LOCAL CONVERGENCE OF STABILIZED SEQUENTIAL QUADRATIC PROGRAMMING AND SEQUENTIAL QUADRATICALLY CONSTRAINED QUADRATIC PROGRAMMING METHODS, AND THEIR EXTENSIONS TO VARIATIONAL PROBLEMS

Doctoral Thesis by
Damián R. Fernández Ferreyra

Supervised by
Mikhail V. Solodov

IMPA - Instituto Nacional de Matemática Pura e Aplicada
Rio de Janeiro - 2008

Acknowledgments

I would like to thank to my adviser, Mikhail Solodov, for showing me modern tools to solve a widely range of relevant problems in nonlinear programming. Thanks to Claudia Sagastizábal for their always welcome suggestions. Thanks to Alfredo Iusem for his attention and willingness to talk about any subject, even beyond the academic affairs. I thank also the significant contributions of the other members of the committee: José Mario Martinez, Luis Mauricio Graña Drummond and Susana Scheimberg.

I can not forget to thank all the students of IMPA with whom I shared my days in Rio de Janeiro, and especially to those with whom I shared many deep and meaningful philosophical conversations at the Bina Bar.

I thank to CNPq and CAPES (PEC-PG – Programa de Estudante-Convênio de Pós-Graduação) for the financial support.

Abstract

This work is devoted to the analysis of local convergence of some recently proposed Newton–type methods for solving optimization problems. We also develop natural extensions of those techniques to the setting of variational problems. We emphasize that all results are new already in the optimization setting, while variational extensions can be considered as additional contributions.

The first method considered is the stabilized Sequential Quadratic Programming (sSQP) algorithm, proposed by S.J. Wright and further studied by W.W. Hager, by A. Fischer, and by S.J. Wright. This method has been developed for preserving superlinear/quadratic convergence in the case of nonunique Lagrange multipliers (i.e., degenerate constraints), which is a challenging problem in contemporary optimization. Previously, convergence of sSQP was proven either under the Mangasarian–Fromovitz constraint qualification and the second-order sufficient condition for optimality, or under the strong second-order sufficient condition for optimality. We prove primal-dual quadratic convergence assuming only the second-order sufficient condition for optimality, without any constraint qualification assumptions, thus improving significantly over any of the previous results. In addition, we introduce a natural extension of sSQP techniques to the variational setting. We show also that in the variational setting, under a suitable second-order condition, the so-called natural residual provides a local error bound for the solution set of the associated Karush–Kuhn–Tucker (KKT) system. This is the first error bound for variational KKT systems that does not assume anything about the constraints of the underlying variational problem.

The second method considered is the Sequential Quadratically Constrained Quadratic Programming (SQCQP) algorithm, studied by M. Anitescu, and by M. Fukushima, Z.-Q. Luo and P. Tseng. Previously, convergence of SQCQP was proven either under the Mangasarian–Fromovitz constraint qualification and the quadratic growth condition, or under the convexity assumptions, the Slater constraint qualification and the strong second-order sufficient condition for optimality. We prove a new primal-dual quadratic convergence result assuming the linear independence constraint qualification, strict complementarity, and the usual second-order sufficient condition for optimality. This result is complementary to the two previous ones, being neither weaker nor stronger. In addition, we obtain superlinear convergence of the primal sequence under a Dennis-Moré type condition, as well as extend the method to variational problems.

Key words. Stabilized sequential quadratic programming, sequential quadratically constrained quadratic programming, Karush-Kuhn-Tucker system, variational inequalities, Newton methods, superlinear convergence, quadratic convergence, error bound, Dennis-Moré condition.

This page intentionally *left* blank.

Contents

Introduction

The finite-dimensional variational inequality problem provides a broad unifying setting for the study of optimization, equilibrium, and related problems, and serves as a useful computational framework for solution of a host of applications in the mathematical sciences.

The subject of variational inequalities has its origin in the calculus of variations associated with the minimization of infinite-dimensional functionals. The systematic study of the subject began in the early 1960s and it was used as an analytic tool for studying free boundary problems defined by nonlinear partial differential operators arising from unilateral problems in elasticity and plasticity theory and in mechanics. In real-life applications of infinite-dimensional problems, the discretization strategy is mainly determined by the solvers available for finite-dimensional problems. Hence, any procedure should aim at incorporating existing software and avoid subproblems for which no computational methods are available.

The development of the finite-dimensional variational inequalities also began in the early 1960s but followed a different path. Unlike its infinite-dimensional counterpart, which was conceived in the area of partial differential systems, the finite-dimensional variational inequality was born in the domain of Mathematical Programming. In the recent monograph, F. Facchinei and J.-S. Pang [10] present state-of-the-art in finite-dimensional variational inequalities and complementarity problems. We shall use not only some results from [10], but also its notation and terminology, for the most part.

This work is organized in three chapters, as follows. In the first chapter, we recall the fundamental Newton method for solving smooth nonlinear equations and the Josephy-Newton method for solving generalized (set-valued) equations. In both cases, we relate local behavior of the method with sensitivity analysis of the problem in question, following the modern view of Variational Analysis [27, 37]. In order to introduce computational methods for optimization and their properties, we present a brief review of nonlinear programming theory and terminology, with emphasis on the use of constraint qualifications and optimality conditions, including the second-order sufficient conditions for optimality. We present the Sequential Quadratic Programming (SQP) method as a particular case of Newton or Josephy-Newton method, depending on whether inequality constraints are present or not in the problem. To deal with nonlinear programs with nonunique Lagrange multipliers (degenerate constraints), we present Wright's stabilized SQP algorithm and give a summary of its local convergence theory. In this introductory chapter, globalization of SQP is considered only as a motivation to present a modification of SQP, called the Sequential Quadratically Constrained Quadratic Programming (SQCQP) method. We describe briefly some facts about local and global convergence of SQCQP. To complete the background material, we introduce the variational problem that will be the main focus of our further development, with optimization being a special case. We also mention relations with other well-known problems and extend some concepts from the optimization case to the variational context.

In the second chapter, we introduce a stabilized Newton method for solving variational problems that generalizes sSQP for optimization. We prove that this method has a local primal-dual quadratic convergence rate, assuming only that the starting point is close to a primal-dual pair satisfying a suitable second-order condition (in the case of optimization, this assumption coincides with the standard second-order sufficient condition for optimality). We emphasize that no constraints qualifications are assumed, which gives a significantly stronger result than any other one in the existing literature.

The third chapter describes an extension of the Sequential Quadratically-Constrained Quadratic-Programming method for variational problems. We show that the generated primal-dual sequence converges locally at a quadratic rate under the assumptions of the linear independence constraint qualification, strict complementarity and a suitable second-order condition. In the case of optimization, this result gives a new property, complementing what has been known before. Also, we provide a necessary and sufficient condition for superlinear convergence of the primal sequence under a Dennis-Moré type condition.

Basic notation and terminology

$\mathbb{R}^n$: the n-dimensional Euclidean space,
$\mathbb{R}^n_+ = \{x \in \mathbb{R}^n \mid x_i \geq 0 \;\; i = 1, \ldots, n\}$: the nonnegative orthant,
$\mathbb{R}_+ = \mathbb{R}^1_+$: nonnegative real numbers,
$\langle \cdot, \cdot \rangle$: the Euclidean inner product,
$\| \cdot \|$: the Euclidean norm,
B: the open unit ball (the underlying space is always clear from the context),
I: the identity matrix (the dimension is always clear from the context),
$M^\top$: the transpose of a matrix $M \in \mathbb{R}^{m \times n}$,
$M_{\mathcal{I}}$: submatrix of $M \in \mathbb{R}^{m \times n}$ with rows indexed by the set $\mathcal{I} \subseteq \{1, \ldots, m\}$,
$x_{\mathcal{I}}$: subvector of $x \in \mathbb{R}^n$ with coordinates indexed by $\mathcal{I} \subseteq \{1, \ldots, n\}$,
$u^\perp = \{v \in \mathbb{R}^l \mid \langle u, v \rangle = 0\}$: orthogonal complement of $u \in \mathbb{R}^l$,
$\xi(t) = o(t)$: any function $\xi : \mathbb{R}_+ \to \mathbb{R}^q$ such that $\lim_{t \to 0} t^{-1} \|\xi(t)\| = 0$,
$\phi(t) = O(t)$: any function $\phi : \mathbb{R}_+ \to \mathbb{R}^q$ such that $\limsup_{t \to 0} t^{-1} \|\phi(t)\| < \infty$,
$t_k \to 0^+$: when $t_k \to 0$ and $\{t_k\} \subset \mathbb{R}_+ \; \forall k$,
$\Psi'(\bar{x}, \bar{\mu})$: the full derivative of $\Psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$ at the point $(\bar{x}, \bar{\mu})$,
$\Psi'_x(\bar{x}, \bar{\mu})$: the partial derivative of Ψ with respect to x at the point $(\bar{x}, \bar{\mu})$,
$\mathrm{dist}(z, S) = \inf_{s \in S} \|z - s\|$: the distance from $z \in \mathbb{R}^l$ to a nonempty set $S \subset \mathbb{R}^l$,
$\Pi_S(z)$: the Euclidean projection of $z \in \mathbb{R}^l$ on the set $S \subset \mathbb{R}^l$,
$K^* = \{u \in \mathbb{R}^l \mid \langle u, v \rangle \geq 0 \quad \forall v \in K\}$: (positive) dual of the cone $K \subset \mathbb{R}^l$,
$\mathrm{gph}\, \Xi = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m \mid y \in \Xi(x)\}$: the graph of a multifunction Ξ from $\mathbb{R}^n$ to subsets of $\mathbb{R}^m$,
$\Xi^{-1}(y) = \{x \in \mathbb{R}^n \,|\, y \in \Xi(x)\}$: inverse graph of the multifunction Ξ,

Recall that a set $K \subset \mathbb{R}^l$ is a cone if $v \in K$ implies that $tv \in K$ for all $t \in \mathbb{R}_+$. A matrix $M \in \mathbb{R}^{l \times l}$ is said to be copositive on a cone $K \subset \mathbb{R}^l$ if $\langle Mv, v \rangle \geq 0$ for all $v \in K$, and strictly copositive if this inequality is strict for all $v \in K \setminus \{0\}$.

Chapter 1

Background Material

In this chapter, we collect some basic facts about Newton-type methods, first and second order optimality conditions, constraint qualifications, and variational problems, that will be used in the sequel. We also survey previous results on stabilized sequential quadratic programming and sequential quadratically constrained quadratic programming that motivate our contributions.

1.1 Newton Method

A very important tool to solve optimization and variational problems is the Newton Method. The classical Newton algorithm for smooth equations serves as a prototype of many procedures, as it reflects some fundamental principles that lead to fast local convergence.

Consider first the problem

$$\text{find } z \in \mathbb{R}^n \text{ s.t. } G(z) = 0, \tag{1.1}$$

where the function $G : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. The main idea of the Newton Method is to replace the function G by its linearization at the current iterate, resulting in an approximate problem that can be solved more easily. Specifically, given an iterate z^k, we define the next *Newton iterate* z^{k+1} as the solution of the *linear* equation

$$\text{find } z \in \mathbb{R}^n \text{ s.t. } G(z^k) + G'(z^k)(z - z^k) = 0. \tag{1.2}$$

To begin our discussion of local analysis of the method, we recall the result that states the well-definedness of the Newton sequence $\{z^k\}$ and its convergence to a zero $\bar{z}$ of G. It, also shows that the convergence rate is fast. To this end, let us formalize this concept: let $\{z^k\} \subset \mathbb{R}^n$ be a sequence of vectors tending to the limit $\bar{z} \neq z^k$ for all k. We will say that the convergence rate is (at least) linear if

$$\limsup_{k\to\infty} \frac{\|z^{k+1} - \bar{z}\|}{\|z^k - \bar{z}\|} < 1;$$

superlinear if

$$\lim_{k\to\infty} \frac{\|z^{k+1} - \bar{z}\|}{\|z^k - \bar{z}\|} = 0;$$

and quadratic if

$$\limsup_{k\to\infty} \frac{\|z^{k+1} - \bar{z}\|}{\|z^k - \bar{z}\|^2} < \infty.$$

In each case, we say that $\{z^k\}$ converges to $\bar{z}$ (at least) linearly, superlinearly and quadratically, respectively.

We can now state the well known convergence result of the Newton Method.

Theorem 1.1.1 *Let $G : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable in a neighborhood of $\bar{z} \in \mathbb{R}^n$ satisfying $G(\bar{z}) = 0$. Suppose that $G'(\bar{z})$ is nonsingular. Then there exists a neighborhood $\mathcal{V}$ of $\bar{z}$ such that if $z^0 \in \mathcal{V}$, the Newton sequence $\{z^k\}$ is well-defined and converges superlinearly to $\bar{z}$. Furthermore, if G' is Lipschitz-continuous in a neighborhood of $\bar{z}$, then the convergence rate is quadratic.*

We remark that the nonsingularity of $G'(\bar{z})$ implies not only local uniqueness of the solution $\bar{z}$, but also that the solution of the equation in (1.1) remains locally unique under small perturbations of G of the form $G(z)+p$, $p \in \mathbb{R}^n$. To see this property, note that by the Implicit Function Theorem there exist $\gamma > 0$ and a function $z : \mathbb{R}^n \to \mathbb{R}^n$, defined implicitly by $0 = G(z(p)) + p$, that is well-defined and Lipschitz-continuous (in particular, single-valued) for $p \in \gamma B$.

Also, the stability property of problem (1.1) can be seen as a certain regularity condition of the solution set of the perturbed problem. To formulate this condition, we say that a closed multifunction Ξ from $\mathbb{R}^n$ to the subsets of $\mathbb{R}^n$ has the *Aubin property* at $(p^0, z^0) \in \operatorname{gph}\Xi$ if there exist $\varepsilon_0, \gamma_0, \ell_0 > 0$ such that

$$\Xi(p^2) \cap (z^0 + \varepsilon_0 B) \subset \Xi(p^1) + \ell_0\|p^2 - p^1\|B \qquad \forall p^1, p^2 \in p^0 + \gamma_0 B. \tag{1.3}$$

Some authors refer to this concept saying that Ξ is *pseudo-Lipschitz-continuous* at the point (p^0, z^0).

For our purposes, if we define the multifunction

$$\Sigma(p) = \{z \in \mathbb{R}^n \mid 0 = G(z) + p\},$$

which is the solution set of the perturbation of problem (1.1), then the nonsingularity of $G'(\bar{z})$ implies that for all $p \in \gamma B$ the multifunction $\Sigma(p)$ is single-valued and has the Aubin property at $(0, \bar{z})$.

Another multifunction that plays an important role in convergence analysis of Newton-type methods is

$$\mathrm{L}\Sigma(p) = \{z \in \mathbb{R}^n \mid 0 = G(\bar{z}) + G'(\bar{z})(z - \bar{z}) + p\},$$

which is the solution set of the perturbed linearization of the problem (1.1) at the point $\bar{z}$.

To arrive to concepts appropriate for convergence analysis of Newton-type methods in the more general settings where nonsingularity of $G'(\bar{z})$ is no longer a relevant regularity condition, we point out that when G is continuously differentiable at $\bar{z}$ and $G(\bar{z}) = 0$, then the following statements are equivalent:

(i) $G'(\bar{z})$ is nonsingular;

(ii) Σ has the Aubin property at $(0, \bar{z})$;

(iii) $\mathrm{L}\Sigma$ has the Aubin property at $(0, \bar{z})$.

In the set-valued settings, where property (i) is no longer the appropriate condition, properties (ii) or (iii) are still relevant and meaningful, and can be used to obtain convergence of Newton-type methods. Furthermore, depending on the specific problem, these properties can usually be translated into constructive conditions on the problem data.

By the Implicit Function Theorem (i)⇒(ii).

For completeness, we give the proof of (ii)⇒(iii), adapting that of [27, Theorem 4.69] (we note that while the assertion is in principle not new, there seems to be no proof in the literature which is as simple as the one we give below for our setting). By (ii), there exist $\varepsilon, \gamma, \ell > 0$ such that

$$\Sigma(p) \cap (\bar{z} + \varepsilon B) \subset \Sigma(\hat{p}) + \ell \|p - \hat{p}\| B \qquad \forall p, \hat{p} \in \gamma B.$$

Take ε small enough such that $\varepsilon\ell < 1$. Since G is continuously differentiable at $\bar{z}$, we have that exists $\delta_0 > 0$ such that

$$\|G(z) - G(\hat{z}) - G'(\bar{z})(z - \hat{z})\| \le \varepsilon \|z - \hat{z}\| \qquad \forall z, \hat{z} \in \bar{z} + \delta_0 B.$$

Choose $\delta < \min\{\delta_0, \varepsilon, \gamma/\varepsilon\}$ and $\beta < \min\{\delta(1 - \varepsilon\ell)/4\ell, \gamma - \varepsilon\delta\}$. Fix $p, \hat{p} \in \beta B$ and let $z \in \mathrm{L}\Sigma(p) \cap (\bar{z} + \frac{\delta}{2} B)$. Note that $z \in \Sigma(q) \cap (\bar{z} + \frac{\delta}{2} B)$ for $q = p - G(z) + G'(\bar{z})(z - \bar{z})$ and that

$$\|q\| \le \|p\| + \|G(\bar{z}) - G(z) - G'(\bar{z})(\bar{z} - z)\| \le \beta + \varepsilon\delta/2 < \gamma.$$

Similarly, we have that $\hat{p} - G(\hat{z}) + G'(\bar{z})(\hat{z} - \bar{z}) \in \gamma B$ for any $\hat{z} \in \bar{z} + \delta B$. Let $z^1 = z$, then by the Aubin property of Σ there exists $z^2 \in \Sigma\left(\hat{p} - G(z^1) + G'(\bar{z})(z^1 - \bar{z})\right)$ such that

$$\|z^1 - z^2\| \le \ell \|q - [\hat{p} - G(z^1) + G'(\bar{z})(z^1 - \bar{z})]\| = \ell \|p - \hat{p}\|.$$

By induction, suppose that there exist $z^2, \ldots, z^{n-1}$ with

$$z^i \in \Sigma\left(\hat{p} - G(z^{i-1}) + G'(\bar{z})(z^{i-1} - \bar{z})\right)$$

and $\|z^i - z^{i-1}\| \leq \ell\|p-\hat{p}\|(\varepsilon\ell)^{i-2}$ for $i = 2, \ldots, n-1$. Then, by the choice of β we have

$$\begin{aligned}\|z^i - \bar{z}\| &\leq \|z^1 - \bar{z}\| + \sum_{j=2}^{i} \|z^j - z^{j-1}\| \leq \frac{\delta}{2} + \ell\|p - \hat{p}\| \sum_{j=2}^{i} (\varepsilon\ell)^{j-2} \\ &\leq \frac{\delta}{2} + \frac{\ell}{1-\varepsilon\ell}\|p - \hat{p}\| \leq \frac{\delta}{2} + \frac{2\ell\beta}{1-\varepsilon\ell} < \delta.\end{aligned}$$

Then, $\hat{p} - G(z^i) + G'(\bar{z})(z^i - \bar{z}) \in \gamma B$ for $i = 2, \ldots, n-1$. Again, by the Aubin property of Σ, there exists $z^n \in \Sigma\left(\hat{p} - G(z^{n-1}) + G'(\bar{z})(z^{n-1} - \bar{z})\right)$ such that

$$\begin{aligned}\|z^n - z^{n-1}\| &\leq \ell\|G(z^{n-2}) - G(z^{n-1}) - G'(\bar{z})(z^{n-2} - z^{n-1})\| \\ &\leq \ell\varepsilon\|z^{n-1} - z^{n-2}\| \leq \ell\|p - \hat{p}\|(\varepsilon\ell)^{n-2}.\end{aligned}$$

We thus get $\|z^n - z^{n-1}\| \to 0$. Hence, $\{z^n\}$ is a Cauchy sequence converging to some $z^\infty \in \bar{z} + \delta B$. By the definition of Σ, we have that

$$0 = G(z^n) - G(z^{n-1}) + G'(\bar{z})(z^{n-1} - \bar{z}) + \hat{p}.$$

Taking limit as $n \to \infty$, we obtain $0 = G'(\bar{z})(z^\infty - \bar{z}) + \hat{p}$, i.e., $z^\infty \in \mathrm{L}\Sigma(\hat{p})$. Since

$$\|z^n - z\| \leq \sum_{i=2}^{n} \|z^i - z^{i-1}\| \leq \ell\|p - \hat{p}\| \sum_{i=2}^{n} (\varepsilon\ell)^{i-2} \leq \frac{\ell}{1-\varepsilon\ell}\|p - \hat{p}\|,$$

thus, taking limits, we have $\|z - z^\infty\| \leq \frac{\ell}{1-\varepsilon\ell}\|p - \hat{p}\|$.

The proof of (iii)⇒(ii) is a particular case of the proof of [9, Theorem 1], that we include here for completeness. By (iii), there exist $\varepsilon, \gamma, \ell > 0$ such that

$$\mathrm{L}\Sigma(p) \cap (\bar{z} + \varepsilon B) \subset \mathrm{L}\Sigma(\hat{p}) + \ell\|p - \hat{p}\|B \qquad \forall p, \hat{p} \in \gamma B.$$

Let $H(z) = G'(\bar{z})(z - \bar{z})$ and take $\delta < \min\{\varepsilon, \gamma/\|G'(\bar{z})\|\}$ (note that if $G'(\bar{z}) = 0$ then $\mathrm{L}\Sigma$ does not satisfy the Aubin property). We will show that $H(\bar{z} + \delta B)$ is open, and therefore, $G'(\bar{z})$ nonsingular. Choose $\hat{z} \in \bar{z} + \delta B$ and define $\hat{p} = -H(\hat{z})$. Consider a sequence $\{p^k\}$ such that $p^k \to \hat{p}$. Since

$$\|\hat{p}\| = \|H(\hat{z})\| \leq \|G'(\bar{z})\|\|\hat{z} - \bar{z}\| \leq \delta\|G'(\bar{z})\| < \gamma,$$

then $p^k, \hat{p} \in \gamma B$ for k large enough. Thus we have

$$\mathrm{L}\Sigma(\hat{p}) \cap (\bar{z} + \varepsilon B) \subset \mathrm{L}\Sigma(p^k) + \ell\|\hat{p} - p^k\|B.$$

By definition of $\hat{p}$ and the choice of δ, we have that $\hat{z} \in \mathrm{L}\Sigma(\hat{p}) \cap (\bar{z} + \varepsilon B)$. Hence, there exists a sequence $\{z^k\}$ such that $z^k \in \mathrm{L}\Sigma(p^k)$ and $\|\hat{z} - z^k\| \leq \ell\|\hat{p} - p^k\|$. This implies that $0 = G'(\bar{z})(z^k - \bar{z}) + p^k$ and $z^k \to \hat{z} \in \bar{z} + \delta B$. Then, for sufficiently large k, we have that $-p^k = H(z^k) \in H(\bar{z} + \delta B)$.

In the sequel, we shall need to solve a particular type of *generalized equation*. To be more specific, we shall be interested in solving the problem

$$\text{find } z \in \mathbb{R}^n \text{ s.t. } 0 \in G(z) + \mathcal{N}_C(z), \tag{1.4}$$

where $G : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable, $C \subset \mathbb{R}^n$ is a nonempty convex set, and

$$\mathcal{N}_C(z) = \begin{cases} \{v \in \mathbb{R}^n \mid \langle v, w - z\rangle \leq 0,\ \forall w \in C\} & \text{if } z \in C, \\ \emptyset & \text{if } z \notin C, \end{cases} \tag{1.5}$$

is the normal cone to C at $z \in \mathbb{R}^n$.

When $C = \mathbb{R}^n$, $\mathcal{N}_C = \{0\}$ and we recover the nonlinear equation $G(z) = 0$. A natural extension of the Newton method to the setting of (1.4) is the well-known *Josephy-Newton* algorithm [21]: at the k-th iteration, define z^{k+1} as solution of the problem

$$\text{find } z \in \mathbb{R}^n \text{ s.t. } 0 \in G(z^k) + G'(z^k)(z - z^k) + \mathcal{N}_C(z). \tag{1.6}$$

To begin the discussion of local behavior of the Josephy-Newton method, let us define the multifunctions associated to solution sets of perturbations of (1.4) and of its linearization:

$$\Sigma(p) = \{z \in \mathbb{R}^n \mid 0 \in G(z) + \mathcal{N}_C(z) + p\}, \tag{1.7}$$

and

$$\mathrm{L}\Sigma(p) = \{z \in \mathbb{R}^n \mid 0 \in G(\bar{z}) + G'(\bar{z})(z - \bar{z}) + \mathcal{N}_C(z) + p\}, \tag{1.8}$$

where $\bar{z} \in \mathbb{R}^n$ is a solution of (1.4). Note that $\bar{z} \in \Sigma(0)$ and $\bar{z} \in \mathrm{L}\Sigma(0)$.

As can be seen, the nonsingularity of $G'(\bar{z})$ does not seem to be an appropriate regularity condition in this case. However, when C is polyhedral, the Aubin property of Σ or of $\mathrm{L}\Sigma$ at $(0, \bar{z})$ are still sufficient conditions to guarantee the well-definedness and convergence of the sequence $\{z^k\}$ generated by (1.6).

In order to extend the classical result about Newton method of Theorem 1.1.1, S.M. Robinson [34] shows that a type of Implicit Function Theorem for generalized equations holds if (1.4) is *strongly regular* at $\bar{z}$, i.e., $\mathrm{L}\Sigma$ is locally single-valued and Lipschitz-continuous around the origin. By [9, Theorem 1], if C is polyhedral, the strong regularity of (1.4) at $\bar{z}$ is equivalent to the Aubin property of $\mathrm{L}\Sigma$ at $(0, \bar{z})$, and, using that G is continuously differentiable at $\bar{z}$, it can be proved (for example [27, Theorem 4.69]) that $\mathrm{L}\Sigma$ has the Aubin property at $(0, \bar{z})$ if and only if Σ has the Aubin property at $(0, \bar{z})$.

Before stating the convergence result, we remark that a condition weaker than the Aubin property of Σ at $(0, \bar{z})$ can be used. Following the work of J.F. Bonnans [3], we say that $\bar{z} \in \Sigma(0)$ is *semistable* if there exist $\varepsilon_1, \tau > 0$ such that

$$\Sigma(p) \cap (\bar{z} + \varepsilon_1 B) \subset \bar{z} + \tau \|p\| B. \tag{1.9}$$

Note that for $\gamma = \varepsilon_1/\tau$ we are involving only $p \in \mathbb{R}^n$ such that $\|p\| \leq \gamma$, because if $\|p\| > \gamma$ the inclusion (1.9) holds trivially. As can be seen, the semistability of $\bar{z} \in \Sigma(0)$

is weaker than the Aubin property of Σ at $(0, \bar{z})$, but it is not sufficient to guarantee the result by itself. Let us say that $\bar{z} \in \Sigma(0)$ is *hemistable* if for all $\alpha > 0$ there exists $\varepsilon_2 > 0$ such that if $\|\hat{z} - \bar{z}\| + \|M - G'(\bar{z})\| < \varepsilon_2$ then

$$\{z \in \mathbb{R}^n \mid 0 \in G(\hat{z}) + M(z - \hat{z}) + \mathcal{N}_C(z)\} \cap (\bar{z} + \alpha B) \neq \emptyset.$$

We can now state the following convergence result.

Theorem 1.1.2 *[3, Theorem 2.2] If $\bar{z}$ is a semistable and hemistable solution of* (1.4), *then there exists $\varepsilon > 0$ such that if $\|z^0 - \bar{z}\| \leq \varepsilon$, we have the following:*

1. *At each step k there exists a solution z^{k+1} of* (1.6) *satisfying $\|z^{k+1} - z^k\| \leq 2\varepsilon$.*
2. *The sequence $\{z^k\}$ defined in this way converges superlinearly (quadratically if G' is locally Lipschitz) to $\bar{z}$.*

Also, in the same work, J.F. Bonnans shows that strong regularity at $\bar{z}$ implies both, semistability and hemistability at $\bar{z}$ (see [3, Remark 2.4 (ii)]). In the case of optimization problems, the conditions of strong regularity, semistability and hemistability, have a natural interpretation in terms of constraint qualifications and second-order conditions. This point will be clear in the sequel.

1.2 Constraint Qualifications and Optimality Conditions

One of the problems that we are interested to solve is the classical nonlinear programming problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = 0, \\ & g(x) \leq 0, \end{array} \tag{1.10}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable functions.

A very important tool in the study of optimization problems are the *first-order necessary optimality conditions.* These conditions can be seen as geometrical conditions in a neighborhood of a local minimizer. To this end, let us define the *tangent cone* of a closed set D at a point $x \in D$ as

$$\mathcal{T}_D(x) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{c} \exists \text{ sequences } d^k \to d \text{ and } t_k \to 0^+ \\ \text{with } x + t_k d^k \in D \text{ for all } k \in \mathbb{N} \end{array} \right\}. \tag{1.11}$$

This cone, also called *Bouligand cone* or *contingent cone*, captures the asymptotic geometry of D around x. Now, we can state the conditions mentioned before.

Theorem 1.2.1 *If $\bar{x} \in D$ is a local minimizer of f on D, then*

$$\langle f'(\bar{x}), d \rangle \geq 0 \qquad \forall d \in \mathcal{T}_D(\bar{x}).$$

In general, a vector $\bar{x} \in D$ satisfying this inequality is called a *stationary point* of the problem of minimizing f over the set D.

When D is represented by finitely many differentiable inequalities and equalities:

$$D = \{x \in \mathbb{R}^n \mid h(x) = 0,\ g(x) \leq 0\}, \tag{1.12}$$

there are several conditions on the constraint functions h and g under which $\mathcal{T}_D(x)$ becomes a polyhedral cone. These conditions on the constraints are known as *constraint qualification* (CQ). One of the more general of these CQs is Abadie's CQ, also known as *quasiregularity*, which simply postulates that $\mathcal{T}_D(x)$ is equal to the *linearization cone* of D at x defined as:

$$\mathcal{L}_D(x) = \left\{ d \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \langle h_j'(x), d \rangle = 0 & j = 1, \ldots, l \\ \langle g_i'(x), d \rangle \leq 0 & \forall\, i \in \mathcal{I}(x) \end{array} \right\},$$

where $\mathcal{I}(x)$ is the *active index set* at x, i.e.,

$$\mathcal{I}(x) = \{i \in \{1, \ldots, m\} \mid g_i(x) = 0\}. \tag{1.13}$$

The linearization cone $\mathcal{L}_D(x)$ is also known as the *cone of first-order feasible variations at x*. Of course, Abadie's CQ is not a constructive condition (it is not a condition on the problem data h and g only). This is the reason why other more constructive CQs (see below) are more important, both in theory of optimality conditions and for convergence analysis of computational methods.

One of the advantages of the polyhedrality of $\mathcal{L}_D(x)$ is that, by Farkas' Lemma, we obtain that

$$\langle f'(x), d \rangle \geq 0 \qquad \forall d \in \mathcal{L}_D(x),$$

if and only if, there exist multipliers (or dual variables) $\lambda \in \mathbb{R}^l$ and $\mu_i \geq 0, i \in \mathcal{I}(x)$, such that

$$0 = f'(x) + \sum_{j=1}^{l} \lambda_j h_j'(x) + \sum_{i \in \mathcal{I}(x)} \mu_i g_i'(x).$$

To write this condition in the standard format, let us introduce the *Lagrangian function* associated to problem (1.10), $L : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ such that

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle. \tag{1.14}$$

Hence, we have that

$$L_x'(x, \lambda, \mu) = f'(x) + h'(x)^\top \lambda + g'(x)^\top \mu.$$

We can now write the previous equivalence in this format:

$$x \in D, \qquad \langle f'(x), d\rangle \geq 0 \qquad \forall d \in \mathcal{L}_D(x) \tag{1.15}$$

if and only if $\exists(\lambda, \mu) \in \mathbb{R}^l \times \mathbb{R}^m$ such that

$$\begin{array}{c} 0 = L'_x(x, \lambda, \mu) \\ 0 = h(x) \\ 0 \leq \mu \perp g(x) \leq 0, \end{array} \tag{1.16}$$

where the notation $\perp$ means "perpendicular" (i.e., $\langle \mu, g(x)\rangle = 0$).

The system (1.16) is known as the *Karush-Kuhn-Tucker (KKT) system* of the nonlinear program (1.10). We call a triple (x, λ, μ) satisfying the KKT system a *KKT triple* of (1.10), and x a *KKT point* for (1.10). For any point $x \in \mathbb{R}^n$, the set

$$\mathcal{M}(x) = \{(\lambda, \mu) \in \mathbb{R}^l \times \mathbb{R}^m \mid (x, \lambda, \mu) \text{ is a KKT triple } \}, \tag{1.17}$$

is called the *Lagrange multipliers set* associated to x.

Clearly, the last requirement in the KKT system, i.e., $0 \leq \mu \perp g(x) \leq 0$, is equivalent to the condition

$$0 \leq \mu_i, \quad g_i(x) \leq 0, \quad \mu_i g_i(x) = 0, \quad i = 1, \ldots, m.$$

This condition is called *complementarity condition.* If, in addition, we have that

$$\mu_i > 0, \quad \forall i \in \mathcal{I}(x), \tag{1.18}$$

then we say that *strict complementarity* (SC) condition holds.

Using the multipliers associated to inequality constraints, we shall introduce the following index sets at (x, μ):

$$\mathcal{I}_+(x, \mu) = \{i \in \mathcal{I}(x) \mid \mu_i > 0\}, \quad \mathcal{I}_0(x, \mu) = \mathcal{I}(x) \setminus \mathcal{I}_+(x, \mu), \tag{1.19}$$

known as the sets of *strongly and weakly active constraints*, respectively. Easily, it can be seen that the strict complementarity condition is equivalent to $\mathcal{I}_0(x, \mu) = \emptyset$.

By definition of the tangent cone, it is not hard to prove that $\mathcal{T}_D(x) \subset \mathcal{L}_D(x)$ for any $x \in D$. Thus, to show that $\bar{x}$ is a quasiregular point, i.e., $\mathcal{T}_D(\bar{x}) = \mathcal{L}_D(\bar{x})$, we only need conditions for the constraint functions h and g at $\bar{x}$ under which $\mathcal{L}_D(\bar{x}) \subset \mathcal{T}_D(\bar{x})$. Those conditions are called constraint qualification conditions; the most widely used are the following:

(ACQ) *Affine CQ*:

$$\begin{array}{c} h_j \quad j = 1, \ldots, l \text{ and } g_i \quad \forall\, i \in \mathcal{I}(\bar{x}) \\ \text{are affine in a neighborhood of } \bar{x}. \end{array} \tag{1.20}$$

(LICQ) *Linear Independence CQ*:

$$h'_j(\bar{x}),\ j = 1, \dots, l \quad \text{and} \quad g'_i(\bar{x}),\ i \in \mathcal{I}(\bar{x}), \\ \text{are linearly independent.} \tag{1.21}$$

(SCQ) *Slater's CQ*:

$$h \text{ is affine, } g_i \text{ is convex for all } i \in \mathcal{I}(\bar{x}) \text{ and} \\ \exists\, \hat{x} \in \mathbb{R}^n \text{ s.t. } h(\hat{x}) = 0 \text{ and } g_i(\hat{x}) < 0 \quad \forall i \in \mathcal{I}(\bar{x}). \tag{1.22}$$

(MFCQ) *Mangasarian-Fromovitz CQ*:

$$h'_j(\bar{x}),\ j = 1, \dots, l, \text{ are linearly independent and} \\ \exists\, \hat{d} \in \mathbb{R}^n \text{ s.t. } h'(\bar{x})\hat{d} = 0 \text{ and } \langle g'_i(\bar{x}), \hat{d} \rangle < 0 \quad \forall i \in \mathcal{I}(\bar{x}). \tag{1.23}$$

Any of these conditions (1.20)–(1.23) is sufficient to guarantee that $\mathcal{L}_D(\bar{x}) \subset \mathcal{T}_D(\bar{x})$, and hence $\mathcal{T}_D(\bar{x}) = \mathcal{L}_D(\bar{x})$ follows.

Now, we can formalize those properties with the Karush-Kuhn-Tucker Theorem.

Theorem 1.2.2 *Let f, h and g be differentiable at a local minimizer $\bar{x}$ of problem* (1.10)*, with the derivative of h being continuous at $\bar{x}$. If $\bar{x}$ satisfies any of the conditions* (1.20)–(1.23) *then there exist vectors $\bar{\lambda} \in \mathbb{R}^l$ and $\bar{\mu} \in \mathbb{R}^m$ such that*

$$0 = L'_x(\bar{x}, \bar{\lambda}, \bar{\mu}) \\ 0 = h(\bar{x}) \\ 0 \le \bar{\mu} \perp g(\bar{x}) \le 0. \tag{1.24}$$

It can be proved that under the LICQ (1.21) at $\bar{x}$, the Lagrange multiplier set $\mathcal{M}(\bar{x})$ is a singleton. With more effort, it can be proven that under the MFCQ (1.23) at $\bar{x}$ we can guarantee that the Lagrange multipliers set $\mathcal{M}(\bar{x})$ is compact. There is another CQ, stronger that MFCQ but weaker than LICQ, under which $\mathcal{M}(\bar{x})$ is a singleton; we state this constraint qualification at a pair $(\bar{x}, \bar{\mu})$:

(SMFCQ) *Strict Mangasarian-Fromovitz CQ*:

$$h'_j(\bar{x}),\ j = 1, \dots, l \quad \text{and} \quad g'_i(\bar{x}),\ i \in \mathcal{I}_+(\bar{x}, \bar{\mu}) \\ \text{are linearly independent and } \exists\, \hat{d} \in \mathbb{R}^n \text{ s.t. } h'(\bar{x})\hat{d} = 0, \\ \langle g'_i(\bar{x}), \hat{d} \rangle = 0 \quad \forall i \in \mathcal{I}_+(\bar{x}, \bar{\mu}) \text{ and } \langle g'_i(\bar{x}), \hat{d} \rangle < 0 \quad \forall i \in \mathcal{I}_0(\bar{x}, \bar{\mu}). \tag{1.25}$$

This condition depends, a priori, on a choice of a multiplier $\bar{\mu}$, but this detail turns out to be not important, because it can be shown that $\mathcal{M}(\bar{x}) = \{(\bar{\lambda}, \bar{\mu})\}$ if and only if the SMFCQ (1.25) holds at $(\bar{x}, \bar{\mu})$.

1.2.1 Second-Order Sufficient Conditions

Second-order sufficient conditions are also important properties for the study of local convergence of iterative methods to solve problem (1.10).

Following the approach appearing in [36], the main idea is to find a condition to impose on a KKT triple $(\bar{x}, \bar{\lambda}, \bar{\mu})$ of the problem (1.10) to ensure that $\bar{x}$ will be a strict local minimizer of (1.10). To this end, let us write the KKT system (1.24) in the following form:

$$0 \in \begin{bmatrix} L'_x(\bar{x}, \bar{\lambda}, \bar{\mu}) \\ -h(\bar{x}) \\ -g(\bar{x}) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \mathcal{N}_{\mathbb{R}^m_+}(\bar{\mu}) \end{bmatrix}, \tag{1.26}$$

where

$$\mathcal{N}_{\mathbb{R}^m_+}(\mu) = \begin{cases} \{\nu \in \mathbb{R}^m \mid \nu \leq 0, \langle \nu, \mu \rangle = 0\} \text{ if } \mu \geq 0, \\ \emptyset, \text{ otherwise,} \end{cases}$$

is the normal cone to the nonnegative orthant $\mathbb{R}^m_+$ at $\mu \in \mathbb{R}^m$. Using (1.5), it can be easily seen that

$$\mathcal{N}_{\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+}(x, \lambda, \mu) = \begin{bmatrix} 0 \\ 0 \\ \mathcal{N}_{\mathbb{R}^m_+}(\mu) \end{bmatrix}.$$

As it was shown in Section 1.1, important aspects of the behavior of a generalized equation are captured in its linearization around a given point. To make use of this linearization, let us define

$$G(x, \lambda, \mu) = \begin{bmatrix} L'_x(x, \lambda, \mu) \\ -h(x) \\ -g(x) \end{bmatrix}.$$

Then, the linearized form of (1.26) at $(\bar{x}, \bar{\lambda}, \bar{\mu})$ will be

$$0 \in G(\bar{x}, \bar{\lambda}, \bar{\mu}) + G'(\bar{x}, \bar{\lambda}, \bar{\mu}) \begin{bmatrix} x - \bar{x} \\ \lambda - \bar{\lambda} \\ \mu - \bar{\mu} \end{bmatrix} + \mathcal{N}_{\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+}(x, \lambda, \mu), \tag{1.27}$$

which is a generalized equation with unknowns $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$. Since

$$G'(x, \lambda, \mu) = \begin{bmatrix} L''_{xx}(x, \lambda, \mu) & h'(x)^\top & g'(x)^\top \\ -h'(x) & 0 & 0 \\ -g'(x) & 0 & 0 \end{bmatrix},$$

rewriting (1.27) we obtain

$$0 \in \begin{bmatrix} f'(\bar{x}) + L''_{xx}(\bar{x}, \bar{\lambda}, \bar{\mu})(x - \bar{x}) + h'(\bar{x})^\top \lambda + g'(\bar{x})^\top \mu \\ -h(\bar{x}) - h'(\bar{x})(x - \bar{x}) \\ -g(\bar{x}) - g'(\bar{x})(x - \bar{x}) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \mathcal{N}_{\mathbb{R}^m_+}(\mu) \end{bmatrix}. \tag{1.28}$$

A brief examination of this last generalized equation is sufficient to note that this is the KKT system associated to the problem

$$\begin{array}{rl} \min\limits_{x\in\mathbb{R}^n} & \langle f'(\bar{x}), x-\bar{x}\rangle + \frac{1}{2}\langle L''_{xx}(\bar{x},\bar{\lambda},\bar{\mu})(x-\bar{x}), x-\bar{x}\rangle \\ \text{s.t.} & h(\bar{x}) + h'(\bar{x})(x-\bar{x}) = 0, \\ & g(\bar{x}) + g'(\bar{x})(x-\bar{x}) \leq 0. \end{array} \tag{1.29}$$

Since we are looking for a condition to ensure that $\bar{x}$ will be a strict local minimizer of (1.10), a simple approach to take is to impose some such conditions on (1.29). Thus, let us consider vectors of the form $x = \bar{x} + u$ and find a condition under which $u = 0$ is the unique solution for u in a neighborhood of zero. For $i \notin \mathcal{I}(\bar{x})$, we have that $g_i(\bar{x}) + \langle g_i'(\bar{x}), u\rangle < 0$ for all u small enough. Hence, for u sufficiently small (and taking into account that $h(\bar{x}) = 0$ and $g_i(\bar{x}) = 0, i \in \mathcal{I}(\bar{x})$), the feasible set of (1.29) will be given by vectors u such that $h'(\bar{x})u = 0$ and $\langle g_i'(\bar{x}), u\rangle \leq 0$, $i \in \mathcal{I}(\bar{x})$. Then $u \in \mathcal{L}_D(\bar{x})$. Since $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a KKT triple, by (1.15) we have that $\langle f'(\bar{x}), u\rangle \geq 0$.

If $\langle f'(\bar{x}), u\rangle > 0$, in the objective function

$$\langle f'(\bar{x}), u\rangle + \frac{1}{2}\langle L''_{xx}(\bar{x},\bar{\lambda},\bar{\mu})u, u\rangle,$$

the first-order term will be dominant for u small enough, implying that for $u \neq 0$ the objective will be strictly greater than at $u = 0$. If $\langle f'(\bar{x}), u\rangle = 0$, then the only way to obtain strict increase in the objective function is to resort to the second-order term and to require that $\langle L''_{xx}(\bar{x},\bar{\lambda},\bar{\mu})u, u\rangle > 0$.

We can now use these observations to formulate an appropriate set of conditions.

Suppose that $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$. We say that the *second-order sufficient condition* (SOSC) holds at $(\bar{x}, \bar{\lambda}, \bar{\mu})$ if

$$\langle L''_{xx}(\bar{x},\bar{\lambda},\bar{\mu})u, u\rangle > 0 \qquad \forall\, u \in \mathcal{C}(\bar{x}; D, f')\backslash\{0\}, \tag{1.30}$$

where

$$\mathcal{C}(\bar{x}; D, f') = \mathcal{L}_D(\bar{x}) \cap f'(\bar{x})^{\perp}. \tag{1.31}$$

The set (1.31) is called the *critical cone* at $\bar{x}$ for the problem (1.10). If $\bar{x}$ is a KKT point, then for any $(\lambda, \mu) \in \mathcal{M}(\bar{x})$ the critical cone can also be written as

$$\mathcal{C}(\bar{x}; D, f') = \left\{ u \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} h'(\bar{x})u = 0, & \\ \langle g_i'(\bar{x}), u\rangle = 0 & \forall\, i \in \mathcal{I}_+(\bar{x}, \mu), \\ \langle g_i'(\bar{x}), u\rangle \leq 0 & \forall\, i \in \mathcal{I}_0(\bar{x}, \mu). \end{array} \right\}. \tag{1.32}$$

There is also another, and stronger, form of (1.30) that is often used in the literature to obtain good behavior (in some sense) of the KKT triple $(\bar{x}, \bar{\lambda}, \bar{\mu})$. This strengthened form is called *strong second-order sufficient condition* (SSOSC), and it states that

$$\langle L''_{xx}(\bar{x},\bar{\lambda},\bar{\mu})u, u\rangle > 0 \qquad \forall\, u \in \mathcal{C}^+(\bar{x}, \bar{\mu})\backslash\{0\}, \tag{1.33}$$

where

$$\mathcal{C}^+(\bar{x},\bar{\mu}) = \{u \in \mathbb{R}^n \mid h'(\bar{x})u = 0,\ \langle g_i'(\bar{x}), u\rangle = 0 \quad \forall\, i \in \mathcal{I}_+(\bar{x},\bar{\mu})\}\,.$$

Clearly, (1.33) is stronger than (1.30), because

$$\mathcal{C}(\bar{x}; D, f') \subset \mathcal{C}^+(\bar{x},\bar{\mu}).$$

It is also important to note that the cone $\mathcal{C}^+(\bar{x},\bar{\mu})$ in SSOSC is a subspace, while the cone $\mathcal{C}(\bar{x}; D, f')$ in SOSC is not, in general. The subspace structure simplifies many issues. For this reason, proving convergence of an algorithm assuming SOSC is almost always significantly more challenging than when SSOSC is assumed (when the two conditions are not the same, of course. Note that they are the same, for example, when there are no inequality constraints or when strict complementarity holds).

1.3 Sequential Quadratic Programming

The *Sequential quadratic programming* (SQP) algorithm is one of the most efficient general-purpose methods for solving nonlinear optimization problems.

To relate this method with the general Newton schemes discussed above, we shall begin with the equality constrained nonlinear program

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & h(x) = 0. \end{array} \tag{1.34}$$

To solve this problem we shall generate a primal-dual sequence such that at the k-th iteration we define (x^{k+1}, λ^{k+1}) where λ^{k+1} is a multiplier associated to a local minimum (or a stationary point) x^{k+1} of

$$\begin{array}{rl} \min\limits_{y\in\mathbb{R}^n} & \langle f'(x^k), y - x^k\rangle + \frac{1}{2}\langle L''_{xx}(x^k,\lambda^k)(y-x^k), y - x^k\rangle \\ \text{s.t.} & h(x^k) + h'(x^k)(y - x^k) = 0. \end{array} \tag{1.35}$$

Note that the ACQ (1.20) guarantees the existence of a multiplier.

If $\bar{x}$ is a local minimizer of (1.34) with an associated multiplier $\bar{\lambda}$, then

$$0 = G(\bar{x},\bar{\lambda}),$$

where

$$G(x,\lambda) = \left[\begin{array}{c} L'_x(x,\lambda) \\ -h(x) \end{array}\right].$$

Since the KKT system associated to (1.35) is

$$\begin{array}{l} 0 = f'(x^k) + L''_{xx}(x^k,\lambda^k)(x^{k+1} - x^k) + h'(x^k)^\top \lambda^{k+1}, \\ 0 = h(x^k) + h'(x^k)(x^{k+1} - x^k), \end{array}$$

taking $w = (x, \lambda)$, this system can be seen to be equivalent to

$$0 = G(w^k) + G'(w^k)(w^{k+1} - w^k).$$

Hence, the sequence generated by solving subproblems (1.35) is the same as the one generated by Newton method (1.2) applied to find a zero of G.

In order to use the convergence result of Theorem 1.1.1, we need the nonsingularity of $G'(\bar{w})$. By simple calculations it can be seen that a sufficient condition for this nonsingularity property hold is that both LICQ (1.21) and SOSC (1.30) hold at $(\bar{x}, \bar{\lambda})$. To summarize, let us state the local convergence result.

Theorem 1.3.1 *Let f and h be twice continuously differentiable in a neighborhood of a solution $\bar{x}$ of* (1.34). *Suppose that LICQ* (1.21) *holds at $\bar{x}$ and that for the unique associated multiplier $\bar{\lambda}$ SOSC* (1.30) *holds at $(\bar{x}, \bar{\lambda})$. Then, there exists a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\lambda})$ such that if $(x^0, \lambda^0) \in \mathcal{V}$ then the sequence $\{(x^k, \lambda^k)\}$ generated by* (1.35) *is well-defined and converges superlinearly to $(\bar{x}, \bar{\lambda})$. Furthermore, if f'' and h'' are Lipschitz-continuous in a neighborhood of $\bar{x}$ then the convergence rate is quadratic.*

For future reference, note that under the assumptions of Theorem 1.3.1, the primal-dual solution is locally unique.

When the problem to solve is (1.10), i.e.,

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & h(x) = 0, \\ & g(x) \leq 0, \end{array}$$

the SQP method generates a sequence $\{(x^k, \lambda^k, \mu^k)\}$ such that for each k, $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is the nearest vector to (x^k, λ^k, μ^k) where $(\lambda^{k+1}, \mu^{k+1})$ are multipliers associated to a local minimum (or stationary point) x^{k+1} of

$$\begin{array}{rl} \min\limits_{y \in \mathbb{R}^n} & \langle f'(x^k), y - x^k \rangle + \frac{1}{2} \langle L''_{xx}(x^k, \lambda^k)(y - x^k), y - x^k \rangle \\ \text{s.t.} & h(x^k) + h'(x^k)(y - x^k) = 0, \\ & g(x^k) + g'(x^k)(y - x^k) \leq 0. \end{array} \tag{1.36}$$

Note that, in contrast with the equality constrained problem, here subproblems may have multiple local solutions under natural assumptions, even in a neighborhood of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ (see [4, Example 13.1]). This is the reason for the "localization" requirement that $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is nearest to (x^k, λ^k, μ^k) among possibly multiple solutions. In practice, this requirement is simply ignored, of course. But as a matter of convergence analysis, such a requirement is absolutely necessary in order to eliminate possible stationary points that are far from the region of the analysis.

Let us consider $\bar{x}$, a local minimum of (1.10), with associated multipliers $(\bar{\lambda}, \bar{\mu})$. Thus, as was noted in (1.26), for $(x, \lambda, \mu) = (\bar{x}, \bar{\lambda}, \bar{\mu})$ we have

$$0 \in G(x, \lambda, \mu) + \mathcal{N}_{\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+}(x, \lambda, \mu), \tag{1.37}$$

where

$$G(x, \lambda, \mu) = \begin{bmatrix} L'_x(x, \lambda, \mu) \\ -h(x) \\ -g(x) \end{bmatrix}.$$

The KKT system associated to the subproblem (1.36) is

$$\begin{array}{l} 0 = f'(x^k) + L''_{xx}(x^k, \lambda^k, \mu^k)(x^{k+1} - x^k) + h'(x^k)^\top \lambda^{k+1} + g'(x^k)^\top \mu^{k+1}, \\ 0 = h(x^k) + h'(x^k)(x^{k+1} - x^k), \\ 0 \leq \mu^{k+1} \perp g(x^k) + g'(x^k)(x^{k+1} - x^k) \leq 0. \end{array}$$

Thus, taking $w = (x, \lambda, \mu)$, this system is equivalent to

$$0 \in G(w^k) + G'(w^k)(w^{k+1} - w^k) + \mathcal{N}_{\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+}(w^{k+1}). \tag{1.38}$$

Hence, the sequence generated by solving subproblems (1.36) is the same as the one generated by the Josephy-Newton method (1.6) with $C = \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m_+$ and G defined above.

In [34, Section 4], S.M. Robinson shows that under the LICQ (1.21) at $\bar{x}$, if the unique $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ satisfies SOSC (1.30) and SC (1.18), then the generalized equation (1.37) is strongly regular at $(\bar{x}, \bar{\lambda}, \bar{\mu})$. He also proved that the strong regularity of (1.37) at $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is guaranteed under LICQ (1.21) together with the SSOSC (1.33). Hence, we can obtain the classical convergence result of SQP method from Theorem 1.1.2.

Theorem 1.3.2 *Let f, h and g be twice continuously differentiable in a neighborhood of a solution $\bar{x}$ of* (1.10)*. Suppose that LICQ* (1.21) *holds at $\bar{x}$ and that for the unique associated multiplier $(\bar{\lambda}, \bar{\mu})$, SSOSC* (1.33)*, or SOSC* (1.30) *and SC* (1.18) *hold. Then there exists a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ such that if $(x^0, \lambda^0, \mu^0) \in \mathcal{V}$, then the sequence $\{(x^k, \lambda^k, \mu^k)\}$, generated by* (1.36)*, is well-defined and converges superlinearly to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Furthermore, if f'', h'' and g'' are Lipschitz-continuous in a neighborhood of $\bar{x}$ then the convergence rate is quadratic.*

As LICQ (1.21) implies SMFCQ (1.25) and SSOSC (1.33) implies SOSC (1.30), we can conclude this section with the more general convergence result, due to J.F. Bonnans. Note that this result also follows from convergence properties of the Josephy-Newton method, because SMFCQ and SOSC imply semistability and hemistability of the associated generalized equation [3].

Theorem 1.3.3 *[3, Theorem 6.1] Let f, h and g be twice continuously differentiable in a neighborhood of a solution $\bar{x}$ of* (1.10)*. Suppose that SMFCQ* (1.25) *holds at $\bar{x}$ and that for the unique associated multiplier $(\bar{\lambda}, \bar{\mu})$ the SOSC* (1.30) *holds. Then there exists a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\lambda}, \bar{\mu})$ such that if $(x^0, \lambda^0, \mu^0) \in \mathcal{V}$ then the sequence $\{(x^k, \lambda^k, \mu^k)\}$, generated by* (1.36)*, is well-defined and converges superlinearly to $(\bar{x}, \bar{\lambda}, \bar{\mu})$. Furthermore, if f'', h'' and g'' are Lipschitz-continuous in a neighborhood of $\bar{x}$ then the convergence rate is quadratic.*

It is important to stress that Theorem 1.3.3 is the sharpest known result on local convergence of SQP, and that it subsumes that the primal-dual solution is isolated (in particular, uniqueness of the multiplier).

1.4 Stabilized Sequential Quadratic Programming

As has been discussed in the previous section, superlinear convergence of the SQP method can be guaranteed only when at a local solution $\bar{x}$ of (1.10) the Lagrange multipliers set $\mathcal{M}(\bar{x})$ is a singleton. Dealing with situations where the latter is not the case, is considered one of the challenges in contemporary optimization. S.J Wright [41] proposed *stabilized Sequential Quadratic Programming* (sSQP) algorithm to tackle problems with nonunique multipliers.

In order to simplify the notation, from now on we shall deal with the inequality constrained problem, which is also consistent with the literature on the subject. We also note that since our eventual convergence result does not assume anything about the constraints (in particular, no CQs), considering inequality constraints only certainly does not lead to any loss of generality (we could just write equalities as inequalities with opposite signs, as violation of LICQ by such a formulation is of no concern for us).

Let the problem be

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & g(x) \leq 0. \end{array} \tag{1.39}$$

As discussed above, the classical SQP method generates a sequence $\{(x^k, \mu^k)\}$ such that for each k, (x^{k+1}, μ^{k+1}) is the nearest vector to (x^k, μ^k) where μ^{k+1} is a multiplier associated to a local minimum (or stationary point) x^{k+1} of

$$\begin{array}{ll} \displaystyle\min_{y \in \mathbb{R}^n} & \langle f'(x^k), y - x^k \rangle + \frac{1}{2}\langle L''_{xx}(x^k, \mu^k)(y - x^k), y - x^k \rangle \\ \text{s.t.} & g(x^k) + g'(x^k)(y - x^k) \leq 0. \end{array} \tag{1.40}$$

To understand the stabilization procedure, note that we can derive the subproblem (1.40) from the following min-max problem:

$$\begin{array}{ll} \displaystyle\min_{y \in \mathbb{R}^n} \max_{\nu \in \mathbb{R}^m_+} & \langle f'(x^k), y - x^k \rangle + \frac{1}{2}\langle L''_{xx}(x^k, \mu^k)(y - x^k), y - x^k \rangle \\ & + \langle \nu, g(x^k) + g'(x^k)(y - x^k) \rangle. \end{array}$$

The idea is to control the behavior of the dual sequence $\{\mu^k\}$ adding a proximal penalty term. Hence, (x^{k+1}, μ^{k+1}) will be a local minimax for the problem

$$\begin{array}{ll} \displaystyle\min_{y \in \mathbb{R}^n} \max_{\nu \in \mathbb{R}^m_+} & \langle f'(x^k), y - x^k \rangle + \frac{1}{2}\langle L''_{xx}(x^k, \mu^k)(y - x^k), y - x^k \rangle \\ & + \langle \nu, g(x^k) + g'(x^k)(y - x^k) \rangle - \frac{\sigma(x^k, \mu^k)}{2}\|\nu - \mu^k\|^2, \end{array}$$

where the dual stabilization parameter $\sigma(x^k, \mu^k) > 0$ is some computable quantity measuring the violation of optimality conditions for (1.39) at (x^k, μ^k). After some calculations, it can be seen that this minimax problem is equivalent to solving the following quadratic programming subproblem:

$$\begin{array}{ll} \min\limits_{(y,\nu)\in\mathbb{R}^n\times\mathbb{R}^m} & \langle f'(x^k), y - x^k\rangle + \frac{1}{2}\langle L''_{xx}(x^k, \mu^k)(y - x^k), y - x^k\rangle + \frac{\sigma(x^k,\mu^k)}{2}\|\nu\|^2 \\ \text{s.t.} & g(x^k) + g'(x^k)(y - x^k) - \sigma(x^k, \mu^k)(\nu - \mu^k) \leq 0. \end{array} \tag{1.41}$$

One advantage of sSQP is easy to see from the mere formulation of the subproblem: unlike in SQP, subproblem (1.41) is always feasible (just take $y = x^k$ and ν large enough). In the case of SQP subproblems, a constraint qualification is needed (at least, MFCQ) to guarantee feasibility.

Concerning local behavior of this method, in [41] superlinear convergence of sSQP has been established under the MFCQ (1.23), SOSC (1.30) for all $\bar{\mu} \in \mathcal{M}(\bar{x})$, and the assumption that the initial dual iterate μ^0 is close enough to a multiplier $\bar{\mu}$ satisfying the strict complementarity condition (1.18). In [42], superlinear convergence of sSQP has been shown without strict complementarity, under MFCQ (1.23) and the SSOSC (1.33). In [43], the assumption of strict complementarity has also been removed from the results of [41], thus showing superlinear convergence under MFCQ (1.23) and SOSC (1.30) for all $\bar{\mu} \in \mathcal{M}(\bar{x})$. If MFCQ is not assumed, then superlinear convergence can be shown under the assumption of SSOSC (1.33) for *some* $\bar{\mu} \in \mathcal{M}(\bar{x})$, provided that μ^0 is close enough to such $\bar{\mu}$ [15]; see also [13].

For completeness, we proceed to formally state the result where no constraint qualification is assumed.

Theorem 1.4.1 *[15, Theorem 1] Let f and g be twice Lipschitz-continuously differentiable functions in a neighborhood of a local minimizer $\bar{x}$ of* (1.39)*, and suppose that there exists $\bar{\mu} \in \mathcal{M}(\bar{x})$ such that $(\bar{x}, \bar{\mu})$ satisfies the SSOSC* (1.33)*.*

Then for any choice of the constant ρ_0 sufficiently large, there exist constants ρ_1, δ and β with the property that $\rho_0\delta \leq \rho_1$ and for each starting guess $(x^0, \mu^0) \in (\bar{x}, \bar{\mu}) + \delta B$, there are iterates (x^k, μ^k) contained in $(\bar{x}, \bar{\mu}) + \delta B$, where (x^k, μ^k) is the unique solution of (1.41) *and $\sigma_k = \sigma(x^k, \mu^k)$ is any scalar that satisfies the condition*

$$\rho_0\|x^k - \bar{x}\| \leq \sigma_k \leq \rho_1.$$

Moreover, the following estimate holds:

$$\|x^{k+1} - \bar{x}\| + \|\mu^{k+1} - \hat{\mu}^{k+1}\| \leq \beta\left(\|x^k - \bar{x}\|^2 + \|\mu^k - \hat{\mu}^k\|^2 + \sigma_k\|\mu^k - \hat{\mu}^k\|\right),$$

where $\hat{\mu}^{k+1}$ and $\hat{\mu}^k$ are the closest elements of $\mathcal{M}(\bar{x})$ to μ^{k+1} and μ^k, respectively.

In Chapter 2, we prove a new superlinear convergence result stronger than Theorem 1.4.1 (or any of the other results cited above), assuming only SOSC (1.30) and not assuming any constraint qualification. As an additional bonus, our result is obtained in the setting of variational problems, for which we introduce a natural extension of sSQP.

1.5 Globalization of the Sequential Quadratic Programming method

There are at least three classes of techniques to globalize a local algorithm: line-search, trust-region and filter. Here, we consider only the line-search approach, mostly to motivate SQCQP, a modification of SQP that avoids the so-called Maratos effect. To this end, let us define the *penalty function* for (1.10)

$$\Upsilon_\beta(x) = f(x) + \beta\rho(x), \tag{1.42}$$

where $\beta > 0$ and $\rho : \mathbb{R}^n \to \mathbb{R}_+$ is zero on the feasible set of (1.10) and positive outside of it. Examples of this kind of penalty functions are the ℓ_1 and ℓ_∞ penalty functions, i.e.,

$$\rho(x) = \sum_{j=1}^{l} |h_j(x)| + \sum_{i=1}^{m} \max\{0, g_i(x)\}$$

and

$$\rho(x) = \max\{|h_1(x)|, \dots, |h_l(x)|, g_1(x), \dots, g_m(x)\},$$

respectively.

At the k-th iteration, the globalized SQP will find the KKT triple $(\tilde{x}^{k+1}, \tilde{\lambda}^{k+1}, \tilde{\mu}^{k+1})$ of subproblem (1.36) as before, and shall define

$$(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) = \left(x^k + t_k(\tilde{x}^{k+1} - x^k), \tilde{\lambda}^{k+1}, \tilde{\mu}^{k+1}\right),$$

where $t_k \in (0, 1]$ is obtained by a line-search procedure in the primal space that guarantees a sufficient decrease of the function

$$t \to \Upsilon_{\beta_k}\left(x^k + t(\tilde{x}^{k+1} - x^k)\right) - \Upsilon_{\beta_k}(x^k).$$

For a line-search procedure and update of the penalty parameter β_k, see, e.g., [4, 15.1].

As can be seen, when $t_k \neq 1$, the generated sequence differs from that in Section 1.3. Hence, we can only guarantee superlinear convergence of the sequence if $t_k = 1$ for all k sufficiently large. Unfortunately, there are problems satisfying all natural assumptions (see [29, Example 18.1]), where $t_k = 1$ is rejected by any penalty function of the form (1.42), even though this step would have made good progress toward a solution. This undesirable phenomenon is called the *Maratos effect* [25].

One of the techniques for avoiding the Maratos effect is using a *second-order correction*, that we proceed to describe. In the literature, it seems to be always introduced for the equality constrained case. We shall merely follow the tradition.

The following argument is informal. It serves as an intuition for introducing corrective second order terms to avoid the Maratos effect. Consider $y \in \mathbb{R}^n$ such that

$$0 = h_j(x^k) + \langle h_j'(x^k), y - x^k\rangle + \frac{1}{2}\langle h_j''(x^k)(y - x^k), y - x^k\rangle \qquad j = 1, \dots, l. \tag{1.43}$$

Neglecting third-order terms we can suppose that for $j = 1, \dots, l$ we have

$$h_j(\tilde{x}^{k+1}) = h_j(x^k) + \langle h_j'(x^k), \tilde{x}^{k+1} - x^k \rangle + \frac{1}{2} \langle h_j''(x^k)(\tilde{x}^{k+1} - x^k), \tilde{x}^{k+1} - x^k \rangle.$$

Also, for y near $\tilde{x}^{k+1}$ we can suppose that

$$\langle h_j''(x^k)(y - x^k), y - x^k \rangle = \langle h_j''(x^k)(\tilde{x}^{k+1} - x^k), \tilde{x}^{k+1} - x^k \rangle.$$

Then, using these relations we conclude that

$$\begin{aligned} 0 &= h_j(x^k) + \langle h_j'(x^k), y - x^k \rangle + \frac{1}{2} \langle h_j''(x^k)(y - x^k), y - x^k \rangle \\ &= h_j(\tilde{x}^{k+1}) - \langle h_j'(x^k), \tilde{x}^{k+1} - x^k \rangle + \langle h_j'(x^k), y - x^k \rangle \\ &= h_j(\tilde{x}^{k+1}) + \langle h_j'(x^k), y - \tilde{x}^{k+1} \rangle. \end{aligned}$$

Hence, the second order correction will consist in defining y^k as the vector $y \in \mathbb{R}^n$ nearest to $\tilde{x}^{k+1}$ that satisfies

$$0 = h(\tilde{x}^{k+1}) + h'(x^k)(y - \tilde{x}^{k+1}),$$

and finding $t_k \in (0, 1]$ that guarantees a sufficient decrease of the function

$$t \to \Upsilon_{\beta_k} \left(x^k + t(\tilde{x}^{k+1} - x^k) + t^2(y^k - \tilde{x}^{k+1}) \right) - \Upsilon_{\beta_k}(x^k).$$

Obviously, after this change, we have to set $x^{k+1} = x^k + t_k(\tilde{x}^{k+1} - x^k) + t_k^2(y^k - \tilde{x}^{k+1})$.

Under some hypothesis, it can be proven that, with this correction, the value $t_k = 1$ will be accepted for all k sufficiently large (see [4, Proposition 15.7]). Note that when $t_k = 1$, we obtain that $x^{k+1} = y^k$ and it is a vector satisfying (1.43), a quadratic approximation of the original constraints. This gives an idea of a modification of the SQP method that will be considered in the next section.

1.6 Sequential Quadratically Constrained Quadratic Programming

The *Sequential Quadratically Constrained Quadratic Programming* (SQCQP) method solves at each iteration a subproblem that involves quadratic constraints and a quadratic objective function. For simplicity, we shall deal with the inequality constrained problem (1.39). To solve

$$\begin{array}{rl} \min & f(x) \\ \text{s.t.} & g(x) \le 0, \end{array}$$

the SQCQP method generates a primal-dual sequence $\{(x^k, \mu^k)\}$ such that for each k, μ^{k+1} is a multiplier associated to x^{k+1}, the nearest to x^k local minimum (or stationary point) of

$$\begin{aligned} \min_{y \in \mathbb{R}^n} \quad & \langle f'(x^k), y - x^k \rangle + \frac{1}{2} \langle f''(x^k)(y - x^k), y - x^k \rangle \\ \text{s.t.} \quad & g_i(x^k) + \langle g_i'(x^k), y - x^k \rangle + \frac{1}{2} \langle g_i''(x^k)(y - x^k), y - x^k \rangle \leq 0 \quad i = 1, \ldots, m. \end{aligned} \tag{1.44}$$

As some previous work on SQCQP and related methods, we mention [31, 32, 40, 22, 1, 14, 38]. In the convex case, subproblem (1.44) can be cast as a second-order cone program [24, 28], which can be solved efficiently by interior-point algorithms (such as [26, 39]). In [1], nonconvex subproblems (1.44) were also handled quite efficiently by using other nonlinear programming techniques. Even though quadratically constrained subproblems are computationally more difficult than linearly constrained ones (as in the more traditional SQP methods), they are manageable by modern computational tools and the extra effort in solving them can be worth it. I.e., at least in some situations, one may expect that fewer subproblems will need to be solved, when compared to SQP. Some numerical validation of this observations can be found in computational experiments of [1].

In order to guarantee global convergence, SQCQP methods require some modifications to subproblem (1.44), as well as a linesearch procedure for an adequately chosen penalty function. (See, for example, [14, 38]). But under certain assumptions, locally, all those modifications reduce precisely to (1.44). Moreover, the unit stepsize satisfies the linesearch criteria under very mild conditions [38, Proposition 8] (in particular, no second-order sufficiency is needed), which is one of the attractive features of SQCQP. Thus, what is relevant for local convergence analysis is precisely the method given by (1.44). Note that, as a consequence of acceptance of the unit stepsize, the Maratos effect [25, 33] does not occur in SQCQP.

As for the local convergence results, in [1] local primal superlinear rate of convergence of a trust-region SQCQP method is obtained under the MFCQ (1.23) and a certain quadratic growth condition. We note that, under MFCQ, quadratic growth is equivalent to the SOSC (1.30), see [5, Theorem 3.70]. Quadratic convergence of the primal-dual sequence is obtained in [14]. The assumptions in [14] are as follows: convexity of f and of g, the Slater condition (1.22) (equivalent to MFCQ in the convex case) and a condition stronger than the SSOSC (1.33) (implying quadratic growth). This set of assumptions is stronger than those in [1], but the assertions in the two papers are different and not comparable, because neither of the two results implies the other one.

We remark, for future reference, that quadratic convergence of $\{(x^k, \mu^k)\}$ to $(\bar{x}, \bar{\mu})$ does not imply even superlinear or linear convergence of $\{x^k\}$ to $\bar{x}$. For example (see [4, Exercise 12.8]), let $\mu^1 \in (0, 1)$ and consider the sequence $\{(x^k, \mu^k)\}_{k \geq 1} \subset \mathbb{R}^2$ generated by

$$\mu^{k+1} = (\mu^k)^2 \quad \text{and} \quad x^{k+1} = \begin{cases} x^k & \text{if } k \text{ is odd,} \\ (\mu^{k+1})^2 & \text{if } k \text{ is even.} \end{cases}$$

Hence, if k is odd we obtain that $(x^{k+1}, \mu^{k+1}) = \big(x^k, (\mu^k)^2\big) = \big((\mu^k)^2, (\mu^k)^2\big)$ and $(x^k, \mu^k) = \big((\mu^k)^2, \mu^k\big)$. If k is even we have $x^{k-1} = (\mu^{k-1})^2 = \mu^k$, then $(x^{k+1}, \mu^{k+1}) = \big((\mu^{k+1})^2, \mu^{k+1}\big) = \big((\mu^k)^4, (\mu^k)^2\big)$ and $(x^k, \mu^k) = (x^{k-1}, \mu^k) = (\mu^k, \mu^k)$.
Thus $(x^k, \mu^k) \to (0, 0)$ and

$$\limsup_{k\to\infty} \frac{\|(x^{k+1}, \mu^{k+1})\|}{\|(x^k, \mu^k)\|^2} = \sqrt{2}, \qquad \limsup_{k\to\infty} \frac{|x^{k+1}|}{|x^k|} = 1.$$

In particular, we have quadratic convergence of $\{(x^k, \mu^k)\}$, but not even linear convergence of $\{x^k\}$.

Thus, if one is interested in primal convergence rate, it must be analyzed separately, independently of any primal-dual analysis.

In Chapter 3, we show a new local convergence result which complements [14] and [1] (neither stronger nor weaker). We prove primal-dual quadratic convergence under LICQ (1.21), SOSC (1.30) and SC (1.18). Additionally, we provide a necessary and sufficient condition for superlinear convergence of the primal sequence. Also, our results hold in the more general variational setting, for which we introduce an extension of SQCQP.

1.7 Variational Problems

Let us consider the following variational problem:

$$\text{find } \bar{x} \in D \text{ s.t. } \langle F(\bar{x}), y - \bar{x} \rangle \geq 0 \quad \forall\, y \in \bar{x} + \mathcal{T}_D(\bar{x}), \tag{1.45}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable, D is given by (1.12) and $\mathcal{T}_D(\bar{x})$ is the tangent cone defined by (1.11).

When D is convex, (1.45) is equivalent to the classical variational inequality [10]:

$$\text{find } \bar{x} \in D \text{ s.t. } \langle F(\bar{x}), x - \bar{x} \rangle \geq 0 \quad \forall x \in D. \tag{1.46}$$

It is good to emphasize that the variational inequality (1.46) makes no sense without the convexity assumption on D, despite sometimes appearing in the literature without this assumption (It is enough to think of minimizing $f(x) = x^2$, $x \in \mathbb{R}$, on the set $D = \{-2; 1; 2\}$, and observe that the (global) minimizer $\bar{x} = 1$ does not satisfy (1.46) for $F(x) = f'(x)$). The sensible formulation for a variational problem over a nonconvex set D is (1.45).

Replacing D by a closed convex cone $K \subset \mathbb{R}^n$, (1.45) is equivalent to

$$\text{find } \bar{x} \in \mathbb{R}^n \text{ s.t. } K^* \ni F(\bar{x}) \perp \bar{x} \in K,$$

where $K^* = \{u \in \mathbb{R}^n \mid \langle u, v\rangle \geq 0 \quad \forall v \in K\}$, which is known as the (generalized) *complementarity problem*. The particular case when $K = \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}_+$ with $n_1 + n_2 = n$ is known as the *mixed complementarity problem* (MCP).

Note that when in (1.45)

$$F(x) = f'(x), \quad x \in \mathbb{R}^n, \tag{1.47}$$

we obtain the first-order necessary optimality conditions described in Theorem 1.2.1. Also, when D is given by (1.12), the concepts of Section 1.2 can be extended to the variational context, as discussed next.

Let us introduce the vector function $\Psi : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n$ such that

$$\Psi(x, \lambda, \mu) = F(x) + h'(x)^\top \lambda + g'(x)^\top \mu. \tag{1.48}$$

Since (1.15) is equivalent to (1.16), we have that if $\bar{x}$ is a quasiregular point (i.e., $\mathcal{T}_D(\bar{x}) = \mathcal{L}_D(\bar{x})$), then $\bar{x}$ is a solution of (1.45) if and only if there exist vectors $\bar{\lambda} \in \mathbb{R}^l$ and $\bar{\mu} \in \mathbb{R}^m$ such that

$$\begin{gathered} 0 = \Psi(\bar{x}, \bar{\lambda}, \bar{\mu}), \\ 0 = h(\bar{x}), \\ 0 \leq \bar{\mu} \perp g(\bar{x}) \leq 0. \end{gathered} \tag{1.49}$$

We shall refer to (1.49) as the KKT system of (1.45). In this context, $(\bar{x}, \bar{\lambda}, \bar{\mu})$ will be a KKT triple of (1.45) and the Lagrange multipliers set $\mathcal{M}$, will be defined as in (1.17).

We say that the *second-order condition* (SOC) holds at $(\bar{x}, \bar{\lambda}, \bar{\mu})$ if

$$\langle \Psi'_x(\bar{x}, \bar{\lambda}, \bar{\mu})u, u\rangle \neq 0 \qquad \forall\, u \in \mathcal{C}(\bar{x}; D, F)\backslash\{0\}, \tag{1.50}$$

where

$$\mathcal{C}(\bar{x}; D, F) = \mathcal{L}_D(\bar{x}) \cap F(\bar{x})^\perp.$$

If $\bar{x}$ is a KKT point, then $\mathcal{C}(\bar{x}; D, F)$ will coincide with the expression (1.32), introduced for the optimization case, for any $(\lambda, \mu) \in \mathcal{M}(\bar{x})$. Note that since the cone $\mathcal{C}(\bar{x}; D, F)$ is convex, (1.50) means that the quadratic form has the same nonzero sign for all $u \in \mathcal{C}(\bar{x}; D, F) \setminus \{0\}$. Sometimes, we shall distinguish the two cases, using SOC$^+$ if

$$\langle \Psi'_x(\bar{x}, \bar{\lambda}, \bar{\mu})u, u\rangle > 0 \qquad \forall\, u \in \mathcal{C}(\bar{x}; D, F)\backslash\{0\}, \tag{1.51}$$

and SOC$^-$ if

$$\langle \Psi'_x(\bar{x}, \bar{\lambda}, \bar{\mu})u, u\rangle < 0 \qquad \forall\, u \in \mathcal{C}(\bar{x}; D, F)\backslash\{0\}. \tag{1.52}$$

As can be seen, in the case of the optimization problem corresponding to (1.47), we have that

$$\Psi(x, \lambda, \mu) = L'_x(x, \lambda, \mu),$$

and hence SOC$^+$ (1.51) reduces to SOSC (1.30).

Chapter 2

Stabilized Newton-Type Method for Variational Problems without Constraint Qualifications

This chapter corresponds to the material from paper [12].

The stabilized version of the sequential quadratic programming algorithm (sSQP) had been developed in order to achieve fast convergence despite possible degeneracy of constraints of optimization problems, when the Lagrange multipliers associated to a solution are not unique. Superlinear convergence of sSQP has been previously established under the second-order sufficient condition for optimality (1.30) and the Mangasarian-Fromovitz constraint qualification (1.23), or under the strong second-order sufficient condition for optimality (1.33) (in that case, without constraint qualification assumptions). We prove a superlinear convergence result stronger than the above, assuming SOSC (1.30) only. In addition, our analysis is carried out in the more general setting of variational problems, for which we introduce a natural extension of sSQP techniques. In the process, we also obtain a new error bound for Karush-Kuhn-Tucker systems for variational problems.

2.1 Introduction

Given smooth mappings $F : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^m$, we consider the following variational inequality (VI) problem:

$$\text{find } x \in D \quad \text{s.t.} \quad \langle F(x), y - x\rangle \geq 0 \quad \forall\, y \in x + \mathcal{T}_D(x), \tag{2.1}$$

where

$$D = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0,\ i = 1, \dots, m\},$$

and $\mathcal{T}_D(x)$ is the (standard) tangent cone to the set D at the point $x \in D$. Throughout this chapter, we assume that F is once and g is twice continuously differentiable.

When for some smooth function $f : \mathbb{R}^n \to \mathbb{R}$ it holds that

$$F(x) = f'(x), \quad x \in \mathbb{R}^n, \tag{2.2}$$

then (2.1) describes (primal) first-order necessary optimality conditions for the optimization problem

$$\min\ f(x) \quad \text{s.t.} \quad x \in D. \tag{2.3}$$

To motivate the development consider, for the moment, the optimization problem (2.3). Iterations of the fundamental sequential quadratic programming method (SQP, e.g., [2]) for (2.3) consist of solving subproblems of the form

$$\begin{array}{ll} \min\limits_{y \in \mathbb{R}^n} & \langle f'(x^k), y - x^k \rangle + \frac{1}{2} \langle L''_{xx}(x^k, \mu^k)(y - x^k), y - x^k \rangle \\ \text{s.t.} & g(x^k) + g'(x^k)(y - x^k) \leq 0, \end{array}$$

where

$$L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, \quad L(x, \mu) = f(x) + \langle \mu, g(x) \rangle,$$

is the Lagrangian of (2.3), and $(x^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^m_+$ is the current primal-dual iterate. Let $\bar{x} \in \mathbb{R}^n$ be a local solution of (2.3), and let $\mathcal{M}(\bar{x})$ be the set of Lagrange multipliers associated to $\bar{x}$. The minimal conditions [3] which guarantee that the SQP method outlined above is locally well-defined and superlinearly convergent are the existence and uniqueness of the Lagrange multiplier $\bar{\mu}$ associated to $\bar{x}$ (also known as the strict Mangasarian-Fromovitz constraint qualification) and the second-order sufficient condition (SOSC)

$$\langle L''_{xx}(\bar{x}, \bar{\mu})d,\ d \rangle > 0 \quad \forall\, d \in \mathcal{C}(\bar{x}; D, f') \setminus \{0\}, \tag{2.4}$$

where

$$\begin{aligned} \mathcal{C}(\bar{x}; D, f') &= \{d \in \mathbb{R}^n \mid \langle f'(\bar{x}), d \rangle = 0,\ \langle g_i'(\bar{x}), d \rangle \leq 0 \quad \forall i \in \mathcal{I}(\bar{x})\} \\ &= \{d \in \mathbb{R}^n \mid \langle g_i'(\bar{x}), d \rangle = 0 \quad \forall i \in \mathcal{I}_+(\bar{x}, \bar{\mu}),\ \langle g_i'(\bar{x}), d \rangle \leq 0 \quad \forall i \in \mathcal{I}_0(\bar{x}, \bar{\mu})\}, \end{aligned} \tag{2.5}$$

is the critical cone of (2.3) at $\bar{x}$, with

$$\mathcal{I} = \mathcal{I}(\bar{x}) = \{i \in \{1, \ldots, m\} \mid g_i(\bar{x}) = 0\}$$

the set of constraints active at $\bar{x}$, and

$$\mathcal{I}_+(\bar{x}, \bar{\mu}) = \{i \in \mathcal{I}(\bar{x}) \mid \bar{\mu}_i > 0\}, \quad \mathcal{I}_0(\bar{x}, \bar{\mu}) = \mathcal{I}(\bar{x}) \setminus \mathcal{I}_+(\bar{x}, \bar{\mu}),$$

being the set of strongly and weakly active constraints, respectively.

We emphasize that convergence of SQP requires certain regularity of constraints (specifically, the strict Mangasarian-Fromovitz constraint qualification).

To deal with the case when constraint qualifications may be violated (and multipliers associated to the primal solution of the optimization problem (2.3) may not be unique), a stabilized version of SQP (sSQP) has been introduced in [41]. This method can be stated [23] in the form of solving subproblems

$$\begin{array}{cl} \min\limits_{(y,\lambda)\in\mathbb{R}^n\times\mathbb{R}^m} & \langle f'(x^k), y-x^k\rangle + \frac{1}{2}\langle L''_{xx}(x^k,\mu^k)(y-x^k), y-x^k\rangle + \frac{\sigma(x^k,\mu^k)}{2}\|\lambda\|^2 \\ \text{s.t.} & g(x^k)+g'(x^k)(y-x^k)-\sigma(x^k,\mu^k)(\lambda-\mu^k)\leq 0, \end{array} \tag{2.6}$$

where $(x^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^m_+$ is again the current primal-dual iterate, while the dual stabilization parameter $\sigma(x^k, \mu^k) > 0$ is some computable quantity measuring violation of optimality conditions for (2.3) at the point (x^k, μ^k). As is easy to see, unlike in SQP, the subproblems (2.6) are always feasible, regardless of constraint qualification. In [41], superlinear convergence of sSQP has been established under the Mangasarian-Fromovitz constraint qualification (MFCQ, which is equivalent to nonemptiness and compactness of the multiplier set $\mathcal{M}(\bar{x})$), SOSC (2.4) for all $\bar{\mu} \in \mathcal{M}(\bar{x})$, and the assumption that the initial dual iterate μ^0 is close enough to a multiplier such that $\bar{\mu}_{\mathcal{I}(\bar{x})} > 0$ (in particular, strict complementarity is assumed). In [42], superlinear convergence of sSQP has been shown without strict complementarity, under MFCQ and the strong second-order sufficient condition (SSOSC)

$$\langle L''_{xx}(\bar{x},\bar{\mu})d,\, d\rangle > 0 \qquad \forall\, d \in \mathcal{C}^+(\bar{x},\bar{\mu}) \setminus \{0\}, \tag{2.7}$$

assumed for all $\bar{\mu} \in \mathcal{M}(\bar{x})$, where

$$\mathcal{C}^+(\bar{x},\bar{\mu}) = \{d \in \mathbb{R}^n \mid \langle g_i'(\bar{x}), d\rangle = 0 \quad \forall i \in \mathcal{I}_+(\bar{x},\bar{\mu})\}.$$

In [43], the assumption of strict complementarity has also been removed from the results of [41], thus showing superlinear convergence under MFCQ and SOSC (2.4) for all $\bar{\mu} \in \mathcal{M}(\bar{x})$. If MFCQ is not assumed, then superlinear convergence can be shown under the assumption of SSOSC (2.7) for *some* $\bar{\mu} \in \mathcal{M}(\bar{x})$, provided that μ^0 is close enough to such $\bar{\mu}$ [15]; see also [13]. In fact, it was posed as an open question in [13, p. 117] whether or not some condition weaker than SSOSC could be used to prove sSQP convergence when no constraints qualifications are assumed. In this chapter, we answer this question in an affirmative manner. We show that if the starting point is close to $(\bar{x}, \bar{\mu})$ satisfying SOSC (2.4), then the sSQP method is well-defined and converges superlinearly. Moreover, our development is carried out for the variational setting, in which sSQP for optimization is a special case.

Let us now go back to the variational problem (2.1). In this context, a natural extension of sSQP is the following iterative procedure, obtained from the variational formulation of optimality conditions for (2.6). To this end, define

$$\Psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n, \quad \Psi(x,\mu) = F(x) + g'(x)^\top \mu.$$

Let $(x^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^m_+$ be the current primal-dual approximation to a solution of (2.1), and define

$$\Phi_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m, \quad \Phi_k(y,\lambda) = \begin{bmatrix} F(x^k) + \Psi'_x(x^k,\mu^k)(y-x^k) \\ \sigma(x^k,\mu^k)\lambda \end{bmatrix},$$

and

$$\Delta_k = \{(y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m \mid g(x^k) + g(x^k)(y - x^k) - \sigma(x^k, \mu^k)(\lambda - \mu^k) \leq 0\},$$

where $\sigma(x^k, \mu^k) > 0$ is the dual stabilization parameter.

Consider *affine* variational subproblems of the form

$$\text{find } (y, \lambda) \in \Delta_k \quad \text{s.t.} \quad \langle \Phi_k(y, \lambda), (z, \nu) - (y, \lambda) \rangle \geq 0 \qquad \forall\, (z, \nu) \in \Delta_k. \tag{2.8}$$

As it can be easily seen, in the optimization case (2.2), the variational subproblem (2.8) is precisely the first-order (primal) necessary optimality condition for the sSQP subproblem (2.6). Thus, our framework contains sSQP for optimization as a special case. Note that the framework makes good sense also in the variational setting, where solving the fully nonlinear problem (2.1) is replaced by solving a sequence of fully affine subproblems (2.8) (the mapping Φ_k is affine and the set Δ_k is polyhedral). As in sSQP, the feasible set in (2.8) is always nonempty. We shall prove that under a suitable second-order condition, the method outlined above is locally well-defined and converges superlinearly to a solution of the Karush-Kuhn-Tucker (KKT) system for (2.1), which is

$$\begin{array}{c} 0 = \Psi(x, \mu) = F(x) + g'(x)^\top \mu, \\ 0 \leq \mu \perp g(x) \leq 0, \end{array} \tag{2.9}$$

where $\mu \perp g(x)$ means that $\langle \mu, g(x) \rangle = 0$. We make the standing assumption that the KKT system (2.9) has a primal-dual solution (in fact, if the constraints are degenerate, there are many dual solutions associated to the same primal solution). The setting of assuming existence of multipliers, while not assuming any specific constraint qualification conditions ensuring such existence, is common when dealing with degenerate problems, e.g., [42, 13, 18, 44, 20, 19].

The rest of the chapter is organized as follows. In Section 2.2, we recall the general iterative framework of Fischer [13] that will be used to prove superlinear convergence of our algorithm. We note that in [13], the general framework has been applied to the method of proximally-regularized linearizations of monotone mixed complementarity problems (MCP), and to sSQP for KKT systems arising from optimization. Compared to the first item, our iterations are different (regularization is in the dual space only), and we do not assume any monotonicity or convexity. Compared to the second item, we cover KKT systems that include variational problems, and prove superlinear convergence under SOSC instead of SSOSC employed in [13]. In Section 2.3, we prove that subproblems (2.8) are locally solvable if $\sigma(\cdot)$ provides a local error bound [30] on the distance to the solution set of the KKT system (2.9). In Section 2.4, among other things, we derive a suitable error bound. The results of Sections 2.3 and 2.4 show that the assumptions of [13], stated in Section 2.2, are verified, which implies superlinear convergence of the method given by (2.8).

2.2 Fischer's general iterative framework for problems with nonisolated solutions

Let $G : \mathbb{R}^q \to \mathbb{R}^l$ be a continuous map, $\mathcal{N}$ be a closed set-valued map from $\mathbb{R}^q$ to the subsets of $\mathbb{R}^l$, and consider the generalized equation (GE)

$$\text{find } w \in \mathbb{R}^q \quad \text{s.t.} \quad 0 \in G(w) + \mathcal{N}(w). \tag{2.10}$$

Denote by Σ_* the (nonempty) solution set of (2.10).

Consider a class of methods that, given $s \in \mathbb{R}^q$, generate the next iterate by solving a subproblem of the form

$$\text{find } w \in \mathbb{R}^q \quad \text{s.t.} \quad 0 \in \mathcal{A}(w, s) + \mathcal{N}(w), \tag{2.11}$$

where $\mathcal{A}(\cdot, s)$ is an approximation of $G(\cdot)$ around s. Denote by

$$Z(s) = \{w \in \mathbb{R}^q \mid 0 \in \mathcal{A}(w, s) + \mathcal{N}(w)\}$$

the solution set of (2.11). In local convergence analysis it is standard to assume that the distance between two consecutive iterates is not too large (without very strong assumptions, subproblems (2.11) may have other solutions that are far away from a given solution of (2.10) that is being approximated; those solutions are irrelevant for local analysis and should be excluded). To this end, define

$$Z_c(s) = \{w \in Z(s) \mid \|w - s\| \leq c \operatorname{dist}(s, \Sigma_*)\},$$

where $c \in [1, +\infty)$ is arbitrary but fixed, and consider the iterative scheme

$$w^{k+1} \in Z_c(w^k),\ k = 0, 1, \ldots \quad , w^0 \in \mathbb{R}^q. \tag{2.12}$$

The following holds (see [13, Theorem 1]).

Theorem 2.2.1 *Let Σ_* be the (nonempty) solution set of* (2.10) *and let $\Sigma_0 \neq \emptyset$ be a closed subset of Σ_*. Suppose that*

1. **(Upper Lipschitz-continuity of the solution set of GE)**
   There exist numbers $\varepsilon_1, \gamma, \ell > 0$ such that, with $Q = \Sigma_0 + \varepsilon_1 B$, it holds that

   $$\Sigma(p) \cap Q \subseteq \Sigma_* + \ell\|p\|B \quad \forall p \in \gamma B,$$

   and

   $$\Sigma(p) = \{w \in \mathbb{R}^q \mid 0 \in G(w) + \mathcal{N}(w) + p\}.$$

2. **(Precision of approximation of $G(\cdot)$ by $\mathcal{A}(\cdot, s)$)**
   There exists $\varepsilon_2 > 0$ such that

   $$\sup\{\|R(w, s)\| \mid w \in s + c\, dist(s, \Sigma_*)B\} \leq o\,(dist(s, \Sigma_*)) \quad \forall s \in \Sigma_0 + \varepsilon_2 B,$$

   where $R(w, s) = G(w) - \mathcal{A}(w, s)$.

3. **(Solvability of subproblems)**
There exists $\varepsilon_3 > 0$ *such that* $Z_c(s) \neq \emptyset$ *for all* $s \in \Sigma_0 + \varepsilon_3 B$.

Then there exists $\varepsilon > 0$ *such that for any* $w^0 \in \Sigma_0 + \varepsilon B$, *the iterates defined by* (2.12) *are well defined and converge superlinearly to some* $w^* \in \Sigma_*$. *Furthermore, the convergence is of order* β *if the function* $o(\cdot)$ *in Item 2 satisfies*

$$o(t) \leq c_0 t^\beta \quad \forall t \in [0, 1],$$

for some $c_0 > 0$ *and* $\beta > 1$ *(in particular, convergence is quadratic if* $\beta = 2$*).*

Let the vectors $l \in \bar{\mathbb{R}}^p$ (lower bounds) and $u \in \bar{\mathbb{R}}^p$ (upper bounds) be given, where $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. Moreover, assume that $l_i < u_i, \quad i = 1, \ldots, p$. Define $C = \{w \in \mathbb{R}^q \mid l_i \leq w_i \leq u_i, \quad i = 1, \ldots, p\}$ and denote by $\mathcal{N}_C(w)$ the normal cone to C at $w \in \mathbb{R}^q$. In the sequel, we shall also make use of the following error bound result.

Theorem 2.2.2 *[13, Theorem 2] Let* $G : \mathbb{R}^q \to \mathbb{R}^q$ *be a locally Lipschitz-continuous function. If* Σ_0 *is bounded and Assumption 1 in Theorem 2.2.1 is satisfied with* $\mathcal{N} = \mathcal{N}_C$, *then there are* $\beta_0 > 0$ *and* $\varepsilon_0 > 0$ *such that*

$$\|s - \Pi_C(s - G(s))\| \geq \beta_0 \mathit{dist}(s, \Sigma_*) \qquad \forall s \in \Sigma_0 + \varepsilon_0 B.$$

To relate the proposed iterative scheme (2.8) to the framework above, define

$$G(x, \mu) = \begin{bmatrix} \Psi(x, \mu) \\ -g(x) \end{bmatrix}, \quad \mathcal{N}(x, \mu) = \begin{bmatrix} 0 \\ \mathcal{N}_{\mathbb{R}^m_+}(\mu) \end{bmatrix}. \tag{2.13}$$

Let $w = (x, \mu) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^q$. Then the the KKT system (2.9) for problem (2.1) is equivalent to solving the generalized equation (2.10) with G and $\mathcal{N}$ given by (2.13).

Since subproblem (2.8) of our method is an affine VI, it is equivalent to solving the KKT system of finding $(y, \lambda, \nu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{aligned} &0 = F(x^k) + \Psi'_x(x^k, \mu^k)(y - x^k) + g'(x^k)^\top \nu, \\ &0 = \sigma(x^k, \mu^k)\lambda - \sigma(x^k, \mu^k)\nu, \\ &0 \leq \nu \perp [g(x^k) + g'(x^k)(y - x^k) - \sigma(x^k, \mu^k)(\lambda - \mu^k)] \leq 0. \end{aligned}$$

Noting that $\lambda = \nu$, by the second relation, the system above is then equivalent to finding $(y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{aligned} 0 &= F(x^k) + \Psi'_x(x^k, \mu^k)(y - x^k) + g'(x^k)^\top \lambda \\ &= \Psi(x^k, \mu^k) + \Psi'_x(x^k, \mu^k)(y - x^k) + g'(x^k)^\top(\lambda - \mu^k), \\ 0 &\leq \lambda \perp [g(x^k) + g'(x^k)(y - x^k) - \sigma(x^k, \mu^k)(\lambda - \mu^k)] \leq 0. \end{aligned} \tag{2.14}$$

Letting now $w = (x, \mu) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^q$, $s \in \mathbb{R}^q$,

$$\mathcal{A}(w, s) = G(s) + \Big(G'(s) + \Lambda(s)\Big)(w - s), \quad \Lambda(s) = \begin{bmatrix} 0 & 0 \\ 0 & \sigma(s)I \end{bmatrix},$$

where G is defined in (2.13), we obtain that solving (2.14) (and thus (2.8)) is equivalent to solving GE subproblems of the form (2.11).

The rest of the work proves that problem (2.10) and subproblem (2.11), corresponding to problem (2.9) and subproblem (2.14), respectively, satisfy the assumptions in Theorem 2.2.1. The hard part is to prove, under a (weak) second-order condition only, the upper Lipschitz-continuity of the solution set of the KKT system (2.9) and, specially, solvability of subproblems (2.14) (Assumptions 1 and 3 of Theorem 2.2.1).

Assumption 2 is easily seen to be satisfied, because

$$\begin{aligned}\|R(w,s)\| &= \|G(w) - \mathcal{A}(w,s)\| \\ &= \|G(w) - G(s) - (G'(s) + \Lambda(s))(w-s)\| \\ &\le \left\| \int_0^1 [G'(s+t(w-s)) - G'(s)](w-s)dt \right\| + \|\Lambda(s)(w-s)\| \\ &\le \left(\int_0^1 \|G'(s+t(w-s)) - G'(s)\|dt + \sigma(s) \right) \|w-s\|,\end{aligned}$$

which implies that

$$\|R(w,s)\| \le o\left(\operatorname{dist}(s,\Sigma_*)\right)$$

when $w \in s + c \operatorname{dist}(s,\Sigma_*)B$ and

$$\sigma(s) \le L_1 \operatorname{dist}(s,\Sigma_*)$$

for some $L_1 > 0$. The latter inequality holds for any reasonable choice of $\sigma(\cdot)$, for example a residual $\sigma(\cdot)$ of the KKT system (by the Lipschitz-continuity); this will be made evident in Section 2.4.

Note also that if, in addition, the derivatives F' and g'' are Lipschitz-continuous, then so is G', and we have that

$$\|R(w,s)\| \le L_2 \operatorname{dist}(s,\Sigma_*)^2. \tag{2.15}$$

2.3 Solvability of subproblems

We now prove that KKT subproblems of the form (2.14) (which are equivalent to affine variational subproblems (2.8)) are locally solvable if a certain second-order condition holds, and if the dual regularization parameters $\sigma(x^k, \mu^k)$ are of the order of the distance to the solution set of the KKT system (2.9) for problem (2.1). A specific computable way of choosing such parameters will be discussed in Section 2.4.

Let $\bar{x}$ be a solution of VI (2.1), and let

$$\mathcal{M}(\bar{x}) = \{\mu \in \mathbb{R}^m \mid (\bar{x}, \mu) \text{ solves } (2.9)\}$$

be the associated (nonempty) set of Lagrange multipliers. Let the sets of active, strongly active and weakly active constraints ($\mathcal{I} = \mathcal{I}(\bar{x})$, $\mathcal{I}_+(\bar{x},\mu)$ and $\mathcal{I}_0(\bar{x},\mu)$, respectively) be defined as in Section 2.1.

We say that $(\bar{x}, \bar{\mu})$, with $\bar{\mu} \in \mathcal{M}(\bar{x})$, satisfies the positive second-order condition (SOC$^+$) for the KKT system (2.9) if

$$\langle \Psi_x'(\bar{x}, \bar{\mu})u, u\rangle > 0 \qquad \forall\, u \in \mathcal{C}(\bar{x}; D, F)\backslash\{0\}, \tag{2.16}$$

where

$$\begin{aligned}\mathcal{C}(\bar{x}; D, F) &= \{u \in \mathbb{R}^n \mid \langle F(\bar{x}), u\rangle = 0,\ \langle g_i'(\bar{x}), u\rangle \le 0 \quad \forall\, i \in \mathcal{I}(\bar{x})\} \\ &= \{u \in \mathbb{R}^n \mid \langle g_i'(\bar{x}), u\rangle = 0 \quad \forall\, i \in \mathcal{I}_+(\bar{x}, \mu),\ \langle g_i'(\bar{x}), u\rangle \le 0 \quad \forall\, i \in \mathcal{I}_0(\bar{x}, \mu)\}\,.\end{aligned} \tag{2.17}$$

(As it is well known, the second equality above does not depend on the choice of $\mu \in \mathcal{M}(\bar{x})$). In the case of the optimization problem (2.3), $\mathcal{C}(\bar{x}; D, F)$ is the standard critical cone (2.5) at $\bar{x}$, and (2.16) is the standard second-order condition (2.4) which is sufficient for optimality of the point $\bar{x}$.

As already mentioned, we assume also that the function $\sigma(\cdot)$ satisfies the error bound property. As Lemma 2.4.3 in Section 2.4 shows that under SOC$^+$ (2.16) the primal part $\bar{x}$ of the solution is locally unique, we can write our error bound in the following form: there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and constants $\beta_2 \ge \beta_1 > 0$ such that for all $(x, \mu) \in \mathcal{V}$ it holds that

$$\beta_1\Big(\|x - \bar{x}\| + \mathrm{dist}(\mu, \mathcal{M}(\bar{x}))\Big) \le \sigma(x, \mu) \le \beta_2\Big(\|x - \bar{x}\| + \mathrm{dist}(\mu, \mathcal{M}(\bar{x}))\Big). \tag{2.18}$$

More details on a computable choice of $\sigma(\cdot)$ will be given in Section 2.4.

We start with extending SOC$^+$ (2.16) from the copositivity property of the matrix in a primal cone to uniform positivity, in a neighborhood of the point $(\bar{x}, \bar{\mu})$, of a certain function in a certain parametric primal-dual cone.

Proposition 2.3.1 *Suppose that SOC$^+$* (2.16) *holds at $(\bar{x}, \bar{\mu})$ and that σ satisfies the second inequality in* (2.18). *Then there exist a constant $\gamma_1 > 0$ and a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that for all $(x, \mu) \in \mathcal{V}$ it holds that*

$$\langle \Psi_x'(x, \mu)u, u\rangle + \sigma(x, \mu)\|v\|^2 \ge \gamma_1\Big(\|u\|^2 + \sigma(x, \mu)\|v\|^2\Big) \quad \forall (u, v) \in K(x, \mu), \tag{2.19}$$

where

$$K(x, \mu) = \left\{(u, v) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \;\middle|\; \begin{array}{ll} \langle g_i'(x), u\rangle = \sigma(x, \mu)v_i & \forall i \in \mathcal{I}_+(\bar{x}, \bar{\mu}) \\ \langle g_i'(x), u\rangle \le \sigma(x, \mu)v_i & \forall i \in \mathcal{I}_0(\bar{x}, \bar{\mu}) \end{array}\right\}. \tag{2.20}$$

Proof. Suppose the contrary, i.e., that there exist $(x^k, \mu^k) \to (\bar{x}, \bar{\mu})$ and $(u^k, v^k) \in K(x^k, \mu^k)$ such that

$$\langle \Psi_x'(x^k, \mu^k)u^k, u^k\rangle + \sigma_k\|v^k\|^2 < \frac{1}{k}(\|u^k\|^2 + \sigma_k\|v^k\|^2), \tag{2.21}$$

where $\sigma_k = \sigma(x^k, \mu^k)$. Evidently, (2.21) subsumes $(u^k, v^k) \neq 0$.
Let $\eta_k = \|(u^k, \sqrt{\sigma_k} v^k)\| > 0$. Passing onto a subsequence, if necessary, we can assume that

$$\frac{1}{\eta_k} \begin{bmatrix} u^k \\ \sqrt{\sigma_k} v^k \end{bmatrix} \to \begin{bmatrix} \bar{u} \\ \bar{w} \end{bmatrix} \neq 0. \tag{2.22}$$

Observe that since $\sigma_k \to 0$ by the second inequality in (2.18), while $\sqrt{\sigma_k} v^k / \eta_k$ is bounded, it holds that

$$\sigma_k \frac{v^k}{\eta_k} = \sqrt{\sigma_k} \frac{\sqrt{\sigma_k} v^k}{\eta_k} \to 0. \tag{2.23}$$

Since $K(x^k, \mu^k)$ is a cone, we have that $(u^k/\eta_k, v^k/\eta_k) \in K(x^k, \mu^k)$. Dividing now the relations in (2.20) by η_k and taking limits, considering (2.23), we obtain that

$$\langle g_i'(\bar{x}), \bar{u} \rangle = 0 \quad \forall\, i \in \mathcal{I}_+(\bar{x}, \bar{\mu}),\ \langle g_i'(\bar{x}), \bar{u} \rangle \leq 0 \quad \forall\, i \in \mathcal{I}_0(\bar{x}, \bar{\mu}),$$

i.e., $\bar{u} \in \mathcal{C}(\bar{x}; D, F)$.

On the other hand, dividing (2.21) by η_k^2 and taking limits we have that

$$\langle \Psi_x'(\bar{x}, \bar{\mu}) \bar{u}, \bar{u} \rangle + \|\bar{w}\|^2 \leq 0. \tag{2.24}$$

This shows that $\langle \Psi_x'(\bar{x}, \bar{\mu}) \bar{u}, \bar{u} \rangle \leq 0$ for $\bar{u} \in \mathcal{C}(\bar{x}; D, F)$. Hence, $\bar{u} = 0$. Now from (2.24) we have that $\bar{w} = 0$ also, in contradiction with (2.22). ∎

Corollary 2.3.2 *Suppose that SOC*$^+$ (2.16) *holds at* $(\bar{x}, \bar{\mu})$ *and that* σ *satisfies the right inequality in* (2.18)*. Then there exists a neighborhood* $\mathcal{V}$ *of* $(\bar{x}, \bar{\mu})$ *such that the matrix*

$$\begin{bmatrix} \Psi_x'(x, \mu) & g_{\mathcal{I}}'(x)^\top \\ -g_{\mathcal{I}}'(x) & \sigma(x, \mu) I \end{bmatrix} \tag{2.25}$$

is nonsingular for all $(x, \mu) \in \mathcal{V}$ *such that* $\sigma(x, \mu) > 0$.

Proof. By Proposition 2.3.1, there exists a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that (2.19) holds. Let $(x, \mu) \in \mathcal{V}$, $\sigma(x, \mu) > 0$, and suppose that (u, v) is a vector in the kernel of the matrix (2.25), i.e.,

$$0 = \Psi_x'(x, \mu) u + g_{\mathcal{I}}'(x)^\top v, \tag{2.26}$$

$$0 = -g_{\mathcal{I}}'(x) u + \sigma(x, \mu) v. \tag{2.27}$$

By (2.27) we have that $\langle g_i'(x), u \rangle = \sigma(x, \mu) v_i$ for all $i \in \mathcal{I}$. This shows that $(u, v) \in K(x, \mu)$ defined in (2.20). Also, multiplying (2.27) by v we have

$$\langle g_{\mathcal{I}}'(x) u, v \rangle = \sigma(x, \mu) \|v\|^2.$$

Multiplying by u both sides in (2.26), we then obtain that

$$0 = \langle \Psi_x'(x, \mu) u, u \rangle + \langle g_{\mathcal{I}}'(x)^\top v, u \rangle = \langle \Psi_x'(x, \mu) u, u \rangle + \sigma(x, \mu) \|v\|^2.$$

Then, by (2.19), we have that $0 \geq \gamma_1(\|u\|^2 + \sigma(x,\mu)\|v\|^2)$. Hence, $u = 0$ and $v = 0$, implying that the matrix in (2.25) is nonsingular. ■

Our proof of existence of solutions of subproblems is done in two steps. We start by showing that a certain part of KKT subproblem (2.14) has a solution. We shall make use of the existence result in [10, Theorem 2.5.10]. More specifically, we shall need a consequence of [10, Theorem 2.5.10], which we state as follows.

Theorem 2.3.3 *Let K be a closed convex cone in $\mathbb{R}^l$ and $M \in \mathbb{R}^{l\times l}$. Suppose that $d = 0$ is the unique solution of the generalized complementarity problem*

$$K \ni d \perp Md \in K^*, \tag{2.28}$$

and that M is copositive on K.

Then for all $q \in \mathbb{R}^l$, the generalized complementarity problem of finding $d \in \mathbb{R}^l$ such that

$$K \ni d \perp Md + q \in K^*$$

has a nonempty compact solution set.

Clearly, if M is strictly copositive on K, then (2.28) has the origin as the unique solution, and all the assumptions of Theorem 2.3.3 hold.

Proposition 2.3.4 *Suppose that SOC$^+$ (2.16) holds at $(\bar{x}, \bar{\mu})$ and that σ satisfies the second inequality in (2.18). Then there exists a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that for all $(x, \mu) \in \mathcal{V}$ with $\sigma(x,\mu) > 0$, the mixed complementarity problem of finding $(y, \lambda_{\mathcal{I}}) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|}$ such that*

$$\begin{aligned} &0 = F(x) + \Psi'_x(x,\mu)(y - x) + g'_{\mathcal{I}}(x)^\top \lambda_{\mathcal{I}}, \\ &0 = g_i(x) + \langle g'_i(x), y - x\rangle - \sigma(x,\mu)(\lambda_i - \mu_i),\ i \in \mathcal{I}_+(\bar{x},\bar{\mu}), \\ &0 \leq \lambda_i \perp [g_i(x) + \langle g'_i(x), y - x\rangle - \sigma(x,\mu)(\lambda_i - \mu_i)] \leq 0,\ i \in \mathcal{I}_0(\bar{x},\bar{\mu}), \end{aligned} \tag{2.29}$$

has a nonempty compact solution set.

Proof. Define

$$M = \begin{bmatrix} \Psi'_x(x,\mu) & 0 \\ 0 & \sigma(x,\mu)I \end{bmatrix}, \quad q = \begin{bmatrix} F(x) - \Psi'_x(x,\mu)x \\ 0 \end{bmatrix},$$

$$b_i = g_i(x) - \langle g'_i(x), x\rangle + \sigma(x,\mu)\mu_i \quad \forall i \in \mathcal{I},$$

and the $|\mathcal{I}| \times (n + |\mathcal{I}|)$ matrix A with rows given by transposing

$$a^i = \begin{bmatrix} g'_i(x) \\ -\sigma(x,\mu)e^i \end{bmatrix},$$

where $e^i \in \mathbb{R}^{|\mathcal{I}|}$ is the i-th vector of the canonical basis. With this notation, it can be seen that (2.29) is equivalent to solving the following affine VI:

$$\text{find } \bar{z} \in Q \text{ s.t. } \langle M\bar{z} + q, z - \bar{z}\rangle \geq 0 \qquad \forall z \in Q, \tag{2.30}$$

where

$$Q = \{z \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \mid A_{\mathcal{I}_+} z + b_{\mathcal{I}_+} = 0, A_{\mathcal{I}_0} z + b_{\mathcal{I}_0} \leq 0\},$$

$\mathcal{I}_+ = \mathcal{I}_+(\bar{x}, \bar{\mu})$, $\mathcal{I}_0 = \mathcal{I}_0(\bar{x}, \bar{\mu})$.

Let $(\tilde{u}, \tilde{v}_{\mathcal{I}})$ be the unique solution of the linear system

$$\begin{bmatrix} \Psi'_x(x,\mu) & g'_{\mathcal{I}}(x)^\top \\ -g'_{\mathcal{I}}(x) & \sigma(x,\mu)I \end{bmatrix} \begin{bmatrix} u \\ v_{\mathcal{I}} \end{bmatrix} = \begin{bmatrix} -F(x) - g'_{\mathcal{I}}(x)^\top \mu_{\mathcal{I}} \\ g_{\mathcal{I}}(x) \end{bmatrix},$$

which exists due to Corollary 2.3.2. Define $\tilde{z} = (x + \tilde{u}, \mu_{\mathcal{I}} + \tilde{v}_{\mathcal{I}})$. For each $i \in \mathcal{I}$ we then have that

$$\begin{aligned} \langle a^i, \tilde{z}\rangle &= \langle g'_i(x), x\rangle - \sigma(x,\mu)\mu_i + \langle g'_i(x), \tilde{u}\rangle - \sigma(x,\mu)\tilde{v}_i \\ &= \langle g'_i(x), x\rangle - \sigma(x,\mu)\mu_i - g_i(x) \\ &= -b_i. \end{aligned}$$

In particular, $\tilde{z} \in Q$ and all the constraints defining the polyhedral set Q are active at $\tilde{z}$. Note that, in the adopted notation, the cone $K = K(x,\mu)$ defined in (2.20) can be written as

$$K = \{d \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \mid A_{\mathcal{I}_+} d = 0, A_{\mathcal{I}_0} d \leq 0\}.$$

Hence,

$$Q = \{z \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \mid A_{\mathcal{I}_+}(z - \tilde{z}) = 0, A_{\mathcal{I}_0}(z - \tilde{z}) \leq 0\} = \tilde{z} + K.$$

We can then write (2.30) in the following form:

$$\text{find } \bar{d} \in K \text{ s.t. } \langle M\bar{d} + M\tilde{z} + q, d - \bar{d}\rangle \geq 0 \qquad \forall d \in K,$$

which is the generalized complementarity problem

$$K \ni \bar{d} \perp M\bar{d} + M\tilde{z} + q \in K^*. \tag{2.31}$$

By Proposition 2.3.1, there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that (2.19) holds for all $(x,\mu) \in \mathcal{V}$. This shows that if $\sigma(x,\mu) > 0$, then M is strictly copositive on the cone K. Now Theorem 2.3.3 implies that (2.31) has a nonempty compact solution set.
■

We next show that the step given by solving the system (2.29), which is part of our subproblem (2.14), satisfies the localization property appearing in the iterative framework of Section 2.2.

Proposition 2.3.5 *Suppose that SOC$^+$ (2.16) holds at $(\bar{x}, \bar{\mu})$ and that σ satisfies both inequalities in (2.18). Then there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and a constant $\gamma_3 > 0$ such that for all $(x, \mu) \in \mathcal{V} \cap (\mathbb{R}^n \times \mathbb{R}^m_+)$ with $\sigma(x, \mu) > 0$, it holds that*

$$\left\| \begin{bmatrix} y - x \\ \lambda_{\mathcal{I}} - \mu_{\mathcal{I}} \end{bmatrix} \right\| \leq \gamma_3 \sigma(x, \mu),$$

where $(y, \lambda_{\mathcal{I}})$ is any solution of (2.29).

Proof. For contradiction purposes, suppose there exists a sequence $\{(x^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m_+$ such that

$$(x^k, \mu^k) \to (\bar{x}, \bar{\mu}) \text{ and } \eta_k = \left\| \begin{bmatrix} y^k - x^k \\ \lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}} \end{bmatrix} \right\| > k\sigma_k,$$

where $\sigma_k = \sigma(x^k, \mu^k) > 0$ and $(y^k, \lambda^k_{\mathcal{I}})$ satisfies

$$0 = F(x^k) + \Psi'_x(x^k, \mu^k)(y^k - x^k) + g'_{\mathcal{I}}(x^k)^\top \lambda^k_{\mathcal{I}}, \tag{2.32}$$
$$0 = g_i(x^k) + \langle g'_i(x^k), y^k - x^k \rangle - \sigma_k(\lambda^k_i - \mu^k_i), \ i \in \mathcal{I}_+(\bar{x}, \bar{\mu}), \tag{2.33}$$
$$0 \leq \lambda^k_i \perp [g_i(x^k) + \langle g'_i(x^k), y^k - x^k \rangle - \sigma_k(\lambda^k_i - \mu^k_i)] \leq 0, \ i \in \mathcal{I}_0(\bar{x}, \bar{\mu}). \tag{2.34}$$

By the assumption above,

$$\frac{\sigma_k}{\eta_k} < \frac{1}{k} \to 0. \tag{2.35}$$

Note first that, by (2.18), it holds

$$\|g_{\mathcal{I}}(x^k)\| = \|g_{\mathcal{I}}(x^k) - g_{\mathcal{I}}(\bar{x})\| \leq c_1\|x^k - \bar{x}\| \leq c_2\sigma_k, \tag{2.36}$$

$$\|g'_{\mathcal{I}}(x^k) - g'_{\mathcal{I}}(\bar{x})\| \leq c_3\|x^k - \bar{x}\| \leq c_4\sigma_k. \tag{2.37}$$

Denote $\hat{\mu}^k = \Pi_{\mathcal{M}(\bar{x})}(\mu^k)$. Since $\hat{\mu}^k_i = 0$ for $i \notin \mathcal{I}$, we have that

$$\begin{aligned} \|F(x^k) + g'_{\mathcal{I}}(x^k)^\top \mu^k_{\mathcal{I}}\| &= \|F(x^k) + g'_{\mathcal{I}}(x^k)^\top \mu^k_{\mathcal{I}} - F(\bar{x}) - g'_{\mathcal{I}}(\bar{x})^\top \hat{\mu}^k_{\mathcal{I}}\| \\ &\leq c_5 \left(\|x^k - \bar{x}\| + \|\mu^k_{\mathcal{I}} - \hat{\mu}^k_{\mathcal{I}}\|\right) \\ &\leq c_6\sigma_k, \end{aligned} \tag{2.38}$$

where the first inequality follows from the Lipschitz-continuity of the functions involved, while the last one from (2.18).

Taking a subsequence, if necessary, we can assume that

$$\frac{1}{\eta_k} \begin{bmatrix} y^k - x^k \\ \lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}} \end{bmatrix} \to \begin{bmatrix} u \\ w \end{bmatrix} \neq 0. \tag{2.39}$$

Using (2.32), we have that

$$0 = F(x^k) + g'_{\mathcal{I}}(x^k)^\top \mu^k_{\mathcal{I}} + \Psi'_x(x^k, \mu^k)(y^k - x^k) + g'_{\mathcal{I}}(x^k)^\top(\lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}}).$$

Dividing by η_k and taking limit as $k \to \infty$, using (2.35) and (2.38) we obtain

$$0 = \Psi'_x(\bar{x}, \bar{\mu})u + g'_{\mathcal{I}}(\bar{x})^\top w. \tag{2.40}$$

By (2.33) and (2.34), using also that $\mu^k_{\mathcal{I}} \geq 0$, we have that

$$\langle \lambda^k_{\mathcal{I}}, g_{\mathcal{I}}(x^k) + g'_{\mathcal{I}}(x^k)(y^k - x^k) - \sigma_k(\lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}})\rangle = 0,$$

$$\langle \mu^k_{\mathcal{I}}, g_{\mathcal{I}}(x^k) + g'_{\mathcal{I}}(x^k)(y^k - x^k) - \sigma_k(\lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}})\rangle \leq 0.$$

Hence,

$$\langle \lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}}, g_{\mathcal{I}}(x^k) + g'_{\mathcal{I}}(x^k)(y^k - x^k) - \sigma_k(\lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}})\rangle \geq 0.$$

Dividing by η_k^2 and letting $k \to \infty$, using (2.35) and (2.36) we obtain that

$$\langle w, g'_{\mathcal{I}}(\bar{x})u\rangle \geq 0. \tag{2.41}$$

Also, from (2.33) and (2.34), dividing by η_k and taking limits we have that

$$\langle g'_i(\bar{x}), u\rangle = 0 \quad \forall i \in \mathcal{I}_+(\bar{x}, \bar{\mu}),\ \langle g'_i(\bar{x}), u\rangle \leq 0 \quad \forall i \in \mathcal{I}_0(\bar{x}, \bar{\mu}).$$

Thus $u \in \mathcal{C}(\bar{x}; D, F)$.

Multiplying by $u^\top$ in (2.40) and using (2.41), we obtain

$$0 \geq \langle \Psi'_x(\bar{x}, \bar{\mu})u, u\rangle,$$

so that SOC$^+$ (2.16) implies $u = 0$. Hence,

$$0 = g'_{\mathcal{I}}(\bar{x})^\top w. \tag{2.42}$$

Consider the QR-factorization of $g'_{\mathcal{I}}(\bar{x})$, that is

$$g'_{\mathcal{I}}(\bar{x}) = [U\ V]\begin{bmatrix} R \\ 0 \end{bmatrix},$$

where $[U\ V] \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$ is an orthogonal matrix and $R^\top$ has a null kernel (in particular, the columns of V give an orthonormal basis for $\ker g'_{\mathcal{I}}(\bar{x})^\top$).

Since

$$g_{\mathcal{I}}(x^k) = g_{\mathcal{I}}(\bar{x}) + g'_{\mathcal{I}}(\bar{x})(x^k - \bar{x}) + O(\|x^k - \bar{x}\|^2) = g'_{\mathcal{I}}(\bar{x})(x^k - \bar{x}) + O(\|x^k - \bar{x}\|^2)$$

and

$$V^\top g'_{\mathcal{I}}(\bar{x}) = 0, \tag{2.43}$$

we have that

$$V^\top g_{\mathcal{I}}(x^k) = O(\|x^k - \bar{x}\|^2).$$

By (2.18), we then have that

$$\|V^\top g_{\mathcal{I}}(x^k)\| \leq c_7 \sigma_k^2. \tag{2.44}$$

By (2.42), we have that $0 = g'_{\mathcal{I}}(\bar{x})^\top w = R^\top U^\top w$. Thus $U^\top w = 0$. Hence,

$$w = UU^\top w + VV^\top w = VV^\top w. \tag{2.45}$$

Let

$$\mathcal{I}^k = \{i \in \mathcal{I} \mid g_i(x^k) + \langle g'_i(x^k), y^k - x^k\rangle - \sigma_k(\lambda_i^k - \mu_i^k) = 0\}.$$

Clearly, there exists an index set $\mathcal{J}$ such that $\mathcal{I}^k = \mathcal{J}$ for infinitely many indices k. From now on, we consider the subsequence such that $\mathcal{I}^k = \mathcal{J}$, without introducing further subindices.

If $i \notin \mathcal{J}$ then $\lambda_i^k = 0$, so that $\lambda_i^k - \mu_i^k = -\mu_i^k \leq 0$. Thus from (2.39), $w_i \leq 0$ for all $i \notin \mathcal{J}$.

Let us define the cone

$$Q = \{\xi \in \mathbb{R}^{|\mathcal{I}|} \mid \xi_i = 0,\, i \in \mathcal{J};\; \xi_i \geq 0,\, i \notin \mathcal{J}\}.$$

Since $w_i \leq 0$ for $i \notin \mathcal{J}$, it holds that

$$-w \in Q^*.$$

By (2.33) and (2.34), we have that

$$-g_{\mathcal{I}}(x^k) - g'_{\mathcal{I}}(x^k)(y^k - x^k) + \sigma_k(\lambda_{\mathcal{I}}^k - \mu_{\mathcal{I}}^k) \in Q.$$

Multiplying this relation by $V^\top$, dividing by $\eta_k\sigma_k$ and using (2.43), gives

$$-\frac{V^\top g_{\mathcal{I}}(x^k)}{\eta_k\sigma_k} - \frac{V^\top(g'_{\mathcal{I}}(x^k) - g'_{\mathcal{I}}(\bar{x}))}{\sigma_k}\frac{(y^k - x^k)}{\eta_k} + \frac{V^\top(\lambda_{\mathcal{I}}^k - \mu_{\mathcal{I}}^k)}{\eta_k} \in V^\top Q.$$

Taking limits, using (2.44), (2.37), (2.35) and the facts that $(y^k - x^k)/\eta_k \to u = 0$ and that the set $V^\top Q$ is closed, we obtain that

$$V^\top w \in V^\top Q.$$

Then there exists $\xi \in Q$ such that $V^\top w = V^\top \xi$. Since $-w \in Q^*$ and $w = VV^\top w$, we conclude that

$$0 \geq \langle w, \xi\rangle = \langle VV^\top w, \xi\rangle = \langle VV^\top \xi, \xi\rangle = \|V^\top \xi\|^2.$$

Thus $V^\top w = V^\top \xi = 0$, so that (2.45) implies $w = 0$.

Then $(u, w) = 0$, in contradiction with (2.39). ∎

We now extend the solution of (2.29) to the solution of our subproblem (2.14), showing also that the needed localization property holds.

Theorem 2.3.6 *Suppose that SOC*$^+$ (2.16) *holds at* $(\bar{x}, \bar{\mu})$ *and that* σ *satisfies both inequalities in* (2.18).

Then there exist a neighborhood $\mathcal{V}$ *of* $(\bar{x}, \bar{\mu})$ *and a constant* $\gamma_4 > 0$ *such that for all* $(x, \mu) \in \mathcal{V} \cap (\mathbb{R}^n \times \mathbb{R}^m_+)$ *with* $\sigma(x, \mu) > 0$, *there exists* $(\bar{y}, \bar{\lambda})$, *a solution of the mixed complementarity problem of finding* $(y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ *such that*

$$\begin{array}{l} 0 = F(x) + \Psi'_x(x, \mu)(y - x) + g'(x)^\top \lambda, \\ 0 \leq \lambda \perp [g(x) + g'(x)(y - x) - \sigma(x, \mu)(\lambda - \mu)] \leq 0, \end{array} \tag{2.46}$$

satisfying

$$\left\| \begin{bmatrix} \bar{y} - x \\ \bar{\lambda} - \mu \end{bmatrix} \right\| \leq \gamma_4 \sigma(x, \mu). \tag{2.47}$$

Proof. By Proposition 2.3.5, there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and a constant $\gamma_3 > 0$ such that

$$\left\| \begin{bmatrix} y - x \\ \lambda_{\mathcal{I}} - \mu_{\mathcal{I}} \end{bmatrix} \right\| \leq \gamma_3 \sigma(x, \mu), \tag{2.48}$$

for any $(x, \mu) \in \mathcal{V}$ such that $\sigma(x, \mu) > 0$ and any solution $(y, \lambda_{\mathcal{I}})$ of (2.29).

Set $\bar{y} = y$, $\bar{\lambda}_{\mathcal{I}} = \lambda_{\mathcal{I}}$ and $\bar{\lambda}_i = 0$ for all $i \notin \mathcal{I}$. With this choice, the system (2.29) gives the equality in (2.46), as well as the complementarity condition in (2.46) for those indices in $\mathcal{I}_0(\bar{x}, \bar{\mu})$.

For $i \notin \mathcal{I}$, we have that

$$\begin{aligned} g_i(x) + \langle g'_i(x), \bar{y} - x \rangle - \sigma(x, \mu)(\bar{\lambda}_i - \mu_i) &= g_i(x) + \langle g'_i(x), \bar{y} - x \rangle + \sigma(x, \mu)\mu_i \\ &\leq g_i(\bar{x})/2 < 0 \end{aligned}$$

if (x, μ) is sufficiently close to $(\bar{x}, \bar{\mu})$ (so that $\sigma(x, \mu)$ is small enough and, consequently, so is $(\bar{y} - x)$, by (2.48)). This verifies the complementarity conditions in (2.46) for the indices not in $\mathcal{I}$.

Given the second relation in (2.29), it remains to check the nonnegativity of $\bar{\lambda}_i$, $i \in \mathcal{I}_+(\bar{x}, \bar{\mu})$. For $i \in \mathcal{I}_+(\bar{x}, \bar{\mu})$, we have that

$$\bar{\lambda}_i = \mu_i + (\bar{\lambda}_i - \mu_i) \geq \bar{\mu}_i/2 > 0$$

if (x, μ) is sufficiently close to $(\bar{x}, \bar{\mu})$ (so that $\sigma(x, \mu)$ is small enough and, consequently, so is $(\bar{\lambda}_{\mathcal{I}} - \mu_{\mathcal{I}})$, by (2.48)).

This concludes the proof of the existence of a solution of (2.46). Finally, let $\hat{\mu} = \Pi_{\mathcal{M}(\bar{x})}(\mu)$. For $i \notin \mathcal{I}$, we have that

$$|\bar{\lambda}_i - \mu_i| = |\mu_i| = |\mu_i - \hat{\mu}_i| \leq \frac{1}{\beta_1}\sigma(x, \mu),$$

by (2.18). Combining this relation with (2.48) we see that (2.47) holds. ∎

Theorem 2.3.6 establishes that Assumption 3 of Theorem 2.2.1 is satisfied for $\Sigma_0 = \{(\bar{x}, \bar{\mu})\}$. In particular, subproblems given by (2.8) (equivalently, by (2.14)) are locally solvable, and the distance between consecutive iterates can be bounded above by a measure of violation of KKT conditions for the original problem (2.1).

2.4 Upper Lipschitz-continuity of the solution set and a new error bound for KKT systems

This section shows satisfaction of Assumption 1 in Theorem 2.2.1, under the SOC

$$\langle \Psi'_x(\bar{x}, \bar{\mu})u, u\rangle \neq 0 \qquad \forall\, u \in \mathcal{C}(\bar{x}; D, F)\backslash\{0\}, \tag{2.49}$$

which is an extension of condition (2.16) used in Section 2.3 (Note that since the cone $\mathcal{C}(\bar{x}; D, F)$ is convex, (2.49) means that the inequality holds for all $u \in \mathcal{C}(\bar{x}; D, F)\backslash\{0\}$ either with the positive sign or with the negative sign). We also show that the so-called natural residual [30]

$$\sigma : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+, \quad \sigma(x, \mu) = \left\| \begin{bmatrix} \Psi(x, \mu) \\ \min\{-g(x), \mu\} \end{bmatrix} \right\|, \tag{2.50}$$

where the minimum is applied component-wise, provides a local error bound (2.18) for the solution set of the KKT system (2.9) under SOC (2.49). Note that, with this choice, the right-most inequality in (2.18) follows from Lipschitz-continuity of the functions involved and the fact that $\sigma(\bar{x}, \bar{\mu}) = 0$ for any $\bar{\mu} \in \mathcal{M}(\bar{x})$.

We start with considering the following problem with affine constraints: find $(x, \mu) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{gathered} 0 = F(x) + A^\top \mu, \\ 0 \leq \mu \perp [Ax + b] \leq 0, \end{gathered} \tag{2.51}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. This is the KKT system associated to the variational problem

$$\text{find } x \in \tilde{D} \quad \text{s.t.} \quad \langle F(x), y - x\rangle \geq 0 \qquad \forall\, y \in \tilde{D}, \tag{2.52}$$

$$\tilde{D} = \{x \in \mathbb{R}^n \mid Ax + b \leq 0\}.$$

We first prove local uniqueness of the primal part of the solution of (2.51), under SOC (1.50). Note that in the case of affine constraints, $\Psi'_x(\bar{x}, \bar{\mu}) = F'(\bar{x})$. Our result is an extension of [16, Proposition 1], where the optimization case (2.3) is considered, under the assumption that $F'(\bar{x}) = f''(\bar{x})$ is strictly copositive on the critical cone (2.5).

Proposition 2.4.1 *Let F be continuously differentiable at a solution $(\bar{x}, \bar{\mu})$ of (2.51) such that*

$$\langle F'(\bar{x})u, u\rangle \neq 0 \qquad \forall\, u \in \mathcal{C}(\bar{x}; \tilde{D}, F)\backslash\{0\}.$$

Then there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that if $x \in \mathcal{V}$ and (x, μ) is a solution of (2.51), then $x = \bar{x}$.

Proof. Suppose the contrary, i.e., that there exists a sequence (x^k, μ^k) of solutions of (2.51) such that $x^k \to \bar{x}$, $x^k \neq \bar{x}$. Taking a subsequence, if necessary, we can assume that

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to u \neq 0.$$

Using that $\mathcal{I}(x^k) \subset \mathcal{I}(\bar{x})$ for k sufficiently large, we have that if $i \notin \mathcal{I}(\bar{x})$ then $i \notin \mathcal{I}(x^k)$ and, hence, $\mu_i^k = 0$. Thus if $i \notin \mathcal{I}(\bar{x})$ then $\mu_i^k (A\bar{x} + b)_i = 0$ for all k sufficiently large. Since this equality holds trivially for $i \in \mathcal{I}(\bar{x})$, we conclude that

$$\langle \mu^k, A\bar{x} + b \rangle = 0 \tag{2.53}$$

for all k sufficiently large.

Since (x^k, μ^k) is a solution of (2.51), we have that

$$\begin{aligned} 0 &= \langle F(x^k) + A^\top \mu^k, x^k - \bar{x} \rangle = \langle F(x^k), x^k - \bar{x} \rangle + \langle \mu^k, A(x^k - \bar{x}) \rangle \\ &= \langle F(x^k), x^k - \bar{x} \rangle + \langle \mu^k, Ax^k + b \rangle - \langle \mu^k, A\bar{x} + b \rangle \\ &= \langle F(x^k), x^k - \bar{x} \rangle, \end{aligned} \tag{2.54}$$

where in the last equation we use (2.53) and the complementarity condition for (x^k, μ^k). Dividing (2.54) by $\|x^k - \bar{x}\|$ and taking limits, we obtain that

$$\langle F(\bar{x}), u \rangle = 0. \tag{2.55}$$

If $i \in \mathcal{I}(\bar{x})$, then $(A(x^k - \bar{x}))_i = (Ax^k + b)_i \leq 0$. Dividing this inequality by $\|x^k - \bar{x}\|$ and taking limits, we obtain that

$$(Au)_i \leq 0 \text{ for all } i \in \mathcal{I}(\bar{x}). \tag{2.56}$$

Together with (2.55) this shows that $u \in \mathcal{C}(\bar{x}; \tilde{D}, F) \setminus \{0\}$.

Also, note that

$$\langle F(x^k), u \rangle = -\langle A^\top \mu^k, u \rangle = -\sum_{i \in \mathcal{I}(x^k)} \mu_i^k (Au)_i \geq 0, \tag{2.57}$$

where in the last inequality we use (2.56) and the fact that $\mathcal{I}(x^k) \subset \mathcal{I}(\bar{x})$.

Using now (2.55) and (2.57) we have that

$$\begin{aligned} 0 &= \langle F(\bar{x}), u \rangle = \langle F(x^k) + F'(x^k)(\bar{x} - x^k), u \rangle + o(\|x^k - \bar{x}\|) \\ &\geq \langle F'(x^k)(\bar{x} - x^k), u \rangle + o(\|x^k - \bar{x}\|). \end{aligned}$$

Dividing this relation by $\|x^k - \bar{x}\|$ and taking the limit, we conclude that

$$0 \leq \langle F'(\bar{x})u, u\rangle.$$

On the other hand, using (2.54) and the fact that $\bar{x}$ is a solution of the variational problem (2.52), while $x^k \in \tilde{D}$, we have

$$\begin{aligned} 0 &= \langle F(x^k), x^k - \bar{x}\rangle \\ &= \langle F(\bar{x}), x^k - \bar{x}\rangle + \langle F'(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|^2) \\ &\geq \langle F'(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|^2). \end{aligned}$$

Dividing this relation by $\|x^k - \bar{x}\|^2$ and letting $k \to \infty$, we obtain in the limit

$$0 \geq \langle F'(\bar{x})u, u\rangle.$$

Hence,

$$\langle F'(\bar{x})u, u\rangle = 0$$

for $u \in \mathcal{C}(\bar{x}; \tilde{D}, F)\backslash\{0\}$, in contradiction with the assumption. ■

Let now $\tilde{F}(x) = Mx + q$, where $M \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$. Consider the KKT system:

$$\begin{aligned} 0 = Mx + q + A^\top \mu, \\ 0 \leq \mu \perp [Ax + b] \leq 0, \end{aligned} \tag{2.58}$$

associated to the affine variational problem

$$\text{find } x \in \tilde{D} \quad \text{s.t.} \quad \langle \tilde{F}(x), y - x\rangle \geq 0 \qquad \forall\, y \in \tilde{D}.$$

Define

$$T(x,\mu) = \begin{bmatrix} Mx + A^\top \mu \\ -Ax \end{bmatrix} \text{ and } \psi = \begin{bmatrix} -q \\ b \end{bmatrix},$$

so that (2.58) is equivalent to the generalized equation

$$\psi \in T(x,\mu) + \mathcal{N}(x,\mu), \tag{2.59}$$

where $\mathcal{N}$ is defined in (2.13).

The following is an extension of [16, Lemma 1], where M is assumed to be symmetric and strictly copositive on $\mathcal{C}(\bar{x}; \tilde{D}, \tilde{F})$, to variational setting.

Lemma 2.4.2 *Suppose that $(\bar{x}, \bar{\mu})$ is a solution of (2.59) for $\bar{\psi}$ and that*

$$\langle Mu, u\rangle \neq 0 \qquad \forall\, u \in \mathcal{C}(\bar{x}; \tilde{D}, \tilde{F})\backslash\{0\}.$$

Then there exist $\beta > 0$ and neighborhoods $\mathcal{V}$ of $\bar{x}$ and $\mathcal{U}$ of $\bar{\psi}$ such that, if (x,μ) is a solution of (2.59) for $\psi \in \mathcal{U}$ and $x \in \mathcal{V}$, then

$$\|x - \bar{x}\| \leq \beta\|\psi - \bar{\psi}\|.$$

Proof. As it is well known [35], the multifunction $\mathcal{F}(x,\mu) = T(x,\mu) + \mathcal{N}(x,\mu)$ and its inverse

$$\mathcal{F}^{-1}(\psi) = \{\omega \in \mathbb{R}^n \times \mathbb{R}^m \mid 0 \in \mathcal{F}(\omega) - \psi\},$$

are polyhedral multifunctions. Furthermore, the function $\mathcal{P}$ such that $\mathcal{P}(x,\mu) = x$ is polyhedral, and so is the composition $\mathcal{P} \circ \mathcal{F}^{-1}$.

By [35, Proposition 1], polyhedral multifunctions are locally upper Lipschitzian at every point, and the Lipschitz constant is independent of the point. Thus, there exist a constant $\beta > 0$ and a neighborhood $\mathcal{U}$ of $\bar{\psi}$ such that

$$\mathcal{P} \circ \mathcal{F}^{-1}(\psi) \subset \mathcal{P} \circ \mathcal{F}^{-1}(\bar{\psi}) + \beta\|\psi - \bar{\psi}\|B \quad \forall\, \psi \in \mathcal{U}. \tag{2.60}$$

Since $\mathcal{P} \circ \mathcal{F}^{-1}(\psi)$ is the set of x-components of solutions of (2.58), by Proposition 2.4.1 there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that

$$\mathcal{P} \circ \mathcal{F}^{-1}(\bar{\psi}) \cap \mathcal{V} = \{\bar{x}\}.$$

Let $\rho = \operatorname{dist}(\bar{x}, \mathcal{P} \circ \mathcal{F}^{-1}(\bar{\psi}) \backslash \{\bar{x}\})$, and choose $\mathcal{V}$ smaller if necessary so that $\mathcal{V} \subset \{\bar{x}\} + \frac{\rho}{3}B$. Choose $\mathcal{U}$ sufficiently small so that

$$\{\bar{x}\} + \beta\|\psi - \bar{\psi}\|B \subset \mathcal{V} \quad \forall\, \psi \in \mathcal{U}.$$

If $\psi \in \mathcal{U}$ and $x \in \mathcal{P} \circ \mathcal{F}^{-1}(\psi) \cap \mathcal{V}$, we obtain from (2.60) that there exist $\hat{x} \in \mathcal{P} \circ \mathcal{F}^{-1}(\bar{\psi})$ and $p \in B$ such that $x = \hat{x} + \beta\|\psi - \bar{\psi}\|p$. But then,

$$\|\bar{x} - \hat{x}\| = \left\|\bar{x} - x + \beta\|\psi - \bar{\psi}\|p\right\| \leq \|x - \bar{x}\| + \beta\|\psi - \bar{\psi}\| \leq \frac{\rho}{3} + \frac{\rho}{3} < \rho,$$

implying that $\hat{x} = \bar{x}$. Hence, $x = \bar{x} + \beta\|\psi - \bar{\psi}\|p$ for some $p \in B$, i.e.,

$$\|x - \bar{x}\| \leq \beta\|\psi - \bar{\psi}\|.$$

■

Thus, for our main problem (2.10) we can state the following error estimates, that verify the upper Lipschitz-continuity property of the solution set of KKT systems.

Lemma 2.4.3 *Let F be differentiable and g twice differentiable at $\bar{x}$, and suppose that there exists $\bar{\mu} \in \mathcal{M}(\bar{x})$ such that $(\bar{x}, \bar{\mu})$ satisfies SOC* (2.49).

Then there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and constants $\gamma, \tau > 0$ such that for every $(x,\mu) \in \mathcal{V}$ and for each $p \in \gamma B$ satisfying

$$0 \in G(x,\mu) + \mathcal{N}(x,\mu) + p, \tag{2.61}$$

where G and $\mathcal{N}$ are defined in (2.13)*, it holds that*

$$\|x - \bar{x}\| + \|\mu - \Pi_{\mathcal{M}(\bar{x})}(\mu)\| \leq \tau\|p\|.$$

Proof. Throughout this proof, the generic constant β is uniformly bounded when $\mathcal{V}$ is sufficiently small.

Consider the affine variational problem (2.59) with

$$M = \Psi'_x(\bar{x}, \bar{\mu}) \quad \text{and} \quad A = g'(\bar{x}).$$

Let $\mathcal{F}(x, \mu) = T(x, \mu) + \mathcal{N}(x, \mu)$.

Define $\psi^1 = T(x, \mu) - G(x, \mu) - p$, where p satisfies (2.61). Then $(x, \mu) \in \mathcal{F}^{-1}(\psi^1)$. Define $\psi^2 = T(\bar{x}, \mu) - G(\bar{x}, \mu)$. Since $F(\bar{x}) + A^\top \bar{\mu} = 0$, we have that

$$\psi^2 = \left[\begin{array}{c} M\bar{x} + A^\top \mu - F(\bar{x}) - A^\top \mu \\ -A\bar{x} + g(\bar{x}) \end{array} \right] = \left[\begin{array}{c} M\bar{x} + A^\top \bar{\mu} \\ -A\bar{x} \end{array} \right] + \left[\begin{array}{c} 0 \\ g(\bar{x}) \end{array} \right].$$

As $g(\bar{x}) \in \mathcal{N}_{\mathbb{R}^m_+}(\bar{\mu})$, this shows that $(\bar{x}, \bar{\mu}) \in \mathcal{F}^{-1}(\psi^2)$.

By the differentiability assumptions, ψ^1 is close to ψ^2 when (x, μ) is close to $(\bar{x}, \bar{\mu})$ and p is close to 0. Consequently, by choosing $\mathcal{V}$ and γ sufficiently small, Lemma 2.4.2 gives us the estimate

$$\|x - \bar{x}\| \le \beta \|\psi^1 - \psi^2\| = \beta \left\| G(x, \mu) - G(\bar{x}, \mu) - (T(x, \mu) - T(\bar{x}, \mu)) + p \right\|, \tag{2.62}$$

for all $(x, \mu) \in \mathcal{V}$ and $p \in \gamma B$.

Given any $\varepsilon > 0$, using the differentiability assumptions and taking $\mathcal{V}$ sufficiently small, we obtain that

$$\|G(x, \mu) - G(\bar{x}, \mu) - T(x - \bar{x}, 0)\| \le \varepsilon \|x - \bar{x}\| \qquad \forall\, (x, \mu) \in \mathcal{V}.$$

Combining this with (2.62), we have that

$$\|x - \bar{x}\| \le \beta\varepsilon \|x - \bar{x}\| + \beta \|p\|.$$

Thus, taking $\varepsilon < 1/\beta$, we obtain

$$\|x - \bar{x}\| \le \frac{\beta}{1 - \varepsilon\beta} \|p\|. \tag{2.63}$$

Consider the decomposition $p = (u, v) \in \mathbb{R}^n \times \mathbb{R}^m$. If $i \notin \mathcal{I}$ then $g_i(\bar{x}) < 0$. Thus, we can take $\mathcal{V}$ and γ small enough, so that $g_i(x) - v_i < 0$. From (2.61) we have that $g(x) - v \in \mathcal{N}_{\mathbb{R}^m_+}(\mu)$. Hence, $\mu_i = 0$ for all $i \notin \mathcal{I}$ and $\mu_i \ge 0$ for $i \in \mathcal{I}$. Since

$$\mathcal{M}(\bar{x}) = \{\nu \in \mathbb{R}^m \mid F(\bar{x}) + g'(\bar{x})^\top \nu = 0;\ \nu_i \ge 0, i \in \mathcal{I};\ \nu_i = 0, i \notin \mathcal{I}\},$$

by Hoffman's error bound for linear systems, we obtain that

$$\|\mu - \hat{\mu}\| \le \beta \|F(\bar{x}) + g'(\bar{x})^\top \mu\| = \beta \|\Psi(\bar{x}, \mu)\|, \tag{2.64}$$

where $\hat{\mu} = \Pi_{\mathcal{M}(\bar{x})}(\mu)$.

From (2.61), we have that $\Psi(x, \mu) + u = 0$. Then, using the differentiability assumptions and taking $\mathcal{V}$ smaller if necessary, we have

$$\|\Psi(\bar{x}, \mu)\| \leq \|\Psi(x, \mu)\| + \|\Psi(\bar{x}, \mu) - \Psi(x, \mu)\| \leq \|u\| + \beta\|x - \bar{x}\|.$$

Since $\|u\| \leq \|p\|$, using (2.63), we obtain

$$\|\Psi(\bar{x}, \mu)\| \leq \beta\|p\|.$$

Combining this with (2.63) and (2.64) gives

$$\|x - \bar{x}\| + \|\mu - \hat{\mu}\| \leq \tau\|p\|,$$

for some $\tau > 0$. ∎

This result shows satisfaction of Assumption 1 in Theorem 2.2.1 for $\Sigma_0 = \{(\bar{x}, \bar{\mu})\}$. Moreover, taking $w = (x, \mu)$ we have that

$$\sigma(w) = \left\| w - \Pi_{\mathbb{R}^n \times \mathbb{R}^m_+}(w - G(w)) \right\|,$$

for σ given by (2.50) and G given by (2.13). Hence, by Theorem 2.2.2, it now also follows that the natural residual (2.50) provides a valid local error bound for the KKT system (2.9). Specifically, we have the following.

Theorem 2.4.4 *Let F be differentiable and g twice differentiable at $\bar{x}$, and suppose there exists $\bar{\mu} \in \mathcal{M}(\bar{x})$ such that $(\bar{x}, \bar{\mu})$ satisfies SOC* (2.49)*.*

Then there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and constants $\beta_2 \geq \beta_1 > 0$ such that for all $(x, \mu) \in \mathcal{V}$ the function σ defined in (2.50) *satisfies the error bound* (2.18)*.*

We note that Theorem 2.4.4 gives the first error bound for KKT systems in variational context that does not subsume some regularity-type assumptions about the constraints. We refer the reader to [17] for a detailed discussion and comparisons of error bounds for KKT systems.

The provided error bound completes the proof of superlinear convergence of our method, that we formalize as follows.

Theorem 2.4.5 *Let F be differentiable and g twice differentiable at $\bar{x}$, and suppose that there exists $\bar{\mu} \in \mathcal{M}(\bar{x})$ such that $(\bar{x}, \bar{\mu})$ satisfies SOC$^+$* (2.16)*.*

Then there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that for any $(x^0, \mu^0) \in \mathcal{V} \cap (\mathbb{R}^n \times \mathbb{R}^m_+)$, the iterates defined by (2.12) *are well defined and converge superlinearly to $(\bar{x}, \mu)$, where μ is some element of $\mathcal{M}(\bar{x})$. Furthermore, the convergence is quadratic if F' and g'' are Lipschitz-continuous in a neighborhood of $\bar{x}$.*

2.5 Concluding remarks

The approach presented here can be used also to prove the uniqueness of solutions of subproblems (2.14), extending the result for optimization under SSOSC (2.7) obtained in [15] (see also [13]). In our case, suppose that

$$\langle \Psi'_x(\bar{x},\bar{\mu})u, u\rangle > 0 \quad \forall\, u \in \mathcal{C}^+(\bar{x},\bar{\mu}) \setminus \{0\}, \tag{2.65}$$

where

$$\mathcal{C}^+(\bar{x},\bar{\mu}) = \{u \in \mathbb{R}^n \mid \langle g'_i(\bar{x}), u\rangle = 0 \quad \forall i \in \mathcal{I}_+(\bar{x},\bar{\mu})\}.$$

Regarding the proof of Proposition 2.3.1, it can be seen that under SSOC (2.65) there exists a constant $\gamma_2 > 0$ such that for all (x,μ) in a neighborhood of $(\bar{x},\bar{\mu})$ it holds that

$$\langle \Psi'_x(x,\mu)u, u\rangle + \sigma(x,\mu)\|v\|^2 \geq \gamma_2\Big(\|u\|^2 + \sigma(x,\mu)\|v\|^2\Big) \quad \forall (u,v) \in K^+(x,\mu), \tag{2.66}$$

where

$$K^+(x,\mu) = \Big\{(u,v) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \mid \langle g'_i(x), u\rangle = \sigma(x,\mu)v_i, \quad i \in \mathcal{I}_+(\bar{x},\bar{\mu})\Big\}. \tag{2.67}$$

Since $\mathcal{C}(\bar{x}; D, F) \subset \mathcal{C}^+(\bar{x},\bar{\mu})$, we have that SOC$^+$ (2.16) and, thus, Propositions 2.3.1 and 2.3.4 hold. In particular, in the proof of Proposition 2.3.4, the generalized complementarity problem

$$\text{find } \bar{d} \quad \text{s.t.} \quad K \ni \bar{d} \perp M\bar{d} + M\tilde{z} + q \in K^*,$$

where K is given by (2.20), has a nonempty compact solution set. Let d^1 and d^2 be solutions of this complementarity problem. Then

$$\begin{aligned} \langle M(d^1 - d^2), d^1 - d^2\rangle &= \langle Md^1 + M\tilde{z} + q - (Md^2 + M\tilde{z} + q), d^1 - d^2\rangle \\ &= -\langle Md^1 + M\tilde{z} + q, d^2\rangle - \langle Md^2 + M\tilde{z} + q, d^1\rangle \\ &\leq 0. \end{aligned} \tag{2.68}$$

Since $K^+(x,\mu)$ is a subspace and $d^1, d^2 \in K \subset K^+(x,\mu)$, we have that

$$d^1 - d^2 \in K^+(x,\mu).$$

Since (2.66) implies that M is strictly copositive on $K^+(x,\mu)$, from (2.68) we conclude that $d^1 - d^2 = 0$. Hence, the mixed complementarity problem (2.29) has a unique solution.

Let us now show that under SSOC (2.65), for (x,μ) sufficiently close to $(\bar{x},\bar{\mu})$ we have that $(\bar{y},\bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m_+$, where $\bar{\lambda}_i = 0, i \notin \mathcal{I}$ and $(\bar{y},\bar{\lambda}_{\mathcal{I}})$ is the solution of (2.29), is the unique solution of (2.46) satisfying (2.47). By Theorem 2.3.6, $(\bar{y},\bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ defined in this way is a solution of (2.46) satisfying (2.47). Conversely, if $(\bar{y},\bar{\lambda}) \in$

$\mathbb{R}^n \times \mathbb{R}^m_+$ is a solution of (2.46) satisfying (2.47), and if (x, μ) is sufficiently close to $(\bar{x}, \bar{\mu})$, we have that

$$g_i(x) + \langle g_i'(x), \bar{y} - x\rangle - \sigma(x, \mu)(\bar{\lambda}_i - \mu_i) < 0,\ i \notin \mathcal{I},$$

$$\bar{\lambda}_i > 0,\ i \in \mathcal{I}_+(\bar{x}, \bar{\mu}).$$

Then by the complementarity conditions in (2.46), we obtain that

$$\bar{\lambda}_i = 0,\ i \notin \mathcal{I},$$

$$g_i(x) + \langle g_i'(x), \bar{y} - x\rangle - \sigma(x, \mu)(\bar{\lambda}_i - \mu_i) = 0,\ i \in \mathcal{I}_+(\bar{x}, \bar{\mu}).$$

Hence, $(\bar{y}, \bar{\lambda}_{\mathcal{I}})$ is a solution of (2.29), which has been established to be unique.

Chapter 3

Local Convergence of Sequential Quadratically Constrained Quadratic Programming type method for Variational Problems

This chapter corresponds to the material from paper [11].

We consider the class of quadratically-constrained quadratic-programming methods in the framework extended from optimization to more general variational problems. Previously, in the optimization case, Anitescu (2002) showed superlinear convergence of the primal sequence under the Mangasarian-Fromovitz constraint qualification and the quadratic growth condition. Quadratic convergence of the primal-dual sequence was established by Fukushima, Luo and Tseng (2003) under the convexity assumptions, the Slater constraint qualification, and a strong second-order sufficient condition. We obtain a new local convergence result, which complements the above (it is neither stronger nor weaker): we prove primal-dual quadratic convergence under the linear independence constraint qualification, strict complementarity, and a second-order sufficiency condition. Additionally, our results apply to variational problems beyond the optimization case. Finally, we provide a necessary and sufficient condition for superlinear convergence of the primal sequence under a Dennis-Moré type condition.

3.1 Introduction

Given sufficiently smooth mappings $F : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ (precise smoothness requirements will be specified later, within the statements of our convergence results), we consider the following variational problem [10]:

$$\text{find } \bar{x} \in D \text{ s.t. } \langle F(\bar{x}), y - \bar{x} \rangle \geq 0 \qquad \forall\, y \in \bar{x} + \mathcal{T}_D(\bar{x}), \tag{3.1}$$

where

$$D = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0,\ i = 1, \ldots, m\}$$

and $\mathcal{T}_D(\bar{x})$ is the (standard) tangent cone to D at $\bar{x} \in D$. When for some smooth function $f : \mathbb{R}^n \to \mathbb{R}$ it holds that

$$F(x) = f'(x), \quad x \in \mathbb{R}^n, \tag{3.2}$$

then (3.1) describes (primal) first-order necessary optimality conditions for the optimization problem

$$\min\ f(x) \quad \text{s.t.} \quad x \in D. \tag{3.3}$$

We consider the following iterative procedure. (As it will be seen below, in the case of the optimization problem (3.3), it reduces to the sequential quadratically-constrained quadratic-programming method, e.g., [1, 14, 38]. In the variational setting, this method appears to be new.) If $x^k \in \mathbb{R}^n$ is the current iterate, then the next iterate x^{k+1} is obtained as a solution of the the following approximation of the variational problem (3.1):

$$\text{find } x \in D_k \text{ s.t. } \langle F_k(x), y - x\rangle \geq 0 \qquad \forall\, y \in x + \mathcal{T}_{D_k}(x), \tag{3.4}$$

where

$$F_k(x) = F(x^k) + F'(x^k)(x - x^k), \quad x \in \mathbb{R}^n,$$

$$D_k = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} g_i(x^k) + \langle g_i'(x^k), x - x^k\rangle + \frac{1}{2}\langle g_i''(x^k)(x - x^k), x - x^k\rangle \leq 0, \\ i = 1, \ldots, m. \end{array} \right\},$$

and $\mathcal{T}_{D_k}(x)$ is the tangent cone to D_k at $x \in D_k$. Subproblem (3.4) can be considered as a "one-step-further" approximation when compared to the classical Josephy-Newton method for variational inequalities [21, 10], where at every step the mapping F is approximated to the first order (as in (3.4)), but the set D is not being simplified (unlike in (3.4)). Specifically, given the current iterate x^k, the Josephy-Newton method solves the following subproblem:

$$\text{find } x \in D \text{ s.t. } \langle F_k(x), y - x\rangle \geq 0 \qquad \forall\, y \in x + \mathcal{T}_D(x). \tag{3.5}$$

It is clear that subproblem (3.4) is structurally simpler than (3.5) (in (3.5) constraints are general nonlinear, while in (3.4) they are quadratic). Thus, in principle, (3.4) should be easier to solve. That said, we shall not be concerned here with specific methods for solving subproblems of the structure of (3.4) (at the very least, the same techniques as for (3.5) can be used). In the case of optimization, as discussed below, specific methods are readily available.

For optimization problems (3.3), an iteration of the sequential quadratically-constrained quadratic-programming method (SQCQP) consists of minimizing a quadratic approximation of the objective function subject to a quadratic approximation of the constraints. Specifically, if $x^k \in \mathbb{R}^n$ is the current iterate, then the next iterate x^{k+1} is obtained as a solution of the following approximation of the original problem:

$$\min \langle f'(x^k), x - x^k \rangle + \frac{1}{2}\langle f''(x^k)(x - x^k), x - x^k \rangle \quad \text{s.t. } x \in D_k. \tag{3.6}$$

Note that taking into account (3.2), the variational subproblem (3.4) describes (primal) first-order necessary optimality conditions for (3.6). Therefore, SQCQP for optimization is a special case in our framework.

As some previous work on SQCQP and related methods, we mention [31, 32, 40, 22, 1, 14, 38]. In the convex case, subproblem (3.6) can be cast as a second-order cone program [24, 28], which can be solved efficiently by interior-point algorithms (such as [26, 39]). In [1], nonconvex subproblems (3.6) were also handled quite efficiently by using other nonlinear programming techniques. Even though quadratically constrained subproblems are computationally more difficult than linearly constrained (as in the more traditional SQP methods, [2]), they are manageable by modern computational tools and the extra effort in solving them can be worth it. I.e., at least in some situations, one may expect that fewer subproblems will need to be solved, when compared to SQP. Some numerical validation of this can be found in computational experiments of [1].

In order to guarantee global convergence, SQCQP methods require some modifications to subproblem (3.6), as well as a linesearch procedure for an adequately chosen penalty function. (See, for example, [14, 38]). But under certain assumptions, locally all those modifications reduce precisely to (3.6). Moreover, the unit stepsize satisfies the linesearch criteria under very mild conditions [38, Proposition 8] (in particular, no second-order sufficiency is needed for this), which is one of the attractive features of SQCQP. Thus, what is relevant for local convergence analysis is precisely the method given by (3.6), and this is the subject of this chapter (except that we consider the more general variational setting of (3.4)). Note that, as a consequence of acceptance of the unit stepsize, the Maratos effect [25, 33] does not occur in SQCQP (of course, Maratos effect can also be avoided in SQP methods, by introducing second-order correction terms in the direction or by using an augmented Lagrangian merit function).

We next survey previous local rate of convergence results and compare them to ours. As already mentioned, in the variational setting, our method appears to be new. Therefore, we limit our discussion to the case of optimization. In [1], local primal superlinear rate of convergence of a trust-region SQCQP method is obtained under the Mangasarian-Fromovitz constraint qualification (MFCQ) and a certain quadratic growth condition. We note that, under MFCQ, quadratic growth is equivalent to the second-order sufficient condition for optimality (SOSC), see [5, Theorem 3.70]. Quadratic convergence of the primal-dual sequence is obtained in [14] (the dual part of the sequence is formed by the Lagrange multipliers associated to solutions of (3.6)). The assumptions in [14] are as follows: convexity of f and of g, the Slater condition (equivalent to MFCQ in the convex case) and a strong second-order sufficient condition (implying quadratic growth). This set of assumptions is stronger than in [1], but the assertions in the two papers are different and not comparable to each other. Thus, neither of the two results implies the other one. To complement the picture, in this chapter we prove a third local convergence result, which is in the same relation to the two previous ones: it neither follows from them nor implies them. Specifically, we shall establish primal-dual quadratic convergence under the linear independence constraint qualification (LICQ), strict complementarity condition, and SOSC. Compared to [14], our assumptions are essentially different (we do not make any convexity assumptions; while [14] makes weaker regularity assumptions). Our assertions are stronger than in [14], because in addition to primal-dual quadratic convergence we also prove superlinear primal convergence. Compared to [1], our assumptions are more restrictive, of course. But our assertions are stronger as well: we prove quadratic primal-dual convergence *and* superlinear primal convergence instead of superlinear primal convergence only. In addition, we shall exhibit a Dennis-Moré type [8] necessary and sufficient condition for superlinear convergence of the primal sequence in the case when the primal-dual convergence is given.

If $\Phi : \mathbb{R}^s \times \mathbb{R}^p \to \mathbb{R}^p$ is Lipschitz continuous in a neighborhood of a point $(\bar{\sigma}, \bar{\xi}) \in \mathbb{R}^s \times \mathbb{R}^p$, by $\partial\Phi(\bar{\sigma}, \bar{\xi})$ we denote the *Clarke generalized Jacobian* of Φ at $(\bar{\sigma}, \bar{\xi})$, i.e.,

$$\partial\Phi(\bar{\sigma}, \bar{\xi}) = \operatorname{conv}\left\{\lim_{l\to\infty} \Phi'(\sigma^l, \xi^l) \mid (\sigma^l, \xi^l) \to (\bar{\sigma}, \bar{\xi}), (\sigma^l, \xi^l) \in \mathcal{N}_\Phi\right\},$$

where conv denotes convex hull of a set, and $\mathcal{N}_\Phi$ is the set of points at which Φ is differentiable (by Rademacher's Theorem, Φ is differentiable almost everywhere in a neighborhood of $(\bar{\sigma}, \bar{\xi})$). In the sequel, we shall make use of the following Implicit Function Theorem.

Theorem 3.1.1 *[6, p. 256] Let $\Phi : \mathbb{R}^s \times \mathbb{R}^p \to \mathbb{R}^p$ be Lipschitz continuous in a neighborhood of a point $(\bar{\sigma}, \bar{\xi}) \in \mathbb{R}^s \times \mathbb{R}^p$ such that $\Phi(\bar{\sigma}, \bar{\xi}) = 0$.*

Suppose that the set of $p \times p$ matrices M, for which there exists a $p \times s$ matrix N such that $[N\,,\ M] \in \partial\Phi(\bar{\sigma}, \bar{\xi})$, has full rank.

Then there exist a neighborhood U_0 of $\bar{\sigma}$, a neighborhood Ω_0 of $\bar{\xi}$, and a unique Lipschitz continuous function $\xi : U_0 \to \Omega_0$ such that $\Phi(\sigma, \xi(\sigma)) = 0$ for all $\sigma \in U_0$.

3.2 Primal-dual quadratic convergence

As it is well-known, under adequate constraint qualifications (which would be the case here), the variational problem (3.1) is equivalent to solving the Karush-Kuhn-Tucker (KKT) system: find $(x,\, \mu) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{gathered} 0 = \Psi(x, \mu) = F(x) + g'(x)^\top \mu, \\ 0 \leq \mu \perp g(x) \leq 0. \end{gathered} \tag{3.7}$$

For the same reason, solutions of subproblem (3.4) are described by the following mixed complementarity problem [10] in $(x,\, \mu) \in \mathbb{R}^n \times \mathbb{R}^m$:

$$F(x^k) + F'(x^k)(x - x^k) + \sum_{i=1}^{m} \mu_i (g_i'(x^k) + g_i''(x^k)(x - x^k)) = 0, \tag{3.8}$$

and for all $i = 1, \ldots, m,$ it holds that

$$g_i(x^k) + \langle g_i'(x^k), x - x^k \rangle + \frac{1}{2} \langle g_i''(x^k)(x - x^k), x - x^k \rangle \le 0, \tag{3.9}$$

$$\mu_i \ge 0, \tag{3.10}$$

$$\mu_i (g_i(x^k) + \langle g_i'(x^k), x - x^k \rangle + \frac{1}{2} \langle g_i''(x^k)(x - x^k), x - x^k \rangle) = 0. \tag{3.11}$$

Note that in the case of the optimization problem (3.3), i.e., when (3.2) holds, the above are precisely the optimality conditions for the SQCQP subproblem (3.6).

Let $(\bar{x}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m$ be some fixed solution of the KKT system (3.7), which by virtue of further assumptions will be locally unique.

We say that LICQ holds at $\bar{x}$ if

$$\{g_i'(\bar{x}) \mid i \in \mathcal{I}\} \text{ is a linearly independent set,} \tag{3.12}$$

where

$$\mathcal{I} = \mathcal{I}(\bar{x}) = \{i \in \{1, \ldots, m\} \mid g_i(\bar{x}) = 0\}$$

is the index set of active constraints at $\bar{x} \in D$. Under LICQ, the multiplier $\bar{\mu}$ associated to the given $\bar{x}$ is unique by necessity. We shall also use the following partitioning of $\mathcal{I}$:

$$\mathcal{I}_+ = \mathcal{I}_+(\bar{x}, \bar{\mu}) = \{i \in \mathcal{I} \mid \bar{\mu}_i > 0\}, \quad \mathcal{I}_0 = \mathcal{I}_0(\bar{x}, \bar{\mu}) = \{i \in \mathcal{I} \mid \bar{\mu}_i = 0\} = \mathcal{I} \setminus \mathcal{I}_+,$$

corresponding to strongly and weakly active constraints, respectively.

We say that $(\bar{x}, \bar{\mu})$ satisfies the second-order condition (SOC) if

$$\langle \Psi_x'(\bar{x}, \bar{\mu})d,\, d \rangle \ne 0 \qquad \forall\, d \in \mathcal{C}(\bar{x}; D, F) \setminus \{0\}, \tag{3.13}$$

where

$$\mathcal{C}(\bar{x}; D, F) = \{u \in \mathbb{R}^n \mid \langle g_i'(\bar{x}), u \rangle = 0 \quad \forall\, i \in \mathcal{I}_+,\ \langle g_i'(\bar{x}), u \rangle \le 0 \quad \forall\, i \in \mathcal{I}_0\}. \tag{3.14}$$

Note that since the cone $\mathcal{C}(\bar{x}; D, F)$ is convex, (3.13) means that the quadratic form has the same nonzero sign for all $d \in \mathcal{C}(\bar{x}; D, F) \setminus \{0\}$. In the case of the optimization problem corresponding to (3.2), $\mathcal{C}(\bar{x}; D, F)$ is the standard critical cone of (3.3) at $\bar{x}$, and

$$\Psi(x,\, \mu) = L_x'(x,\, \mu),$$

where

$$L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, \quad L(x,\, \mu) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x)$$

is the Lagrangian of (3.3). Then (3.13) with the positive sign reduces to the classical second-order sufficient condition for optimality

$$\langle L''_{xx}(\bar{x}, \bar{\mu})d,\, d\rangle > 0 \qquad \forall\, d \in \mathcal{C}(\bar{x}; D, F) \setminus \{0\}.$$

Finally, we say that the condition of strict complementarity holds at $(\bar{x}, \bar{\mu})$ if $\mathcal{I}_0 = \emptyset$ or, equivalently,

$$\bar{\mu}_i > 0 \qquad \forall\, i \in \mathcal{I}. \tag{3.15}$$

We are now in position to state our first convergence result. Since we are not making any convexity/monotonicity type assumptions, even under the stated below conditions at $(\bar{x}, \bar{\mu})$, the mixed complementarity problem (3.8)-(3.11) (or the optimization subproblem (3.6)) may have solutions "of no interest", far from x^k (or $\bar{x}$). We therefore talk about the specific solution closest to x^k. This is typical in results of this nature.

Theorem 3.2.1 *Let $(\bar{x}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m$ be a solution of the KKT system* (3.7)*. Suppose that F is differentiable and g is twice differentiable in some neighborhood of $\bar{x}$, and that the first derivative of F and the second derivative of g are Lipschitz continuous in this neighborhood. Suppose further that LICQ* (3.12)*, SOC* (3.13) *and the strict complementarity condition* (3.15) *are satisfied.*

Then there exists a neighborhood $\mathcal{U}$ of $\bar{x}$ such that if $x^k \in \mathcal{U}$, then the mixed complementarity problem (3.8)-(3.11) *has a solution $(x^{k+1}, \mu^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^m$. Moreover, if $x^0 \in \mathcal{U}$ and, for each $k \geq 0$, x^{k+1} is the closest to x^k solution of* (3.8)-(3.11)*, then there exists a neighborhood $\mathcal{V}$ of $\bar{\mu}$ such that* (3.8)-(3.11) *defines a unique sequence $\{(x^{k+1}, \mu^{k+1})\}$ which stays in $\mathcal{U} \times \mathcal{V}$ and converges quadratically to $(\bar{x}, \bar{\mu})$.*

Proof. We first prove existence of a solution for the mixed complementarity problem (3.8)-(3.11), starting with the equations (3.8) and (3.11). To this end, we shall apply the Implicit Function Theorem (Theorem 3.1.1) to the mapping $\Phi : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ defined by

$$\Phi(x; y, \mu) = \begin{bmatrix} \Psi(x, \mu) + \Psi'_x(x, \mu)(y - x) \\ \mu_1(g_1(x) + \langle g'_1(x), y - x\rangle + \frac{1}{2}\langle g''_1(x)(y - x), y - x\rangle) \\ \vdots \\ \mu_m(g_m(x) + \langle g'_m(x), y - x\rangle + \frac{1}{2}\langle g''_m(x)(y - x), y - x\rangle) \end{bmatrix}. \tag{3.16}$$

Thinking of $x \in \mathbb{R}^n$ as a parameter, the system $\Phi(x; y, \mu) = 0$ has $n + m$ equations and $n + m$ unknowns $(y, \mu) \in \mathbb{R}^n \times \mathbb{R}^m$.

Since $(\bar{x}, \bar{\mu})$ is a solution of the KKT system (3.7), we have that $\Phi(\bar{x}; \bar{x}, \bar{\mu}) = 0$. By our smoothness hypotheses on F and g, Φ is Lipschitz continuous in a neighborhood of $(\bar{x};\bar{x},\bar{\mu})$. Moreover, since Φ is continuously differentiable with respect to y and μ, it easily follows that $\partial\Phi(\bar{x};\bar{x},\bar{\mu})$ is the set of matrices $[N, M]$, where M is given by

$$M = \left(\Phi'_y, \Phi'_\mu\right)(\bar{x};\bar{x},\bar{\mu}) = \begin{bmatrix} \Psi'_x(\bar{x},\bar{\mu}) & g'_1(\bar{x}) & g'_2(\bar{x}) & \dots & g'_m(\bar{x}) \\ \bar{\mu}_1 g'_1(\bar{x})^\top & g_1(\bar{x}) & 0 & \dots & 0 \\ \bar{\mu}_2 g'_2(\bar{x})^\top & 0 & g_2(\bar{x}) & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \bar{\mu}_m g'_m(\bar{x})^\top & 0 & \dots & 0 & g_m(\bar{x}) \end{bmatrix}, \tag{3.17}$$

and

$$N \in \operatorname{conv}\left\{ \lim_{l\to\infty} \Phi'_x(x^l;y^l,\mu^l) \mid (x^l;y^l,\mu^l) \to (\bar{x};\bar{x},\bar{\mu}), (x^l;y^l,\mu^l) \in \mathcal{D}_\Phi \right\}.$$

To apply Theorem 3.1.1, it remains to show that M is nonsingular. Suppose that $M\begin{bmatrix} v \\ w \end{bmatrix} = 0$, where $v \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$. Then we have

$$\Psi'_x(\bar{x},\bar{\mu})v + \sum_{i=1}^m w_i g'_i(\bar{x}) = 0, \tag{3.18}$$

$$\bar{\mu}_i\langle g'_i(\bar{x}), v\rangle + w_i g_i(\bar{x}) = 0, \quad i = 1,\dots,m. \tag{3.19}$$

Since $g_i(\bar{x}) < 0$ and $\bar{\mu}_i = 0$ for all $i \notin \mathcal{I}$, and by the strict complementarity condition (3.15), $g_i(\bar{x}) = 0$ and $\bar{\mu}_i > 0$ for all $i \in \mathcal{I}$, it follows from (3.19) that

$$\begin{aligned} \langle g'_i(\bar{x}), v\rangle &= 0, \quad \forall\, i \in \mathcal{I}, \\ w_i &= 0, \quad \forall\, i \notin \mathcal{I}. \end{aligned} \tag{3.20}$$

Since strict complementarity means that $\mathcal{I}_0 = \emptyset$, from (3.14) and (3.20) we have that $v \in \mathcal{C}(\bar{x}; D, F)$. Multiplying both sides in (3.18) by $v^\top$, we obtain

$$\begin{aligned} 0 &= \langle \Psi'_x(\bar{x},\bar{\mu})v, v\rangle + \sum_{i\in\mathcal{I}} w_i\langle g'_i(\bar{x}), v\rangle + \sum_{i\notin\mathcal{I}} w_i\langle g'_i(\bar{x}), v\rangle \\ &= \langle \Psi'_x(\bar{x},\bar{\mu})v, v\rangle, \end{aligned}$$

where the second equality holda by (3.20). Since $v \in \mathcal{C}(\bar{x}; D, F)$, SOC (3.13) implies that $v = 0$. Now by (3.18) and (3.20), using also that $v = 0$, we obtain that

$$0 = \sum_{i\in\mathcal{I}} w_i g'_i(\bar{x}).$$

Then LICQ (3.12) implies that $w_i = 0$ for all $i \in \mathcal{I}$. Taking into account (3.20), we conclude that $w = 0$, so that $(v, w) = 0$. Hence, M is nonsingular.

Then, by Theorem 3.1.1, there exist a neighborhood $\mathcal{U}_0$ of $\bar{x}$ in $\mathbb{R}^n$, a neighborhood Ω_0 of $(\bar{x},\bar{\mu})$ in $\mathbb{R}^n \times \mathbb{R}^m$, and a Lipschitz continuous function $\xi : \mathcal{U}_0 \to \Omega_0$ such that $\Phi(x;\xi(x)) = 0$ for all $x \in \mathcal{U}_0$, where $\xi(x) = (y(x), \mu(x))$ and $\xi(\bar{x}) = (\bar{x},\bar{\mu})$.

Furthermore, ξ is unique in the sense that if $\hat{x} \in \mathcal{U}_0$, $(\hat{y}, \hat{\mu}) \in \Omega_0$ and $\Phi(\hat{x}; \hat{y}, \hat{\mu}) = 0$, then $(\hat{y}, \hat{\mu}) = \xi(\hat{x})$.

Using the continuity of y and μ at $\bar{x}$ and the strict complementarity condition (3.15), it follows that the sets

$$\mathcal{U}_1 = \{x \in \mathcal{U}_0 \mid g_i(x) + \langle g_i'(x), y(x) - x\rangle + \frac{1}{2}\langle g_i''(x)(y(x) - x), y(x) - x\rangle < 0, \forall i \notin \mathcal{I}\},$$

$$\mathcal{U}_2 = \{x \in \mathcal{U}_0 \mid \mu_i(x) > 0, \forall i \in \mathcal{I}\},$$

are nonempty and open (and they contain $\bar{x}$). Furthermore, since Ω_0 is a neighborhood of $(\bar{x}, \bar{\mu})$, there exist a neighborhood $\mathcal{W}$ of $\bar{x}$ in $\mathbb{R}^n$ and a neighborhood $\mathcal{V}$ of $\bar{\mu}$ in $\mathbb{R}^m$ such that $\mathcal{W} \times \mathcal{V} \subset \Omega_0$. Let

$$\mathcal{U}_3 = \{x \in \mathcal{U}_1 \cap \mathcal{U}_2 \mid \xi(x) \in \mathcal{W} \times \mathcal{V}\}.$$

If $x \in \mathcal{U}_3$, then $(y(x), \mu(x)) \in \mathcal{W} \times \mathcal{V}$ and since $\Phi(x; \xi(x)) = 0$, using the definitions of $\mathcal{U}_1$ and $\mathcal{U}_2$, we conclude that

$$\begin{aligned} 0 &= g_i(x) + \langle g_i'(x), y(x) - x\rangle + \frac{1}{2}\langle g_i''(x)(y(x) - x), y(x) - x\rangle, \forall\, i \in \mathcal{I}, \\ 0 &= \mu_i(x), \forall\, i \notin \mathcal{I}. \end{aligned} \tag{3.21}$$

Now, combining $\Phi(x; \xi(x)) = 0$ with (3.21) and with the definitions of $\mathcal{U}_1$ and $\mathcal{U}_2$, we obtain that $(y(x), \mu(x))$ is a solution of the mixed complementarity problem (3.8)-(3.11).

Now let $x^k \in \mathcal{U}_3$, $k \geq 0$. We next show that if x^{k+1} is the closest to x^k solution of (3.8)-(3.11) and μ^{k+1} is the associated multiplier, then these are uniquely defined by $x^{k+1} = y(x^k)$ and $\mu^{k+1} = \mu(x^k)$. First, note that the gradients of constraints in (3.9), which are active at $y(x^k)$ form the set $\{g_i'(x^k) + g_i''(x^k)(y(x^k) - x^k) \mid i \in \mathcal{I}\}$. For x^k sufficiently close to $\bar{x}$, this is a small perturbation of the linearly independent set in the LICQ condition (3.12). Thus, it is linearly independent itself, which implies that $\mu(x^k)$ is in fact the unique multiplier associated to $y(x^k)$. Taking $\mathcal{U}_0$ sufficiently small (so that $\mathcal{U}_3$ is sufficiently small), it can also be seen that the closest to x^k solution (among all the solutions of (3.8)-(3.11)) is precisely $y(x^k)$, since it is the only solution in $\mathcal{W}$. From now on, $x^k \in \mathcal{U}_3$, $x^{k+1} = y(x^k)$ and $\mu^{k+1} = \mu(x^k)$.

By (3.8), we have that

$$\begin{aligned} 0 &= F(x^k) + g'(x^k)^\top \mu^{k+1} + \left(F'(x^k) + \sum_{i=1}^m \mu_i^{k+1} g_i''(x^k)\right)(x^{k+1} - x^k) \\ &= F(x^k) + g_{\mathcal{I}}'(x^k)^\top \mu_{\mathcal{I}}^k + \left(F'(x^k) + \sum_{i\in\mathcal{I}} \mu_i^k g_i''(x^k)\right)(x^{k+1} - x^k) \\ &\quad + g_{\mathcal{I}}'(x^k)^\top(\mu_{\mathcal{I}}^{k+1} - \mu_{\mathcal{I}}^k) + \sum_{i\in\mathcal{I}}(\mu_i^{k+1} - \mu_i^k) g_i''(x^k)(x^{k+1} - x^k), \end{aligned} \tag{3.22}$$

where we have taken into account that $\mu_i^{k+1} = 0$ for all $i \notin \mathcal{I}$.

By (3.21), we also have that

$$0 = g_i(x^k) + \langle g_i'(x^k), x^{k+1} - x^k \rangle + \frac{1}{2} \langle g_i''(x^k)(x^{k+1} - x^k), x^{k+1} - x^k \rangle, \forall\, i \in \mathcal{I}. \quad (3.23)$$

Defining

$$H : \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \to \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|}, \quad H(z) = \begin{bmatrix} F(x) + g_{\mathcal{I}}'(x)^\top \mu_{\mathcal{I}} \\ g_{\mathcal{I}}(x) \end{bmatrix}, \quad z = (x, \mu_{\mathcal{I}}),$$

relations (3.22) and (3.23) can be written as

$$0 = H(z^k) + H'(z^k)(z^{k+1} - z^k) + E_{k,k+1}, \quad (3.24)$$

where

$$E_{k,k+1} = \begin{bmatrix} \sum_{i \in \mathcal{I}} (\mu_i^{k+1} - \mu_i^k) g_i''(x^k)(x^{k+1} - x^k) \\ \frac{1}{2} \langle g_i''(x^k)(x^{k+1} - x^k), x^{k+1} - x^k \rangle, i \in \mathcal{I} \end{bmatrix}.$$

Note that (3.24) is not a Newton equation, as it is not linear with respect to z^{k+1}. However, we shall relate it, a posteriori, to a specially perturbed Newton type iterative process. The rest of the proof makes this precise and establishes the quadratic rate of convergence.

First, note that $H(\bar{z}) = 0$. By a proof similar to that for the nonsingularity of the matrix M defined in (3.17), it can be seen that the matrix

$$H'(\bar{z}) = \begin{bmatrix} \Psi_x'(\bar{x}, \bar{\mu}) & g_{\mathcal{I}}'(\bar{x})^\top \\ g_{\mathcal{I}}'(\bar{x}) & 0 \end{bmatrix}$$

is nonsingular (in the above formula for $H'(\bar{z})$, we have used the fact that $\bar{\mu}_i = 0$ for all $i \notin \mathcal{I}$). Since $H'(\bar{z})$ is nonsingular, there exists a constant $\eta > 0$ such that

$$\bar{z} \in \tilde{\mathcal{U}}_4 = \{ z \in \mathbb{R}^{n+|\mathcal{I}|} \mid \|H'(z)^{-1}\| < \eta \}.$$

Since F' and $g_i'', i = 1, \ldots, m$, are Lipschitz continuous functions in a neighborhood of $\bar{x}$, taking $\rho > 0$ sufficiently small, there exists a constant $L > 0$ such that $\|H'(w) - H'(z)\| \leq L \|w - z\|$ for all $w, z \in \bar{z} + \rho B$.

We next show that if $z^k \in \bar{z} + \rho B$ then there exists a constant $c > 0$ such that $\|E_{k,k+1}\| \leq c \|z^{k+1} - z^k\|^2$ for all $k \geq 1$, where $z^k = (x^k, \mu_{\mathcal{I}}^k)$. Since $g_i'', i = 1, \ldots, m$, are continuous at $\bar{x}$, there exists a constant $\gamma > 0$ such that $\|g_i''(x)\| \leq \gamma, i = 1, \ldots, m$, for all $x \in \mathbb{R}^n$ such that $\|x - \bar{x}\| \leq \rho$. Since $z^k \in \bar{z} + \rho B$ implies $\|x^k - \bar{x}\| < \rho$, we have that

$$\|E_{k,k+1}\| \quad \leq \quad \sqrt{n} \gamma \|\mu_{\mathcal{I}}^{k+1} - \mu_{\mathcal{I}}^k\| \, \|x^{k+1} - x^k\| + \frac{\gamma}{2} \sum_{i \in \mathcal{I}} \|x^{k+1} - x^k\|^2$$

$$\begin{aligned}
&\le \sqrt{n}\gamma(\max\{\|\mu_{\mathcal{I}}^{k+1}-\mu_{\mathcal{I}}^{k}\|\,,\,\|x^{k+1}-x^k\|\})^2+\frac{\gamma m}{2}\|x^{k+1}-x^k\|^2\\
&\le \sqrt{n}\gamma\|z^{k+1}-z^k\|^2+\frac{\gamma m}{2}\|x^{k+1}-x^k\|^2\\
&\le \gamma(\sqrt{n}+m/2)\|z^{k+1}-z^k\|^2\\
&= c\|z^{k+1}-z^k\|^2, \qquad (3.25)
\end{aligned}$$

where the monotonicity of the norm has been used repeatedly.

Let $r=1/(2\eta(L+4c))$, and define $\mathcal{U}_5=\{x\in\mathcal{U}_3 \mid \|y(x)-\bar{x}\|^2+\|\mu(x)-\bar{\mu}\|^2<r^2\}$, $\tilde{\mathcal{U}_5}=\mathcal{U}_5\times\mathbb{R}^{|\mathcal{I}|}$. Then there exists $\delta>0$ such that $\bar{z}+\delta B\subset\tilde{\mathcal{U}_4}\cap\tilde{\mathcal{U}_5}$.

Let $\varepsilon=\min\{\delta,r,\rho\}$, and define

$$\mathcal{U}=\{x\in\mathbb{R}^n \mid \|y(x)-\bar{x}\|^2+\|\mu(x)-\bar{\mu}\|^2<\varepsilon^2\}.$$

Then $x^0\in\mathcal{U}$ implies that $\|z^1-\bar{z}\|<\varepsilon$.

Now, proceeding by induction, we will show that if $\|z^k-\bar{z}\|<\varepsilon$ then $\|z^{k+1}-\bar{z}\|<\varepsilon$. By the construction of the set $\mathcal{U}$, if $\|z^k-\bar{z}\|<\varepsilon$ then the following properties hold:

$$\|H'(z^k)^{-1}\|<\eta, \qquad (3.26)$$

$$\|z^k-\bar{z}\|<r=\frac{1}{2\eta(L+4c)}, \qquad (3.27)$$

$$\|z^{k+1}-\bar{z}\|<r<\frac{1}{4\eta c}, \qquad (3.28)$$

where in (3.26) we use that $z^k\in\tilde{\mathcal{U}_4}$, (3.27) holds since $\varepsilon\le r$, and (3.28) follows from $x^k\in\mathcal{U}_5$. Also, because $x^k\in\mathcal{U}_3$, by (3.24) it follows that

$$z^{k+1}=z^k-H'(z^k)^{-1}(H(z^k)-H(\bar{z})+E_{k,k+1}),$$

where $H(\bar{z})=0$ was also taken into account. We further obtain

$$\begin{aligned}
\|z^{k+1}-\bar{z}\| &= \|z^k-\bar{z}-H'(z^k)^{-1}(H(z^k)-H(\bar{z})+E_{k,k+1})\|\\
&\le \|H'(z^k)^{-1}\|\,\|H'(z^k)(z^k-\bar{z})+H(\bar{z})-H(z^k)-E_{k,k+1}\|\\
&\le \eta\left\|\int_0^1[H'(z^k)-H'(\bar{z}+t(z^k-\bar{z}))](z^k-\bar{z})dt-E_{k,k+1}\right\|\\
&\le \eta\left(L\|z^k-\bar{z}\|^2\int_0^1(1-t)dt+c\|z^{k+1}-z^k\|^2\right)\\
&\le \frac{\eta L}{2}\|z^k-\bar{z}\|^2+2\eta c\|z^{k+1}-\bar{z}\|^2+2\eta c\|z^k-\bar{z}\|^2\\
&\le \eta\left(\frac{L}{2}+2c\right)\|z^k-\bar{z}\|^2+\frac{1}{2}\|z^{k+1}-\bar{z}\|,
\end{aligned}$$

where the second inequality follows from (3.26) and the Mean-Value Theorem, in the third inequality we use the Lipschitz continuity of H' and (3.25), for the fourth inequality we use that $\|z^{k+1}-z^k\|^2\le 2(\|z^{k+1}-\bar{z}\|^2+\|z^k-\bar{z}\|^2)$, and the fifth inequality follows from $2\eta c\|z^{k+1}-\bar{z}\|<\frac{1}{2}$, which is ensured by (3.28).

Now, rearranging terms in the relation above, we deduce that

$$\|z^{k+1} - \bar{z}\| \leq \eta(L + 4c)\|z^k - \bar{z}\|^2. \tag{3.29}$$

Then, by (3.27), we have $\|z^{k+1} - \bar{z}\| < \frac{1}{2}\|z^k - \bar{z}\| < \varepsilon$.

In consequence, if $x^0 \in \mathcal{U}$ then $(x^{k+1}, \mu^{k+1}) \in \mathcal{U} \times \mathcal{V}$ for all $k \geq 0$, and since $\mu_i^{k+1} = \bar{\mu}_i = 0$ for all $i \notin \mathcal{I}$, we have

$$\|(x^{k+1}, \mu^{k+1}) - (\bar{x}, \bar{\mu})\| = \|z^{k+1} - \bar{z}\| < \frac{1}{2}\|z^k - \bar{z}\| < \ldots < \left(\frac{1}{2}\right)^k \|z^1 - \bar{z}\|,$$

so that $\{(x^{k+1}, \mu^{k+1})\}$ converges to $(\bar{x}, \bar{\mu})$. Then, by (3.29), we conclude that the rate of convergence is quadratic.

■

3.3 Primal superlinear convergence

Recall that quadratic convergence of $\{(x^k, \mu^k)\}$ to $(\bar{x}, \bar{\mu})$ does not imply even superlinear or linear convergence of $\{x^k\}$ to $\bar{x}$. Assuming that some type of convergence occurs, we next give necessary and sufficient conditions for superlinear convergence of the primal sequence. This condition is of the Dennis-Moré type [8], allowing for using approximations of derivatives. Specific update rules of quasi-Newton type are certainly of great interest, yet this is beyond the scope of this work. But we note that our analysis covers those situations where computing the derivatives involves computational work and the precision of approximation can be controlled. Such is the case, for example, when the derivatives are approximated by finite-difference procedures. The accuracy parameter can be controlled using estimates for the distance to the solution via error bounds (see [10] for a discussion of error bounds for variational problems and [17, 7] for detailed comparisons in the context of KKT systems specifically). In particular, such estimates give some idea of how precise should be the approximation in order to conform to conditions (3.36) or (3.37) below.

Let H_k be some approximation of $F'(x^k)$ and $G_{i,k}$ be some approximation of $g_i''(x^k)$ (of course, this includes the possibility of exact derivatives, as in the setting of Section 3.2). We consider a sequence $\{(x^k, \mu^k)\}$ generated by the following process. Given $x^k \in \mathbb{R}^n$, $H_k \in \mathbb{R}^{n\times n}$ and $G_{i,k} \in \mathbb{R}^{n\times n}, i = 1, \ldots, m$, find $(x^{k+1}, \mu^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that:

$$F(x^k) + H_k(x^{k+1} - x^k) + \sum_{i=1}^{m} \mu_i^{k+1}(g_i'(x^k) + G_{i,k}(x^{k+1} - x^k)) = 0, \tag{3.30}$$

and for all $i = 1, \ldots, m,$ it holds that

$$g_i(x^k) + \langle g_i'(x^k), x^{k+1} - x^k\rangle + \frac{1}{2}\langle G_{i,k}(x^{k+1} - x^k), x^{k+1} - x^k\rangle \leq 0, \tag{3.31}$$

$$\mu_i^{k+1} \geq 0, \tag{3.32}$$

$$\mu_i^{k+1}(g_i(x^k) + \langle g_i'(x^k), x^{k+1} - x^k\rangle + \frac{1}{2}\langle G_{i,k}(x^{k+1} - x^k), x^{k+1} - x^k\rangle) = 0. \tag{3.33}$$

In the sequel, we shall consider separately the two possible cases in SOC (3.13) (i.e., SOC$^+$, when (3.13) holds with the positive sign and SOC$^-$, when it holds with the negative sign). Note also that since the cone $\mathcal{C}(\bar{x}; D, F)$ is closed, those two cases can be stated as follows: there exists $t > 0$ such that

$$\langle \Psi'_x(\bar{x}, \bar{\mu})u,\, u\rangle \geq t\|u\|^2 \qquad \forall\, u \in \mathcal{C}(\bar{x}; D, F), \tag{3.34}$$

and

$$-\langle \Psi'_x(\bar{x}, \bar{\mu})u,\, u\rangle \geq t\|u\|^2 \qquad \forall\, u \in \mathcal{C}(\bar{x}; D, F). \tag{3.35}$$

Theorem 3.3.1 *Let* $(\bar{x}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m$ *be a solution of the KKT system* (3.7)*. Suppose that* F *is differentiable and* g *is twice differentiable in some neighborhood of* $\bar{x}$*. Suppose further that a sequence* $\{(x^k, \mu^k)\}$*, generated according to* (3.30)-(3.33) *with uniformly bounded* $G_{i,k}$*,* $i = 1, \ldots, m$*, converges to* $(\bar{x}, \bar{\mu})$*.*

If $\{x^k\}$ *converges superlinearly to* $\bar{x}$ *then*

$$\Pi_{\mathcal{C}}\left[\left(\Psi'_x(\bar{x}, \bar{\mu}) - M_k\right)\left(x^{k+1} - x^k\right)\right] = o(\|x^{k+1} - x^k\|), \tag{3.36}$$

where $\Pi_{\mathcal{C}}[\cdot]$ *denotes the orthogonal projector onto the cone* $\mathcal{C}(\bar{x}; D, F)$ *defined in* (3.14) *and*

$$M_k = H_k + \sum_{i=1}^{m} \mu_i^{k+1} G_{i,k}.$$

Conversely, if LICQ (3.12) *and SOC* (3.13) *are satisfied, then the rate of convergence of* $\{x^k\}$ *to* $\bar{x}$ *is superlinear if SOC*$^+$ (3.34) *and* (3.36) *hold, or if SOC*$^-$ (3.35) *holds and*

$$\Pi_{\mathcal{C}}\left[\left(M_k - \Psi'_x(\bar{x}, \bar{\mu})\right)\left(x^{k+1} - x^k\right)\right] = o(\|x^{k+1} - x^k\|). \tag{3.37}$$

Proof. Denote $d^k = x^{k+1} - x^k$. By (3.30), we have that

$$0 = F(x^k) + H_k d^k + \sum_{i=1}^{m} \mu_i^{k+1}(g_i'(x^k) + G_{i,k} d^k) = \Psi(x^k, \mu^{k+1}) + M_k d^k. \tag{3.38}$$

Also, we have

$$\begin{aligned}
\Psi(x^k, \bar{\mu}) &= \Psi(x^k, \mu^{k+1}) + \sum_{i=1}^{m}(\bar{\mu}_i - \mu_i^{k+1}) g_i'(x^k) \\
&= \Psi(x^k, \mu^{k+1}) + \sum_{i=1}^{m}(\bar{\mu}_i - \mu_i^{k+1}) g_i'(\bar{x}) + o(\|x^k - \bar{x}\|) \\
&= -M_k d^k + \sum_{i=1}^{m}(\bar{\mu}_i - \mu_i^{k+1}) g_i'(\bar{x}) + o(\|x^k - \bar{x}\|, \qquad (3.39)
\end{aligned}$$

where the last equality is by (3.38).

Suppose first that $\{x^k\}$ converges to $\bar{x}$ superlinearly, i.e., $x^{k+1} - \bar{x} = o(\|x^k - \bar{x}\|)$. Since $\Psi(\bar{x}, \bar{\mu}) = 0$, it holds that

$$\begin{aligned} \Psi(x^k, \bar{\mu}) &= \Psi(\bar{x}, \bar{\mu}) + \Psi'_x(\bar{x}, \bar{\mu})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= -\Psi'_x(\bar{x}, \bar{\mu})d^k + \Psi'_x(\bar{x}, \bar{\mu})(x^{k+1} - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= -\Psi'_x(\bar{x}, \bar{\mu})d^k + o(\|x^k - \bar{x}\|). \end{aligned} \tag{3.40}$$

Combining (3.40) and (3.39), we obtain

$$(\Psi'_x(\bar{x}, \bar{\mu}) - M_k)\, d^k = \sum_{i=1}^{m} (\mu_i^{k+1} - \bar{\mu}_i) g'_i(\bar{x}) + o(\|x^k - \bar{x}\|). \tag{3.41}$$

Taking into account that $G_{i,k}$ are uniformly bounded and using the continuity argument in (3.33), we conclude that for all sufficiently large k, it holds that $\mu_i^{k+1} - \bar{\mu}_i = 0, \forall i \notin \mathcal{I}$, and $\mu_i^{k+1} - \bar{\mu}_i = \mu_i^{k+1} \geq 0, \forall\, i \in \mathcal{I}_0$. Then, by (3.41), for all $v \in \mathcal{C}(\bar{x}; D, F)$ it holds that

$$\begin{aligned} \Big\langle (\Psi'_x(\bar{x}, \bar{\mu}) - M_k)\, d^k - o(\|x^k - \bar{x}\|), v \Big\rangle &= \sum_{i=1}^{m} (\mu_i^{k+1} - \bar{\mu}_i) \langle g'_i(\bar{x}), v \rangle \\ &= \sum_{i \in \mathcal{I}} (\mu_i^{k+1} - \bar{\mu}_i) \langle g'_i(\bar{x}), v \rangle \\ &= \sum_{i \in \mathcal{I}_0} (\mu_i^{k+1} - \bar{\mu}_i) \langle g'_i(\bar{x}), v \rangle \\ &= \sum_{i \in \mathcal{I}_0} \mu_i^{k+1} \langle g'_i(\bar{x}), v \rangle \leq 0, \end{aligned} \tag{3.42}$$

where we have used that $\langle g'_i(\bar{x}), v \rangle = 0, \forall\, i \in \mathcal{I}_+,\ \langle g'_i(\bar{x}), v \rangle \leq 0, \forall\, i \in \mathcal{I}_0$ (see (3.14)). By properties of projection operator onto a convex cone, inequality (3.42) means that

$$\Pi_{\mathcal{C}} \left[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k) d^k - o(\|x^k - \bar{x}\|) \right] = 0.$$

Then, by nonexpansiveness of the projection operator, it follows that

$$\begin{aligned} \left\| \Pi_{\mathcal{C}} \left[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k) d^k \right] \right\| &= \Big\| \Pi_{\mathcal{C}} \left[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k) d^k - o(\|x^k - \bar{x}\|) \right] \\ &\quad - \Pi_{\mathcal{C}} \left[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k) d^k \right] \Big\| \\ &= o(\|x^k - \bar{x}\|). \end{aligned}$$

It remains to show that $o(\|x^k - \bar{x}\|) = o(\|d^k\|)$. For this, note that

$$\begin{aligned} \frac{o(\|x^k - \bar{x}\|)}{\|d^k\|} &\leq \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\| - \|x^{k+1} - \bar{x}\|} = \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\| - o(\|x^k - \bar{x}\|)} \\ &= \frac{o(\|x^k - \bar{x}\|)/\|x^k - \bar{x}\|}{1 - o(\|x^k - \bar{x}\|)/\|x^k - \bar{x}\|} \to 0 \text{ as } k \to \infty. \end{aligned}$$

This concludes the proof of (3.36).

We now prove the sufficiency part, assuming LICQ and SOC. Denote

$$\Gamma_{i,k} = g_i(x^k) + \langle g_i'(x^k), d^k\rangle + \frac{1}{2}\langle G_{i,k}d^k, d^k\rangle.$$

By the continuity argument (taking also into account uniform boundedness of $G_{i,k}$), $\{\Gamma_{i,k}\}$ converges to $g_i(\bar{x})$, as $k \to \infty$. Thus for all k sufficiently large, taking into account (3.33), we have that

$$\begin{aligned} &\Gamma_{i,k} < 0,\ \mu_i^{k+1} = 0, \quad \forall\, i \notin \mathcal{I}, \\ &\Gamma_{i,k} = 0,\ \mu_i^{k+1} > 0, \quad \forall\, i \in \mathcal{I}_+. \end{aligned} \tag{3.43}$$

By the Mean-Value Theorem, for each $i = 1, \ldots, m$, there exists a vector $z^{i,k}$ in the line segment joining x^k and $\bar{x}$, such that

$$g_i(x^k) = g_i(\bar{x}) + \langle g_i'(\bar{x}), x^k - \bar{x}\rangle + \frac{1}{2}\langle g_i''(z^{i,k})(x^k - \bar{x}), x^k - \bar{x}\rangle.$$

Note that $\{z^{i,k}\}$ converges to $\bar{x}$ when $k \to \infty$. For $i \in \mathcal{I}$, we then obtain

$$\begin{aligned} \Gamma_{i,k} &= g_i(\bar{x}) + \langle g_i'(\bar{x}), x^k - \bar{x}\rangle + \frac{1}{2}\langle g_i''(z^{i,k})(x^k - \bar{x}), x^k - \bar{x}\rangle + \langle g_i'(x^k), d^k\rangle + \frac{1}{2}\langle G_{i,k}d^k, d^k\rangle \\ &= \langle g_i'(\bar{x}), x^{k+1} - \bar{x}\rangle + w_i^k, \end{aligned} \tag{3.44}$$

where

$$w_i^k = \langle g_i'(x^k) - g_i'(\bar{x}), d^k\rangle + \frac{1}{2}\langle g_i''(z^{i,k})(x^k - \bar{x}), x^k - \bar{x}\rangle + \frac{1}{2}\langle G_{i,k}d^k, d^k\rangle.$$

Clearly,

$$w_i^k = o(\|x^k - \bar{x}\|) + o(\|d^k\|).$$

By LICQ (3.12), for each k, there exists $u^k \in \mathbb{R}^n$ such that

$$g_{\mathcal{I}}'(\bar{x})u^k = w_{\mathcal{I}}^k, \quad \text{where} \quad u^k = o(\|x^k - \bar{x}\|) + o(\|d^k\|). \tag{3.45}$$

Let $v^k = x^{k+1} - \bar{x} + u^k$. Then by (3.45) and (3.44), we have

$$\langle g_i'(\bar{x}), v^k\rangle = \langle g_i'(\bar{x}), x^{k+1} - \bar{x}\rangle + w_i^k = \Gamma_{i,k} \quad \forall\, i \in \mathcal{I}. \tag{3.46}$$

Since $\Gamma_{i,k} = 0, \forall\, i \in \mathcal{I}_+$ (by (3.43)) and $\Gamma_{i,k} \le 0, \forall\, i \in \mathcal{I}_0$ (by (3.31)), relation (3.46) shows that $v^k \in \mathcal{C}(\bar{x}; D, F)$. Since $\Psi(\bar{x}, \bar{\mu}) = 0$, we have that

$$0 = \langle \Psi(\bar{x}, \bar{\mu}), v^k\rangle = \langle F(\bar{x}), v^k\rangle + \sum_{i \in \mathcal{I}_+} \bar{\mu}_i \langle g_i'(\bar{x}), v^k\rangle = \langle F(\bar{x}), v^k\rangle.$$

We then obtain

$$\begin{aligned}
\langle \Psi(\bar{x}, \mu^{k+1}), v^k \rangle &= \langle F(\bar{x}), v^k \rangle + \sum_{i=1}^{m} \mu_i^{k+1} \langle g_i'(\bar{x}), v^k \rangle \\
&= \sum_{i \notin \mathcal{I}} \mu_i^{k+1} \langle g_i'(\bar{x}), v^k \rangle + \sum_{i \in \mathcal{I}} \mu_i^{k+1} \Gamma_{i,k} \\
&= 0, \qquad (3.47)
\end{aligned}$$

where we have used (3.46) for the second equality, and (3.43) with (3.33) for the last equality.

Also,

$$\begin{aligned}
\Psi(x^{k+1}, \mu^{k+1}) &= \Psi(x^k, \mu^{k+1}) + \Psi_x'(x^k, \mu^{k+1}) d^k + o(\|d^k\|) \\
&= (\Psi_x'(x^k, \mu^{k+1}) - M_k) d^k + o(\|d^k\|) \\
&= (\Psi_x'(\bar{x}, \bar{\mu}) - M_k) d^k + (\Psi_x'(x^k, \mu^{k+1}) - \Psi_x'(\bar{x}, \bar{\mu})) d^k + o(\|d^k\|) \\
&= (\Psi_x'(\bar{x}, \bar{\mu}) - M_k) d^k + o(\|d^k\|), \qquad (3.48)
\end{aligned}$$

where (3.38) has been used in the second equality, and the last equality is by the continuity of Ψ_x'. Let $p^k = v^k / \|v^k\|$. Multiplying both sides in (3.48) by p^k (which is bounded), we conclude that

$$\left\langle \Psi(x^{k+1}, \mu^{k+1}), p^k \right\rangle = \left\langle (\Psi_x'(\bar{x}, \bar{\mu}) - M_k) d^k, p^k \right\rangle + o(\|d^k\|). \qquad (3.49)$$

On the other hand,

$$\begin{aligned}
\left\langle \Psi(x^{k+1}, \mu^{k+1}), p^k \right\rangle &= \left\langle \Psi(\bar{x}, \mu^{k+1}), p^k \right\rangle + \left\langle \Psi_x'(\bar{x}, \mu^{k+1})(x^{k+1} - \bar{x}), p^k \right\rangle + o(\|x^{k+1} - \bar{x}\|) \\
&= \left\langle \Psi_x'(\bar{x}, \mu^{k+1})(x^{k+1} - \bar{x}), p^k \right\rangle + o(\|x^{k+1} - \bar{x}\|) \\
&= \left\langle \Psi_x'(\bar{x}, \bar{\mu})(x^{k+1} - \bar{x}), p^k \right\rangle + o(\|x^{k+1} - \bar{x}\|), \qquad (3.50)
\end{aligned}$$

where the second equality follows from (3.47), and the last follows from the continuity of Ψ_x' and boundedness of $\{p^k\}$.

Combining (3.49) and (3.50), we conclude that

$$\left\langle \Psi_x'(\bar{x}, \bar{\mu})(x^{k+1} - \bar{x}), p^k \right\rangle = \left\langle (\Psi_x'(\bar{x}, \bar{\mu}) - M_k) d^k, p^k \right\rangle + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|). \qquad (3.51)$$

Suppose now that SOC holds. Then for the case (3.35) and (3.37), by (3.51) and (3.45), we have

$$\begin{aligned}
t\|v^k\| &\le -\langle \Psi_x'(\bar{x}, \bar{\mu}) v^k, p^k \rangle \\
&= \left\langle \Psi_x'(\bar{x}, \bar{\mu})(\bar{x} - x^{k+1}), p^k \right\rangle - \left\langle \Psi_x'(\bar{x}, \bar{\mu}) u^k, p^k \right\rangle \\
&= \left\langle (M_k - \Psi_x'(\bar{x}, \bar{\mu})) d^k, p^k \right\rangle + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) \\
&\le \left\langle \Pi_{\mathcal{C}} \left[(M_k - \Psi_x'(\bar{x}, \bar{\mu})) d^k \right], p^k \right\rangle + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) \\
&= o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|),
\end{aligned}$$

where the second inequality follows from the fact that for any closed convex cone K and $v \in K$, it holds that $\langle x, v\rangle \leq \langle \Pi_K[x], v\rangle\, \forall x$. Similarly, for the case (3.34) and (3.36) we obtain

$$\begin{aligned} t\|v^k\| &\leq \langle \Psi'_x(\bar{x},\bar{\mu})v^k, p^k\rangle \\ &= \left\langle (\Psi'_x(\bar{x},\bar{\mu}) - M_k)d^k, p^k \right\rangle + o(\|d^k\|) + o(\|x^{k+1}-\bar{x}\|) + o(\|x^k-\bar{x}\|) \\ &\leq \left\langle \Pi_{\mathcal{C}}\left[(\Psi'_x(\bar{x},\bar{\mu}) - M_k)d^k\right], p^k \right\rangle + o(\|d^k\|) + o(\|x^{k+1}-\bar{x}\|) + o(\|x^k-\bar{x}\|) \\ &= o(\|d^k\|) + o(\|x^{k+1}-\bar{x}\|) + o(\|x^k-\bar{x}\|). \end{aligned}$$

Summarizing, in both cases $v^k = o(\|d^k\|) + o(\|x^{k+1}-\bar{x}\|) + o(\|x^k-\bar{x}\|)$.

Since $x^{k+1}-\bar{x} = v^k - u^k = o(\|x^{k+1}-\bar{x}\|) + o(\|x^k-\bar{x}\|) + o(\|d^k\|)$, there exists a sequence $\{t_k\}$ converging to 0 such that

$$\begin{aligned} \|x^{k+1}-\bar{x}\| &\leq t_k \left(\|x^{k+1}-\bar{x}\| + \|x^k-\bar{x}\| + \|d^k\|\right) \\ &\leq 2t_k \left(\|x^{k+1}-\bar{x}\| + \|x^k-\bar{x}\|\right). \end{aligned} \tag{3.52}$$

Since $t_k < 1/2$ for k sufficiently large, rearranging terms in (3.52) we obtain

$$\frac{\|x^{k+1}-\bar{x}\|}{\|x^k-\bar{x}\|} \leq \frac{2t_k}{1-2t_k} \to 0 \text{ as } k \to \infty.$$

In consequence, $x^{k+1}-\bar{x} = o(\|x^k-\bar{x}\|)$, i.e., $\{x^k\}$ converges superlinearly to $\bar{x}$. ■

In particular, Theorem 3.3.1 shows superlinear convergence of the primal sequence $\{x^k\}$ to $\bar{x}$ in the setting of Theorem 3.2.1, where $H_k = F'(x^k)$, $G_{i,k} = g''_i(x^k), i = 1,\ldots,m$, so that $M_k = H_k + \sum_{i=1}^m \mu_i^{k+1} G_{i,k} \to F'(\bar{x}) + \sum_{i=1}^m \bar{\mu}_i g''_i(\bar{x}) = \Psi'_x(\bar{x},\bar{\mu})$ as $k \to \infty$. In this case, conditions (3.36) and (3.37) are automatically satisfied.

We also note that in the setting of Theorem 3.2.1 (or more generally, when the cone $\mathcal{C}(\bar{x}; D, F)$ is a subspace), we do not have to consider separately the two cases of SOC (SOC$^+$ (3.34) and SOC$^-$ (3.35)) neither the two cases of the Dennis-Moré condition ((3.36) and (3.37)). Indeed, when $\mathcal{C}(\bar{x}; D, F)$ is a subspace, we have $\langle x, v\rangle = \langle \Pi_{\mathcal{C}}[x], v\rangle$ for all $v \in \mathcal{C}(\bar{x}; D, F)$. We can further state the SOC (3.13) as

$$|\langle \Psi'_x(\bar{x},\bar{\mu})v, v\rangle| \geq t\|v\|^2 \quad \forall\, v \in \mathcal{C}(\bar{x}; D, F),$$

and modify the corresponding parts of the proof of Theorem 3.3.1, as follows.

For the necessary part, note that for any $x \in \mathbb{R}^n$, there exists the unique decomposition $x = v + v^*$ with $v = \Pi_{\mathcal{C}}[x] \in \mathcal{C}(\bar{x}; D, F)$ and $v^* \in \mathcal{C}(\bar{x}; D, F)^{\perp}$. Clearly, changing the sign, one has $-x = -v - v^*$, where $-v = \Pi_{\mathcal{C}}[-x] \in \mathcal{C}$ and $-v^* \in \mathcal{C}^{\perp}$. Hence, $\|\Pi_{\mathcal{C}}[x]\| = \|\Pi_{\mathcal{C}}[-x]\|$ for any $x \in \mathbb{R}^n$. It follows that in this case, conditions (3.36) and (3.37) are equivalent.

For the sufficient part, we have that

$$\begin{aligned}
t\|v^k\| &\leq \left|\langle \Psi'_x(\bar{x},\bar{\mu})v^k, p^k\rangle\right| \\
&\leq \left|\left\langle (\Psi'_x(\bar{x},\bar{\mu}) - M_k)d^k, p^k\right\rangle\right| + o(\|d^k\|) + o(\|x^{k+1}-\bar{x}\|) + o(\|x^k-\bar{x}\|) \\
&= \left|\left\langle \Pi_{\mathcal{C}}\left[(\Psi'_x(\bar{x},\bar{\mu}) - M_k)d^k\right], p^k\right\rangle\right| + o(\|d^k\|) + o(\|x^{k+1}-\bar{x}\|) + o(\|x^k-\bar{x}\|) \\
&= o(\|d^k\|) + o(\|x^{k+1}-\bar{x}\|) + o(\|x^k-\bar{x}\|),
\end{aligned}$$

and the rest of the proof applies.

3.4 Concluding remarks

We have established a new result on the quadratic convergence of the primal-dual sequence of the sequential quadratically-constrained quadratic-programming method. A necessary and sufficient characterization of the superlinear convergence of the primal sequence has also been provided. Additionally, the class of methods under consideration has been extended from the optimization setting to the more general variational problems.

Bibliography

[1] M. Anitescu. A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming. *SIAM Journal on Optimization*, 12:949–978, 2002.

[2] B.T. Boggs and J.W. Tolle. Sequential quadratic programming. *Acta Numerica*, 4:1–51, 1996.

[3] J. F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization*, 29:161–186, 1994.

[4] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Numerical Optimization: Theoretical and Practical Aspects*. Springer–Verlag, Berlin, Germany, 2006. Second Edition.

[5] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer–Verlag, New York, NY, 2000.

[6] F.H. Clarke. *Optimization and Nonsmooth Analysis*. SIAM Publications, Philadelphia, 1990.

[7] A.N. Daryina, A.F. Izmailov, and M.V. Solodov. A class of active-set Newton methods for mixed complementarity problems. *SIAM Journal on Optimization*, 15:409–429, 2004.

[8] J.E. Dennis and J.J. Moré. A characterization of superlinear convergence and its application to quasi-Newton methods. *Mathematics of Computation*, 28:549–560, 1974.

[9] A.L. Dontchev and R.T. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM Journal on Optimization*, 6(4):1087–1105, 1996.

[10] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer–Verlag, New York, NY, 2003.

[11] D. Fernández and M.V. Solodov. On local convergence of sequential quadratically-constrained quadratic-programming type methods, with an extension to variational problems. *Computational Optimization and Applications*, 39(2):143–160, 2008.

[12] D. Fernández and M.V. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems without constraint qualifications, 2007. Submitted to Mathematical Programming.

[13] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Mathematical Programming*, 94:91–124, 2002.

[14] M. Fukushima, Z.-Q. Luo, and P. Tseng. A sequential quadratically constrained quadratic programming method for differentiable convex minimization. *SIAM Journal on Optimization*, 13:1098–1119, 2003.

[15] W.W. Hager. Stabilized sequential quadratic programming. *Computational Optimization and Applications*, 12:253–273, 1999.

[16] W.W. Hager and M.S. Gowda. Stability in the presence of degeneracy and error estimation. *Mathematical Programming*, 85:181–192, 1999.

[17] A.F. Izmailov and M.V. Solodov. Karush-Kuhn-Tucker systems: regularity conditions, error bounds and a class of Newton-type methods. *Mathematical Programming*, 95:631–650, 2003.

[18] A.F. Izmailov and M.V. Solodov. Newton-type methods for optimization problems without constraint qualifications. *SIAM Journal on Optimization*, 16:210–228, 2004.

[19] A.F. Izmailov and M.V. Solodov. Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. *Computational Optimization and Applications*, 2007. DOI 10.1007/s10107-007-0158-9.

[20] A.F. Izmailov and M.V. Solodov. On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. *Mathematical Programming*, 2007. DOI 10.1007/s10107-007-0158-9.

[21] N.H. Josephy. Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin, 1979.

[22] S. Kruk and H. Wolkowicz. Sequential, quadratically constrained, quadratic programming for general nonlinear programming. In H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors, *Handbook of Semidefinite Programming*, pages 563–575. Kluwer Academic Publishers, 2000.

[23] D.-H. Li and L. Qi. A stabilized SQP method via linear equations. Technical Report Applied mathematics technical report AMR00/5, The University of New South Wales, 2000.

[24] M.S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret. Applications of second-order cone programming. *Linear Algebra and its Applications*, 284:193–228, 1998.

[25] N. Maratos. *Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization Problems*. PhD thesis, Imperial College, University of London, 1978.

[26] R.D.C. Monteiro and T. Tsuchiya. Polynomial convergence of primal-dual algorithms for the second-order cone programs based on the MZ-family of directions. *Mathematical Programming*, 88:61–83, 2000.

[27] B.S. Mordukhovich. *Variational analysis and generalized differentiation I*. Springer-Verlag, Berlin, 2006.

[28] Y.E. Nesterov and A.S. Nemirovskii. *Interior point polynomial methods in convex programming: theory and applications*. SIAM Publications, Philadelphia, PA, 1993.

[29] J. Nocedal and S.J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

[30] J.-S. Pang. Error bounds in mathematical programming. *Mathematical Programming*, 79:299–332, 1997.

[31] V.M. Panin. A second-order method for discrete min-max problem. *USSR Computational Mathematics and Mathematical Physics*, 19:90–100, 1979.

[32] V.M. Panin. Some methods of solving convex programming problems. *USSR Computational Mathematics and Mathematical Physics*, 21:57–72, 1981.

[33] M.J.D. Powell. Variable metric methods for constrained optimization. In A. Bachem, M. Grötschel, and B. Korte, editors, *Mathematical Programming: The State of Art*, pages 288–311. Springer–Verlag, Berlin, 1983.

[34] S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–62, 1980.

[35] S.M. Robinson. Some continuity properties of polyhedral multifunctions. *Mathematical Programming Study*, 14:206–214, 1981.

[36] S.M. Robinson. Generalized equations and their solutions II. Applications to nonlinear programming. *Mathematical Programming Study*, (19):200–221, 1982.

[37] R.T. Rockafellar and J.-B. Wets. *Variational Analysis*. Spronger-Verlag, New York, 1997.

[38] M.V. Solodov. On the sequential quadratically constrained quadratic programming methods. *Mathematics of Operations Research*, 29:64–79, 2004.

[39] T. Tsuchiya. A convergence analysis of the scale-invariant primal-dual path-following algorithms for second-order cone programming. *Optimization Methods and Software*, 11:141–182, 1999.

[40] E.J. Wiest and E. Polak. A generalized quadratic programming-based phase-I–phase-II method for inequality-constrained optimization. *Applied Mathematics and Optimization*, 26:223–252, 1992.

[41] S.J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Computational Optimization and Applications*, 11:253–275, 1998.

[42] S.J. Wright. Modifying SQP for degenerate problems. *SIAM Journal on Optimization*, 13:470–497, 2002.

[43] S.J. Wright. Constraint identification and algorithm stabilization for degenerate nonlinear programs. *Mathematical Programming*, 95(1, Ser. B):137–160, 2003.

[44] S.J. Wright. An algorithm for degenerate nonlinear programming with rapid local convergence. *SIAM Journal on Optimization*, 15 (2005), pp. 673–696.