness* and *robust stability* results for them with respect to *general perturbations*. Given a parameter space  $\mathbb{R}^d$  and a nominal parameter  $\bar{p} \in \mathbb{R}^d$ , define the mappings  $\tilde{\varphi}_0 : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}$  and  $\tilde{f} : \mathbb{R}^n \times \mathbb{R}^d \rightarrow \mathbb{R}^m$  such that

$$\tilde{\varphi}_0(x, \bar{p}) = \varphi_0(x) \quad \text{and} \quad \tilde{f}(x, \bar{p}) = f(x) \quad \text{for all } x \in \mathbb{R}^n, \quad (4.10)$$

where  $\varphi_0$  and  $f$  are taken from (3.5). Assume that both  $\tilde{\varphi}_0$  and  $\tilde{f}$  are  $\mathcal{C}^2$ -smooth with respect to both variables  $x$  and  $p$ . Then we associate with (3.5) its perturbed version

$$\min_{x \in \mathbb{R}^n} \tilde{\varphi}_0(x, p) \quad \text{subject to} \quad \tilde{f}(x, p) \in \Theta. \quad (4.11)$$

The next theorem is a nonpolyhedral extension of [15, Theorem 1.21] obtained there for NLPs.

**Theorem 4.4 (well-posedness and robust stability of solution maps under general perturbations).** *Let  $(\bar{x}, \bar{\lambda})$  be a solution to the KKT system (3.6) under the standing assumptions (H1) and (H2). Suppose that in addition to (4.10) we have the equalities*

$$\nabla_x \tilde{\varphi}_0(\bar{x}, \bar{p}) = \nabla \varphi_0(\bar{x}) \quad \text{and} \quad \nabla_x \tilde{f}(\bar{x}, \bar{p}) = \nabla f(\bar{x}). \quad (4.12)$$

*Then the following assertions hold:*

- (i) *If  $\bar{x}$  is a strict local minimizer of (4.11) with  $p = \bar{p}$  and if the basic constraint qualification (3.14) is satisfied, then for any  $p \in \mathbb{R}^d$  sufficiently close to  $\bar{p}$  problem (4.11) admits a local minimizer  $x_p$  converging to  $\bar{x}$  as  $p \rightarrow \bar{p}$ .*
- (ii) *If the second-order sufficient condition (4.6) and the strong Robinson constraint qualification (3.12) are satisfied, and if*

$$\nabla_{xx}^2 \tilde{\varphi}_0(\bar{x}, \bar{p}) + \nabla_{xx}^2 \langle \bar{\lambda}, \tilde{f} \rangle(\bar{x}, \bar{p}) = \nabla_{xx}^2 L(\bar{x}, \bar{\lambda}), \quad (4.13)$$

*then there exist constants  $\varepsilon > 0$  and  $\ell \geq 0$  such that*

$$\emptyset \neq \Upsilon(p) \cap \mathbb{B}_\varepsilon(\bar{x}, \bar{\lambda}) \subset \{(\bar{x}, \bar{\lambda})\} + \ell \|p - \bar{p}\| \quad \text{for all } p \in \mathbb{B}_\varepsilon(\bar{p}), \quad (4.14)$$

*where  $\Upsilon : \mathbb{R}^d \rightrightarrows \mathbb{R}^n \times \mathbb{R}^m$  is the solution map to the KKT system of (4.11) defined by*

$$\Upsilon(p) := \left\{ (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \mid \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla_x \tilde{\varphi}_0(x, p) + \nabla_x \tilde{f}(x, p)^* \lambda \\ -\tilde{f}(x, p) \end{bmatrix} + \begin{bmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{bmatrix} \right\}.$$

**Proof.** Since  $\bar{x}$  is a strict local minimum of (4.11) as  $p = \bar{p}$ , we get a number  $\varepsilon > 0$  such that  $\bar{x}$  is the unique minimizer for the problem

$$\min_{x \in \mathbb{R}^n} \tilde{\varphi}_0(x, \bar{p}) \quad \text{subject to} \quad \tilde{f}(x, \bar{p}) \in \Theta, \quad x \in \mathbb{B}_\varepsilon(\bar{x}). \quad (4.15)$$

Consider the auxiliary constrained optimization problem

$$\min_{x \in \mathbb{R}^n} \tilde{\varphi}_0(x, p) \quad \text{subject to} \quad \tilde{f}(x, p) \in \Theta, \quad x \in \mathbb{B}_\varepsilon(\bar{x}). \quad (4.16)$$

It follows from (3.14), (4.10), and (4.12) that

$$N_\Theta \tilde{f}(\bar{x}, \bar{p}) \cap \ker \nabla_x \tilde{f}(\bar{x}, \bar{p})^* = \{0\}, \quad (4.17)$$

which is equivalent to the Robinson constraint qualification for the constraint system  $\tilde{f}(x, p) \in \Theta$  at  $(\bar{x}, \bar{p})$ , i.e., to the condition

$$0 \in \text{int}(\tilde{f}(\bar{x}, \bar{p}) + \nabla_x \tilde{f}(\bar{x}, \bar{p}) \mathbb{R}^n - \Theta).$$

Thus it follows from [3, Theorem 2.86] that for all  $(x, p)$  sufficiently close to  $(\bar{x}, \bar{p})$  we have

$$\text{dist}(x; \Gamma(p)) = O(\text{dist}(\tilde{f}(x; p); \Theta)), \quad (4.18)$$

where  $\Gamma(p) := \{x \in \mathbb{R}^n \mid \tilde{f}(x, p) \in \Theta\}$ . This guarantees that whenever  $p$  is sufficiently close to  $\bar{p}$  the feasible region of (4.16), namely  $\Gamma(p) \cap \mathbb{B}_\varepsilon(\bar{x})$ , is nonempty for all small positive numbers  $\varepsilon$ . Appealing to the classical Weierstrass theorem, we deduce that (4.16) attains a (global) minimizer  $\bar{x}_p$  for all  $p$  sufficiently close to  $\bar{p}$ .

Now we claim that  $\bar{x}_p \rightarrow \bar{x}$  as  $p \rightarrow \bar{p}$ . Indeed, the failure of it gives us a sequence  $p_k \rightarrow \bar{p}$  for which  $\bar{x}_{p_k}$ , an optimal solution of (4.16) for  $p = p_k$ , does not converge to  $\bar{x}$  as  $k \rightarrow \infty$ . Since  $\bar{x}_{p_k} \in \mathbb{B}_\varepsilon(\bar{x})$ , there exists a subsequence of  $\bar{x}_{p_k}$  that converges to some  $\tilde{x} \in \mathbb{B}_\varepsilon(\bar{x})$  with  $\tilde{x} \neq \bar{x}$ . Without relabeling, assume with no loss of generality that  $\bar{x}_{p_k} \rightarrow \tilde{x}$  for all  $k \rightarrow \infty$  and then find by (4.18) such  $\tilde{x}_{p_k} \in \Gamma(p_k)$  that

$$\|\tilde{x}_{p_k} - \bar{x}\| = O(\text{dist}(\tilde{f}(\bar{x}, p_k); \Theta)) = O(\|p_k - \bar{p}\|).$$

Remembering that  $p_k \rightarrow \bar{p}$  as  $k \rightarrow \infty$ , it is no harm to suppose that  $\tilde{x}_{p_k} \in \mathbb{B}_\varepsilon(\bar{x})$  for all  $k \in \mathbb{N}$ . Since  $\tilde{x}_{p_k}$  is an optimal solution of (4.16) for  $p = p_k$ , we have  $\tilde{\varphi}_0(\bar{x}_{p_k}, p_k) \leq \tilde{\varphi}_0(\tilde{x}_{p_k}, p_k)$ . Passing to the limit as  $k \rightarrow \infty$  brings us to  $\tilde{\varphi}_0(\tilde{x}, \bar{p}) \leq \tilde{\varphi}_0(\bar{x}, \bar{p})$ , which contradicts the fact that  $\bar{x}$  is the unique minimizer of problem (4.15). Thus we are done with (i).

Turing next to (ii), pick any  $\varepsilon > 0$  such that (i) holds whenever  $p \in \mathbb{B}_\varepsilon(\bar{p})$ . We claim that the set of Lagrange multipliers for (4.11) at  $(\bar{x}, \bar{p})$  is the singleton  $\{\bar{\lambda}\}$ , which follows from the combination of Theorem 3.5 with conditions (3.12), (4.10), and (4.12). It allows us to deduce from (4.13), the second-order sufficient condition (4.6), and Theorem 3.4(ii) that  $\bar{x}$  is a strict local minimizer of (4.11) for  $p = \bar{p}$ . Employing now assertion (i) of this theorem, we conclude that for any  $p$  sufficiently close to  $\bar{p}$  problem (4.11) admits a local minimizer  $x_p$  with  $x_p \rightarrow \bar{x}$  as  $p \rightarrow \bar{p}$ . It follows from Proposition 3.6(i) combined with conditions (3.12) and (4.12) that the basic constraint qualification (4.17) for problem (4.15) is satisfied. Shrinking the closed ball  $\mathbb{B}_\varepsilon(\bar{p})$  if necessary, we assume without loss of generality that the basic constraint qualification (4.17) holds for the perturbed problem (4.16) at  $(x_p, p)$  for all  $p \in \mathbb{B}_\varepsilon(\bar{p})$ . This ensures the existence of a Lagrange multiplier  $\lambda_p$  associated with the local minimizer  $x_p$  for (4.11) whenever  $p \in \mathbb{B}_\varepsilon(\bar{p})$ . Shrinking again the ball  $\mathbb{B}_\varepsilon(\bar{p})$ , we conclude that the Lagrange multipliers  $\lambda_p$  are uniformly bounded for all  $p \in \mathbb{B}_\varepsilon(\bar{p})$ . Remembering that the set of Lagrange multipliers for (4.11) at  $(\bar{x}, \bar{p})$  is the singleton  $\{\bar{\lambda}\}$  and using it together with the boundedness of  $\{\lambda_p\}$  for  $p \in \mathbb{B}_\varepsilon(\bar{p})$  yield the convergence  $\lambda_p \rightarrow \bar{\lambda}$  as  $p \rightarrow \bar{p}$ . Since  $(x_p, \lambda_p) \in \Upsilon(p)$  and  $(x_p, \lambda_p) \rightarrow (\bar{x}, \bar{\lambda})$  as  $p \rightarrow \bar{p}$ , by decreasing  $\varepsilon > 0$  if necessary we get  $\Upsilon(p) \cap \mathbb{B}_\varepsilon(\bar{x}, \bar{\lambda}) \neq \emptyset$  for all  $p \in \mathbb{B}_\varepsilon(\bar{p})$ .

To justify finally the inclusion in (4.14), observe first by (4.10) and (4.12) that  $(\bar{x}, \bar{\lambda}) \in \Upsilon(\bar{p})$ . Employing the latter together with (4.13) and [8, Corollary 4E.3] tells us that the mapping  $\Upsilon$  enjoys the isolated calmness property at  $(\bar{p}, (\bar{x}, \bar{\lambda}))$  provided that the implication

$$\begin{cases} \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w + \nabla f(\bar{x})^*u = 0, \\ u \in DN_\Theta(f(\bar{x}), \bar{\lambda})(\nabla f(\bar{x})w) \end{cases} \implies w = 0, u = 0 \quad (4.19)$$

holds. To verify (4.19), we proceed similarly to the proof of implication (ii)  $\implies$  (i) in Theorem 4.2 and thus complete the proof of the theorem.  $\square$

The following remark is important to understand the meaning of the robust stability property established in Theorem 4.4(ii) and its comparison with known results in this direction.

**Remark 4.5 (robust stability of KKT systems with respect to general vs. canonical perturbations).** The properties established in (4.14) of Theorem 4.4(ii) were investigated before for nonlinear programming problems in [7, Theorem 2.6]. Then they were extended in [6, Theorem 24] to  $\mathcal{C}^2$ -cone reducible constrained optimization problems in the case where the parameterized problem (4.11) has the *canonically perturbed* form (4.2) meaning that

$$\tilde{\varphi}_0(x, p) = \varphi_0(x) - \langle v, x \rangle \text{ and } \tilde{f}(x, p) = f(x) + w \text{ with } p := (v, w) \in \mathbb{R}^n \times \mathbb{R}^m.$$

The properties in (4.14) of Theorem 4.4(ii) were called in [6] the *robust isolated calmness*. We can see that the difference between (4.14) and the isolated calmness of the solution map  $\Upsilon$  is the additional highly important assertion that  $\Upsilon(p) \cap \mathbb{B}_\varepsilon(\bar{x}, \bar{\lambda}) \neq \emptyset$  for all  $p$  sufficiently close to  $\bar{p}$ . In general, the isolated calmness and its robust counterpart (4.14) are *not equivalent* as shown in [23, Example 6.4] for a three-dimensional second-order cone program.

On the other hand, it follows from the previous Theorem 4.2 that the assumptions imposed in Theorem 4.4(ii) yields the isolated calmness property at  $((0, 0), (\bar{x}, \bar{\lambda}))$  of the solution map (4.1) for the KKT system associated with the canonically perturbed form (4.2). Combining it with Theorem 4.4(ii) tells us that for *canonical perturbations* the isolated calmness and robust isolated properties of KKT solution maps are *equivalent*. Our main reason to consider *general perturbations* as in (4.11) and to justify the robust isolated calmness for the associated KKT systems is that this allows us to verify the *solvability* of subproblems in the basic SQP method for parabolically regular constrained optimization problems as developed in the next section. It does not seem to be possible while considering only canonical perturbations.

Note to this end that the assumptions (4.10), (4.12), and (4.13) imposed in Theorem 4.4 are automatically satisfied in our application to the aforementioned solvability of the basic SQP method for the original problem (3.5) that is addressed below in Theorem 5.2.

## 5 Superlinear Convergence of the Basic SQP Method

This section is devoted to establish the *primal-dual superlinear convergence* of the basic SQP method for the general class of parabolically regular constrained optimization problems of type (3.5). Various versions of the SQP method have been among the most effective numerical algorithms in constrained optimization. The principal idea of SQP algorithms is to solve a sequence of quadratic approximations, called subproblems, whose optimal solutions converge under appropriate assumptions to an optimal solution of the original problem.

Given the current iterate  $x_k$ , the *generic SQP subproblems* for (3.5) are formulated as

$$\begin{cases} \min_{x \in \mathbb{R}^n} & \varphi_0(x_k) + \langle \nabla \varphi_0(x_k), x - x_k \rangle + \frac{1}{2} \langle H_k(x - x_k), x - x_k \rangle \\ \text{subject to} & f(x_k) + \nabla f(x_k)(x - x_k) \in \Theta, \end{cases} \quad (5.1)$$

where  $H_k \in \mathbb{R}^{n \times n}$  for all  $k = 0, 1, \dots$ . The KKT system associated with each subproblem (5.1) is described by the generalized equation

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla_x L(x_k, \lambda_k) \\ -f(x_k) \end{bmatrix} + \begin{bmatrix} H_k & \nabla f(x_k)^* \\ -\nabla f(x_k) & 0 \end{bmatrix} \begin{bmatrix} x - x_k \\ \lambda - \lambda_k \end{bmatrix} + \begin{bmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{bmatrix}. \quad (5.2)$$

The *generic SQP method* for the original problem (3.5) is therefore as follows:

**Algorithm 5.1 (generic SQP method).** Choose  $(x_k, \lambda_k) \in \mathbb{R}^n \times \mathbb{R}^m$  and set  $k = 0$ .

1. If  $(x_k, \lambda_k)$  satisfies the KKT system (3.6), then stop.

2. Choose some matrix  $H_k \in \mathbb{R}^{n \times n}$  and compute the next iterate  $(x_{k+1}, \lambda_{k+1})$  as a solution to the generalized equation (5.2).
3. Increase  $k$  by 1 and then go back to Step 1.

In *quasi-Newton SQP* algorithms, the matrix  $H_k$  is chosen as a quasi-Newton approximation of the Hessian  $\nabla_{xx}^2 L(x_k, \lambda_k)$ . An efficient way to construct  $H_k$  is the *BFGS* (Broyden–Fletcher–Goldfarb–Shanno) method; see [9, 15, 18] for more details on this and related algorithms. Instructive discussion and suggestions on how to construct quasi-Newton SQP approximations can be found in [2]. Following [15], we refer to the *basic SQP method* if  $H_k$  are chosen as

$$H_k = \nabla_{xx}^2 L(x_k, \lambda_k) \quad \text{for all } k = 0, 1, \dots \quad (5.3)$$

It has been well recognized, 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 the KKT system (3.6) as the generalized equation

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla_x L(x, \lambda) \\ -f(x) \end{bmatrix} + \begin{bmatrix} 0 \\ N_{\Theta}^{-1}(\lambda) \end{bmatrix}. \quad (5.4)$$

Employing the Newton method for this generalized equation leads us to the aforementioned basic SQP method for the constrained problem (3.5); see [15, Section 3.1] for more details.

The Newton method for generalized equations has been investigated extensively since the late 1970s; see, e.g., [8, 9, 15, 18] and the references therein. Josephy's observation in [16] significantly advanced the topic by showing that *strong regularity* in the sense of Robinson [27] ensures the superlinear convergence of the Newton method for variational inequalities. As proved in [27, Theorem 4.1] for NLPs, strong regularity is guaranteed under a stronger form of the second-order sufficient condition (4.6) together with the linear independence constraint qualification. The next improvement in the setting of NLPs was achieved by Bonnans in [1] by showing that the basic SQP method converges superlinearly if the second-order sufficient condition (4.6) with a *polyhedral* set  $\Omega$  therein is satisfied and the set of Lagrange multipliers is a *singleton*.

In this section we are going to extend Bonnans' result to nonpolyhedral constrained optimization problems (3.5) in the general case of *parabolically regular* sets  $\Theta$ . To do it, we intend to employ the results from the previous section, which tell us that an extra *calmness* assumption on the multiplier mapping (3.10) is needed. To achieve our goal, we utilize—besides Theorems 4.2 and 4.4 obtained above—the result of Theorem 3.2 from the book by Izmailov and Solodov [15] in which superlinear convergence of the Newton method for abstract generalized equations was established under the following two assumptions. The first one, called “semistability” in [15], reduces for (5.4) to the isolated calmness of the solution map  $S$  from (4.1). The second assumption of [15, Theorem 3.2], called “hemistability” therein, is defined in [15, Definition 3.1] for arbitrary generalized equations of the type

$$0 \in g(x) + F(x), \quad x \in \mathbb{R}^n, \quad (5.5)$$

with  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and  $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ . A feasible solution  $\bar{x}$  to (5.5) is said to be *hemistable* if for any  $x$  close to  $\bar{x}$  the perturbed generalized equation

$$0 \in g(x) + \nabla g(x)\eta + F(x + \eta)$$

has a solution  $\eta_x$  such that  $\eta_x \rightarrow 0$  as  $x \rightarrow \bar{x}$ . We show below that the isolated calmness of the solution map (4.1) at  $((0, 0), (\bar{x}, \bar{\lambda}))$  yields the hemistability of the solution pair  $(\bar{x}, \bar{\lambda})$  for (5.4).

The first major theorem of this section justifies the crucial *well-posedness* of subproblems (5.1) in the basic SQP case (5.3) under parabolic regularity, in the sense of the existence of optimal

solutions to (5.1) (i.e., *solvability* of these subproblems) and *stability* of the corresponding KKT systems with respect to general parameter perturbations. The proofs of these results are based on the robust stability characterizations established in Section 4, which are in turn employ recent developments on parabolic regularity [19] and noncriticality of Lagrange multipliers [25].

**Theorem 5.2 (solvability and stability of subproblems in the basic SQP method).** *Let  $(\bar{x}, \bar{\lambda})$  be a solution to the KKT system (3.6) under the standing assumptions (H1) and (H2). Suppose in addition that the second-order sufficient condition (4.6) and the strong Robinson constraint qualification (3.12) are satisfied. Then there is  $\varepsilon > 0$  such that for all  $(u, \mu) \in \mathbb{B}_\varepsilon(\bar{x}, \bar{\lambda})$  the following assertions hold:*

(i) *The perturbed optimization problem*

$$\begin{cases} \min_{x \in \mathbb{R}^n} & \varphi_0(u) + \langle \nabla \varphi_0(u), x - u \rangle + \frac{1}{2} \langle \nabla_{xx}^2 L(u, \mu)(x - u), x - u \rangle \\ \text{subject to} & f(u) + \nabla f(u)(x - u) \in \Theta \end{cases} \quad (5.6)$$

*always admits a local optimal solution.*

(ii) *The KKT system of (5.6), which can be formulated as*

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla_x L(u, \mu) \\ -f(u) \end{bmatrix} + \begin{bmatrix} \nabla_{xx}^2 L(u, \mu) & \nabla f(u)^* \\ -\nabla f(u) & 0 \end{bmatrix} \begin{bmatrix} x - u \\ \lambda - \mu \end{bmatrix} + \begin{bmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{bmatrix}, \quad (5.7)$$

*has a solution  $(x, \lambda)$  that converges to  $(\bar{x}, \bar{\lambda})$  as  $(u, \mu) \rightarrow (\bar{x}, \bar{\lambda})$ .*

**Proof.** Set  $p := (u, \mu)$  with  $\bar{p} := (\bar{x}, \bar{\lambda})$ , denote

$$\varphi(x, p) := \varphi_0(u) + \langle \nabla \varphi_0(u), x - u \rangle + \frac{1}{2} \langle \nabla_{xx}^2 L(u, \mu)(x - u), x - u \rangle, \quad g(x, p) := f(u) + \nabla f(u)(x - u),$$

and view (5.6) as the parameterized optimization problem

$$\min_{x \in \mathbb{R}^n} \varphi(x, p) \quad \text{subject to } g(x, p) \in \Theta \quad (5.8)$$

with respect to the *general* parameter perturbations  $p$ . We claim now that  $x := \bar{x}$  is a *strict local minimizer* for (5.8) associated with  $\bar{p}$ . To this end, observe first that the KKT system of (5.8) associated with  $\bar{p}$  has the representation

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \nabla \varphi(\bar{x}) + \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})(x - \bar{x}) + \nabla f(\bar{x})^* \lambda \\ -f(\bar{x}) - \nabla(\bar{x})(x - \bar{x}) \end{bmatrix} + \begin{bmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{bmatrix}. \quad (5.9)$$

We clearly have that  $(\bar{x}, \bar{\lambda})$  is a solution to (5.9), which hence implies that  $\bar{x}$  is a stationary point for (5.8) associated with  $\bar{p}$  and that  $\bar{\lambda}$  is a Lagrange multiplier associated with  $\bar{x}$  for the latter problem. Furthermore, it is not hard to observe from the KKT system (5.9) with  $p = \bar{p}$  that the set of Lagrange multipliers for the latter problem at  $(\bar{x}, \bar{p})$  agrees with that for problem (3.5) at  $\bar{x}$ . Combining these observations tells us that the set of Lagrange multipliers for the parameterized problem (5.8) at  $(\bar{x}, \bar{p})$  is the singleton  $\{\bar{\lambda}\}$ . Since

$$\nabla_{xx}^2 \varphi(\bar{x}, \bar{p}) + \nabla_{xx}^2 \langle \bar{\lambda}, g \rangle(\bar{x}, \bar{p}) = \nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) \quad \text{and} \quad \nabla_x g(\bar{x}, \bar{p}) = \nabla f(\bar{x}),$$

and since the second-order sufficient condition (4.6) holds for (3.5) at  $(\bar{x}, \bar{\lambda})$ , we get

$$\begin{aligned} & \langle [\nabla_{xx}^2 \varphi(\bar{x}, \bar{p}) + \nabla_{xx}^2 \langle \bar{\lambda}, g \rangle(\bar{x}, \bar{p})] w, w \rangle + d^2 \delta_\Theta(f(\bar{x}, \bar{p}), \bar{\lambda}) (\nabla_x g(\bar{x}, \bar{p}) w) \\ &= \langle \nabla_{xx}^2 L(\bar{x}, \bar{\lambda}) w, w \rangle + d^2 \delta_\Theta(f(\bar{x}), \bar{\lambda}) (\nabla f(\bar{x}) w) > 0 \\ & \text{for all } w \in \mathbb{R}^n \setminus \{0\} \text{ with } \nabla_x g(\bar{x}, \bar{p}) w = \nabla f(\bar{x}) w \in K_\Theta(f(\bar{x}), \bar{\lambda}). \end{aligned}$$

This verifies therefore that the second-order sufficient condition of type (4.6) for problem (5.8) is satisfied at  $((\bar{x}, \bar{p}), \bar{\lambda})$ . Employing further Theorem 3.4(ii) ensures that  $\bar{x}$  a strict local minimizer for (5.8) as  $p = \bar{p}$ . Then it follows from Theorem 4.4 that there exists  $\varepsilon > 0$  such that for all  $(u, \mu) = p \in \mathbb{B}_\varepsilon(\bar{p})$  the parameterized problem (5.8) admits a local minimizer  $x_p$  with  $x_p \rightarrow \bar{x}$  as  $p \rightarrow \bar{p}$ . This concludes the proof of assertion (i).

Turing next to (ii), Turing now to (ii), observed first that the KKT system of (5.6) is the same is the KKT system of the parametric problem (5.8). By Proposition 4.4(ii), for any  $p = (u, \mu)$  close to  $\bar{p} := (\bar{x}, \bar{\lambda})$  the generalized equation (5.7) admits a solution  $(x_p, \lambda_p)$ . Moreover, (4.14) ensures that  $(x_p, \lambda_p) \rightarrow (\bar{x}, \bar{\lambda})$  as  $p \rightarrow \bar{p}$ , which completes the proof of (ii).  $\square$

To the best of our knowledge, the solvability of subproblems in the basic SQP method for NLPs was first established by Robinson [26, Theorem 3.1] who assumed, in addition to the second-order sufficient condition, the validity of the linear independence constraint qualification together with the strict complementarity condition. This result was improved by Bonnans [1, Proposition 6.3] for NLPs satisfying the second-order sufficient condition (4.6) and having a unique Lagrange multiplier. Theorem 5.2 significantly extends Bonnans' result to a general class of nonpolyhedral problems with parabolically regular constraint sets.

We are finally in a position to derive the superlinear convergence of the basic SQP method in our general setting. To be precise, the aimed convergence is understood as the *superlinear* convergence of the primal-dual iterates  $(x_k, \lambda_k)$  to a given KKT solution  $(\bar{x}, \bar{\lambda})$  meaning that

$$\|(x_{k+1}, \lambda_{k+1}) - (\bar{x}, \bar{\lambda})\| = o(\|(x_k, \lambda_k) - (\bar{x}, \bar{\lambda})\|) \text{ as } k \rightarrow \infty.$$

**Theorem 5.3 (superlinear convergence of the basic SQP method under parabolic regularity).** *Let  $(\bar{x}, \bar{\lambda})$  be a solution to the KKT system (3.6) under the validity of the standing assumptions (H1) and (H2). Suppose in addition that the second-order sufficient condition (4.6) is satisfied at  $(\bar{x}, \bar{\lambda})$ , that the multiplier mapping  $M_{\bar{x}}$  is calm at  $((0, 0), \bar{\lambda})$ , and that  $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ . 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 5.1 with  $H_k$  taken from (5.3) generates an iterative sequence  $\{(x_k, \lambda_k)\} \subset \mathbb{R}^n \times \mathbb{R}^m$  that converges to  $(\bar{x}, \bar{\lambda})$ , and the rate of convergence is superlinear.*

**Proof.** Due to our results achieved above, we can fit into the framework of [15, Theorem 3.2] for the KKT system (3.6), which can equivalently written as the generalized equation (5.4). Indeed, the semistability and hemistability properties of  $(\bar{x}, \bar{\lambda})$  imposed in [15, Theorem 3.2] are the isolated calmness of the solution map  $S$  and the KKT well-posedness formulated in Theorem 5.2(ii), respectively. Then both of these requirements are satisfied by the results of Theorems 4.2 and 5.2. It allows us to deduce the claimed superlinear convergence of primal-dual iterates from the abstract result of [15, Theorem 3.2] for arbitrary generalized equations.  $\square$

Let us now compare Theorem 5.3 with known results in the literature.

**Remark 5.4 (comparison with known results on superlinear convergence of the basic SQP method).** We address here the following two issues:

(i) Theorem 5.3 significantly extends the sharpest local superlinear convergence result obtained in [1, Theorem 5.1] for NLPs to a large class of parabolically regular constrained optimization problems. The additional calmness requirement for the multiplier mapping (3.10) automatically holds not only for polyhedral problems, but also in the essentially more general settings; see the discussion above in front of Proposition 4.1. Note also that the results on SQP superlinear convergence for the particular class of second-order cone programs were given in [17, 32] under the strong regularity of the KKT system (3.6) that is a significantly more involved assumption in comparison with those imposed in Theorem 5.3.

(ii) The superlinear convergence of a sequence of the SQP iterates  $(x_k, \lambda_k)$ , while *not the existence* of such a sequence, can be derived from [4, Theorem 6.4] in which the superlinear convergence of the Newton method for the generalized equation (5.5) was obtained when the mapping  $g + F$  is strongly metrically subregular. Adopting the latter result to our framework of the generalized equation (5.4), the assumed strong metric subregularity is equivalent to the isolated calmness of the solution map  $S$  from (4.1). Furthermore, by Theorem 4.2 we can equivalently translate the imposed assumptions in Theorem 5.3 to the requirements that the solution map  $S$  enjoys the isolated calmness property and the point  $\bar{x}$  is a local minimizer for (3.5). This indicates that, in contrast to [4], we impose the additional assumption on local optimality of  $\bar{x}$ , and so one might ask whether it is possible to drop this condition with no harm. However, it happens not to be correct since [4, Theorem 6.4] *assumes* the existence of iterates  $(x_k, \lambda_k)$  (i.e., *solvability* of the subproblems) that always stay in a neighborhood of  $(\bar{x}, \bar{\lambda})$ . Below we present an example of an NLP problem considered in [15, Example 3.3], which shows that the isolated calmness of the solution map  $S$  alone is not enough to guarantee the solvability of the subproblems in our setting. Thus the *local optimality* of  $\bar{x}$  turns out to be an *essential* requirement to construct a well-posed and superlinearly convergent sequence of iterates in the basic SQP method even for simple problems of nonlinear programming.

**Example 5.5 (failure of solvability for subproblems in the basic SQP method).** Consider the one-dimensional nonlinear program

$$\min -\frac{1}{2}x + \frac{1}{6}x^3 \quad \text{subject to} \quad x \geq 0. \quad (5.10)$$

By definition (4.1) of the solution map  $S$  for this problem we have

$$S(v, w) = \{(x, \lambda) \in \mathbb{R}^2 \mid v = -x + \frac{1}{2}x^2 + \lambda \quad \text{and} \quad x + w \in N_{\mathbb{R}_-}(\lambda)\},$$

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 check that  $\Lambda(\bar{x}) = \{\bar{\lambda}\}$  and that  $\bar{x}$  is *not* a local minimizer for problem (5.10); indeed,  $x = 2$  is the unique minimizer for this problem. We show nevertheless that the solution map  $S$  enjoys the isolated calmness property at  $((0, 0), (\bar{x}, \bar{\lambda}))$ . Let us verify to this end that the KKT mapping  $G$  from (3.11) in the case of (5.10) is strongly metrically subregular at  $((\bar{x}, \bar{\lambda}), (0, 0))$  by using the criterion in (4.5). Observe that the mapping  $G$  for this framework reduces to

$$G(x, \lambda) = \begin{bmatrix} -x + \frac{1}{2}x^2 + \lambda \\ -x \end{bmatrix} + \begin{bmatrix} 0 \\ N_{\mathbb{R}_-}(\lambda) \end{bmatrix},$$

and its graphical derivative is easily calculated by

$$DG((\bar{x}, \bar{\lambda}), (0, 0))(\xi, \eta) = \begin{bmatrix} -1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} + \begin{bmatrix} 0 \\ DN_{\mathbb{R}_-}(\bar{\lambda}, \bar{x})(\eta) \end{bmatrix}.$$

Since  $DN_{\mathbb{R}_-}(\bar{\lambda}, \bar{x})(\eta) = N_{\mathbb{R}_-}(\eta)$ , we arrive at

$$(0, 0) \in DG((\bar{x}, \bar{\lambda}), (0, 0))(\xi, \eta) \iff \xi = \eta \quad \text{and} \quad \xi \in N_{\mathbb{R}_-}(\eta)$$

from which it clearly follows that  $\xi = \eta = 0$ . Then (4.5) tells us that the mapping  $G$  is strongly metrically subregular at  $((\bar{x}, \bar{\lambda}), (0, 0))$ .

Next we are going to show that the generalized equation (5.7) has no solution for all  $(u, \mu) \in \mathbb{B}_{1/2}(\bar{x}, \bar{\lambda})$  with  $u \neq \bar{x}$ . To furnish it, pick  $(u, \mu) \in \mathbb{B}_{1/2}(\bar{x}, \bar{\lambda})$  with  $u \neq \bar{x}$  and observe that for (5.10) the generalized equation in question is represented by

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} -u + \frac{1}{2}u^2 + \mu \\ -u \end{bmatrix} + \begin{bmatrix} -1 + u & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x - u \\ \lambda - \mu \end{bmatrix} + \begin{bmatrix} 0 \\ N_{\mathbb{R}_-}(\lambda) \end{bmatrix}.$$

This leads us to the relationships

$$\lambda = \frac{1}{2}u^2 - x(u - 1), \quad x\lambda = 0, \quad x \geq 0, \quad \lambda \leq 0.$$

From the second condition above we deduce that either  $x = 0$  or  $\lambda = 0$ . If the former holds, it follows from the first equation that  $\lambda = \frac{1}{2}u^2 > 0$ , a contradiction. If the latter holds, we get  $x = \frac{u^2}{u-1} < 0$ , a contradiction as well. This justifies that the KKT system associated with the subproblems of the basic SQP method for (5.10) has no solution for such a pair  $(u, \mu)$ .

Since we have  $\Lambda(\bar{x}) = \{\bar{\lambda}\}$ , the basic constraint qualification (3.14) is satisfied. This tells us that if subproblems (5.6) corresponding to (5.10) have local optimal solutions associated with  $(u, \mu)$ , then we would end up with having a solution for the generalized equation (5.7), which is not possible. This demonstrates the failure of solvability for subproblems in the basic SQP.

Let us mention that assuming *metric regularity* vs. strong metric subregularity of the mapping  $G = S^{-1}$  from (3.11) ensures the solvability of subproblems in the basic SQP method; cf. [8, Theorem 6D.2]. However, the next theorem shows that in our framework the metric regularity yields the strong metric subregularity of the mapping  $G$ , which is equivalent to the isolated calmness of solution map  $S$  from (4.1). Recall to this end that the isolated calmness of  $S$  alone does not ensure the existence of the SQP iterates as shown in Example 5.5.

**Theorem 5.6 (strong metric subregularity from metric regularity for KKT systems).** *Let  $(\bar{x}, \bar{\lambda})$  be a solution to the KKT system (3.6) under the standing assumptions (H1) and (H2). If the mapping  $G$  from (3.11) is metrically regular around  $((\bar{x}, \bar{\lambda}), (0, 0))$ , then it is strongly metrically subregular at this point.*

**Proof.** According to the coderivative criterion for metric regularity (2.6), the mapping  $G$  enjoys this property around  $((\bar{x}, \bar{\lambda}), (0, 0))$  if and only if

$$(0, 0) \in D^*G((\bar{x}, \bar{\lambda}), (0, 0))(w, u) \implies w = 0, u = 0. \quad (5.11)$$

It is not hard to see that for any  $(w, u) \in \mathbb{R}^n \times \mathbb{R}^m$  the coderivative of  $G$  is calculated by

$$D^*G((\bar{x}, \bar{\lambda}), (0, 0))(w, u) = \begin{bmatrix} \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w - \nabla f(\bar{x})^*u \\ \nabla f(\bar{x})w + D^*N_\Theta^{-1}(\bar{\lambda}, f(\bar{x}))(u) \end{bmatrix}.$$

As follows from [8, Theorem 4G.1],  $G$  is strongly metrically subregular at  $((\bar{x}, \bar{\lambda}), (0, 0))$  if and only if we have the implication

$$(0, 0) \in DG((\bar{x}, \bar{\lambda}), (0, 0))(w, u) \implies w = 0, u = 0. \quad (5.12)$$

Furthermore, it is easy to calculate as in (4.5) that

$$DG((\bar{x}, \bar{\lambda}), (0, 0))(w, u) = \begin{bmatrix} \nabla_{xx}^2 L(\bar{x}, \bar{\lambda})w + \nabla f(\bar{x})^*u \\ -\nabla f(\bar{x})w + DN_\Theta^{-1}(\bar{\lambda}, \Phi(\bar{x}))(u) \end{bmatrix}$$

for all  $(w, u) \in \mathbb{R}^n \times \mathbb{R}^m$ . We get from Theorem 3.2(ii) under our standing assumptions (H1) and (H2) that the normal cone mapping  $N_\Theta$  is proto-differentiable at  $f(\bar{x})$  for  $\bar{\lambda}$ . The latter together with [29, Theorem 13.57] gives us the *derivative-coderivative inclusion*

$$DN_\Theta(f(\bar{x}), \bar{\lambda})(\xi) \subset D^*N_\Theta(f(\bar{x}), \bar{\lambda})(\xi) \quad \text{for all } \xi \in \mathbb{R}^m. \quad (5.13)$$

Pick now  $(w, u) \in \mathbb{R}^n \times \mathbb{R}^m$  from the left-hand side of implication (5.12) and deduce from the derivative-coderivative inclusion (5.13) that  $(w, -u)$  satisfies the relation on the left-hand side

of implication (5.11). Hence  $(w, u) = (0, 0)$ , which yields (5.12), and thus ensures the strong metric subregularity of  $G$  at  $((\bar{x}, \bar{\lambda}), (0, 0))$ .  $\square$

A similar result to the one in Theorem 5.6 was obtained in [8, Lemma 4F.8] for NLPs by using the graphical derivative characterization of the isolated calmness in (4.5). It was extended in [6, Corollary 25] to  $\mathcal{C}^2$ -cone reducible constrained optimization problems by using a delicate topological result by Fusek [10]. The obtained Theorem 5.6 further extends the aforementioned results to the general case of parabolically regular constrained problems by employing a completely different device based on the coderivative criterion, the derivative-coderivative inclusion, and our recent second-order developments on parabolic regularity.

## 6 Conclusions

This paper provides first applications of the novel theory of parabolic regularity in second-order variational analysis to characterizations of robust stability properties of KKT systems in general nonpolyhedral problems of constrained optimization with the subsequent usage of them to the development and justification of numerical algorithms. In particular, we derive complete second-order characterizations of robust isolated calmness for KKT systems with respect to arbitrary (not just canonical) perturbations. The obtained results are applied to developing the basic SQP method of solving constrained optimization problems governed by parabolically regular sets. This approach allows us to fully justify well-posedness of the basic SQP method (including solvability and stability of the corresponding subproblems) and its primal-dual superlinear convergence for parabolically regular constrained problems. Most of the obtained results are new not only in the general framework under consideration, but also for any conventional setting of nonpolyhedral constrained optimization.

Our future research aims at further developments of the presented approach to make it powerful in applications to more involved and efficient versions of the generic SQP algorithm as well as in other primal-dual methods of constrained optimization.

## References

- [1] J. F. BONNANS, *Local analysis of Newton-type methods for variational inequalities and nonlinear programming*, Appl. Math. Optim., 29 (1994), pp. 161–186.
- [2] J. F. BONNANS, J. CH. GILBERT, C. LEMARÉCHAL AND C. SAGASTIZABAL, *Numerical Optimization: Theoretical and Practical Aspects*, 2nd ed., Springer, Berlin, 2006.
- [3] J. F. BONNANS AND A. SHAPIRO, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
- [4] R. CIBULKA, A. L. DONTCHEV AND A. KRUGER, *Strong metric subregularity of mappings in variational analysis and optimization*, J. Math. Anal. Appl., 457 (2018), 1247–1282.
- [5] Y. CUI, D. SUN AND K.-C. TOH, *On the R-superlinear convergence of the KKT residues generated by the augmented Lagrangian method for convex composite conic programming*, published online in Math. Program.; <https://doi.org/10.1007/s10107-018-1300-6>.
- [6] 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), pp. 67–90.
- [7] A. L. DONTCHEV AND R. T. ROCKAFELLAR, *Characterizations of Lipschitzian stability in nonlinear programming*, in Mathematical Programming With Data Perturbations (A. V. Fiacco, ed.), pp. 65–82, Marcel Dekker, New York, 1997.
- [8] A. L. DONTCHEV AND R. T. ROCKAFELLAR, *Implicit Functions and Solution Mappings: A View from Variational Analysis*, 2nd edition, Springer, New York, 2014.

[9] F. FACCHINEI AND J.-S. PANG, *Finite-Dimensional Variational Inequalities and Complementary Problems*, published in two volumes, Springer, New York, 2003.

[10] P. FUSEK, *Isolated zeros of Lipschitzian metrically regular  $\mathbb{R}^n$ -functions*, Optimization, 49 (2001), pp. 425–446.

[11] H. GFRERER, *First-order and second-order characterizations of metric subregularity and calmness of constraint set mappings*, SIAM J. Optim., 21 (2011), pp. 1439–1474.

[12] N. T. V. HANG, B. S. MORDUKHOVICH AND M. E. SARABI, *Second-order variational analysis in second-order cone programming*, published online in Math. Program.; <https://doi.org/10.1007/s10107-018-1345-6>.

[13] R. HENRION AND J. V. OUTRATA, *Calmness of constraint systems with applications*, Math. Program., 104 (2005), pp. 437–464.

[14] A. F. IZMAILOV, *On the analytical and numerical stability of critical Lagrange multipliers*, Comput. Math. Math. Phys., 45 (2005), pp. 930–946.

[15] A. F. IZMAILOV AND M. V. SOLODOV, *Newton-Type Methods for Optimization and Variational Problems*, Springer, New York, 2014.

[16] 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).

[17] H. KATO AND M. FUKUSHIMA, *An SQP-type algorithm for nonlinear second-order cone programs*, Optim. Lett. 1 (2007), pp. 129–144.

[18] D. KLATTE AND B. KUMMER, *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods, and Applications*, Kluwer, Boston, MA, 2002.

[19] A. MOHAMMADI, B. S. MORDUKHOVICH AND M. E. SARABI, *Parabolic regularity in geometric variational analysis*, preprint (2019), 48 pp.; [arxiv.org/abs/1909.00241](https://arxiv.org/abs/1909.00241).

[20] B. S. MORDUKHOVICH, *Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions*, Trans. Amer. Math. Soc., 340 (1993), pp. 1–35.

[21] B. S. MORDUKHOVICH, *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Application*, Springer, Berlin, 2018.

[22] B. S. MORDUKHOVICH, *Variational Analysis and Applications*, Springer, Cham, Switzerland, 2018.

[23] B. S. MORDUKHOVICH, J. V. OUTRATA AND H. RAMÍREZ, *Graphical derivatives and stability analysis for parameterized equilibria with conic constraints*, Set-Valued Var. Anal., 23 (2015), pp. 687–704.

[24] B. S. MORDUKHOVICH AND M. E. SARABI, *Critical multipliers in variational systems via second-order generalized differentiation*, Math. Program., 169 (2018), pp. 605–648.

[25] B. S. MORDUKHOVICH AND M. E. SARABI, *Criticality of Lagrange multipliers in variational systems*, SIAM J. Optim., 29 (2019), pp. 1524–1557.

[26] S. M. ROBINSON, *Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms*, Math. Program. 7 (1974), pp. 1–16.

[27] S. M. ROBINSON, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 3–62.

[28] S. M. ROBINSON, *Some continuity properties of polyhedral multifunctions*, Math. Program. Study, 14 (1981), pp. 206–214.

[29] R. T. ROCKAFELLAR AND R. J-B. WETS, *Variational Analysis*, Springer, Berlin, 1998.

[30] A. SHAPIRO, *Duality, optimality conditions and perturbation analysis*, in: Semidefinite Programming and , Applications Handbook, Kluwer Academic Publishers, Boston, pp. 67–92, 2000.

[31] J. J. YE AND X. Y. YE, *Necessary optimality conditions for optimization problems with variational inequality constraints*, Math. Oper. Res., 22 (1997), pp. 977–997.

- [32] Y. WANG AND L. ZHANG, *Properties of equation reformulation of the Karush-Kuhn-Tucker condition for nonlinear second-order cone optimization problems*, Math. Meth. Oper. Res., 70 (2009), pp. 195–218.
- [33] X. Y. ZHENG AND K. F. NG, *Metric subregularity and calmness for nonconvex generalized equations in Banach spaces*, SIAM J. Optim., 20 (2010), pp. 2119–2136.