

STABILIZED SEQUENTIAL QUADRATIC PROGRAMMING: A SURVEY*

Damián Fernández[†] and Mikhail Solodov[‡]

September 30, 2013

ABSTRACT

We review the motivation for, the current state-of-the-art in convergence results, and some open questions concerning the stabilized version of the sequential quadratic programming algorithm for constrained optimization. We also discuss the tools required for its local convergence analysis, globalization challenges, and extensions of the method to the more general variational problems.

Key words: Newton method, (stabilized) sequential quadratic programming, constrained optimization, variational problem, second-order sufficiency, (non)critical Lagrange multipliers.

AMS subject classifications: 90C30, 90C33, 90C55, 65K05.

* The second author is supported in part by CNPq Grant 302637/2011-7, by PRONEX–Optimization, and by FAPERJ.

[†] CIEM-CONICET, FaMAF, Universidad Nacional de Córdoba, Medina Allende s/n, 5000 Córdoba, Argentina.

Email: dfernandez@famaf.unc.edu.ar

[‡] IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.

Email: solodov@impa.br

1 Introduction

Consider the optimization problem

$$\begin{aligned} & \text{minimize}_x && f(x) \\ & \text{subject to} && h(x) = 0, \quad g(x) \leq 0, \end{aligned} \tag{1.1}$$

where $f : \mathbf{R}^n \rightarrow \mathbf{R}$, $h : \mathbf{R}^n \rightarrow \mathbf{R}^l$ and $g : \mathbf{R}^n \rightarrow \mathbf{R}^m$ are twice continuously differentiable. Let $L : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \rightarrow \mathbf{R}$,

$$L(x, \lambda, \mu) = f(x) + \langle \lambda, h(x) \rangle + \langle \mu, g(x) \rangle$$

be the Lagrangian of problem (1.1), where $\langle \cdot, \cdot \rangle$ stands for the inner product (the space is always clear from the context). Stationary points of problem (1.1) and the associated Lagrange multipliers are characterized by the Karush–Kuhn–Tucker (KKT) optimality system

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x) \rangle = 0. \tag{1.2}$$

We denote by $\mathcal{M}(\bar{x})$ the set of Lagrange multipliers associated with $\bar{x} \in \mathbf{R}^n$, that is, the pairs $(\lambda, \mu) \in \mathbf{R}^l \times \mathbf{R}^m$ satisfying (1.2) for $x = \bar{x}$.

The fundamental Newtonian approach to solving (1.1) is the sequential quadratic programming (SQP) algorithm [2, 17]. Given the current primal-dual iterate $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, an iteration of SQP generates the next iterate $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ as a stationary point and associated Lagrange multipliers of the quadratic programming (QP) subproblem

$$\begin{aligned} & \text{minimize}_x && \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle H_k(x - x^k), x - x^k \rangle \\ & \text{subject to} && h(x^k) + h'(x^k)(x - x^k) = 0, \\ & && g(x^k) + g'(x^k)(x - x^k) \leq 0, \end{aligned} \tag{1.3}$$

where H_k is a symmetric $n \times n$ matrix. The basic Newtonian scheme corresponds to taking

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k). \tag{1.4}$$

In fact, if there are no inequality constraints, it can be seen that computing (x^{k+1}, λ^{k+1}) by solving (1.3) with the choice (1.4) is equivalent to the usual Newton iteration from the point (x^k, λ^k) , applied to the equation given by the first two equalities in the KKT system (1.2).

To motivate the stabilized modification of SQP, we start with some comments about convergence properties of SQP itself. The first relevant observation is that without constraint qualification (CQ) [37] assumptions, the QP (1.3) can simply be infeasible and thus the method be not well-defined. Indeed, it can be informally stated that one of the roles of CQs is precisely to ensure that the first-order approximation of the constraints, like in (1.3), be consistent (and adequately approximate local structure of the feasible set of the original problem (1.1) around the given point).

For a point \bar{x} feasible in (1.1), denote by

$$A(\bar{x}) = \{i = 1, \dots, m \mid g_i(\bar{x}) = 0\}$$

the set of inequality constraints active at \bar{x} . If \bar{x} is a stationary point of (1.1) and $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$, denote further by

$$A_+(\bar{x}, \bar{\mu}) = \{i \in A(\bar{x}) \mid \bar{\mu}_i > 0\}, \quad A_0(\bar{x}, \bar{\mu}) = A(\bar{x}) \setminus A_+(\bar{x}, \bar{\mu})$$

the sets of strongly and weakly active constraints, respectively.

The linear independence constraint qualification (LICQ) is said to hold at \bar{x} if

$$\text{rank} \begin{pmatrix} h'(\bar{x}) \\ g'_{A(\bar{x})}(\bar{x}) \end{pmatrix} = l + |A(\bar{x})|, \quad (1.5)$$

where from now on, the notation M_J refers to the submatrix of the matrix M comprised by the rows of M indexed by the set J . In particular, the LICQ condition (1.5) says that the gradients of all of the equality constraints together with the gradients of all of the active inequality constraints form a linearly independent set in \mathbf{R}^n . The Mangasarian–Fromovitz constraint qualification (MFCQ) is said to hold at \bar{x} if

$$\text{rank } h'(\bar{x}) = l \text{ and } \exists \bar{\xi} \in \ker h'(\bar{x}) \text{ such that } g'_{A(\bar{x})}(\bar{x})\bar{\xi} < 0, \quad (1.6)$$

where for a matrix M we denote its null space by $\ker M = \{\xi \mid M\xi = 0\}$. Both LICQ and MFCQ imply that for a local solution \bar{x} of (1.1) the multiplier set $\mathcal{M}(\bar{x})$ is nonempty (for this specific property, weaker or different conditions can be used as well; see [37]). Note that MFCQ is equivalent to the requirement that $\mathcal{M}(\bar{x})$ be nonempty and bounded. The so-called strict MFCQ (SMFCQ) consists of saying that, in addition to (1.6), the multiplier associated to \bar{x} is unique ($\mathcal{M}(\bar{x})$ is a singleton). In the absence of (active) inequality constraints MFCQ, SMFCQ and LICQ are all equivalent (to the regularity condition $\text{rank } h'(\bar{x}) = l$), but otherwise MFCQ is a weaker assumption than SMFCQ which, in turn, is weaker than LICQ.

We say that for a given stationary point \bar{x} of problem (1.1) and for an associated multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ the second-order sufficient optimality condition (SOSC) holds if

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C(\bar{x}) \setminus \{0\}, \quad (1.7)$$

where

$$C(\bar{x}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0, \ g'_{A(\bar{x})}(\bar{x})\xi \leq 0, \ \langle f'(\bar{x}), \xi \rangle \leq 0\}$$

is the critical cone of problem (1.1) at \bar{x} . We note that SOSC implies that \bar{x} is a strict local minimizer in (1.1).

The sharpest local superlinear convergence result for SQP is provided by the analysis in [3]. It assumes SMFCQ (uniqueness of the multiplier $(\bar{\lambda}, \bar{\mu})$ associated to \bar{x}) and SOSC (1.7). Earlier results all required, in addition to SOSC, the stronger LICQ and the strict complementarity condition $\bar{\mu}_{A(\bar{x})} > 0$ (such statements are standard; see, e.g., [1, 33]). In particular, we emphasize that convergence of SQP requires certain regularity of constraints (a CQ).

The stabilized version of SQP (sSQP) had been developed with the goal to guarantee fast convergence rate despite possible degeneracy of constraints (i.e., when usual CQs may

not hold), and in particular when the Lagrange multipliers associated to a solution are not unique. The method was introduced in [38] for the case of inequality constraints, in the form of iteratively solving the min-max subproblems

$$\begin{aligned} \text{minimize}_{x \in \mathbf{R}^n} \quad & \max_{\mu \in \mathbf{R}_+^m} \left\{ \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \mu^k)(x - x^k), x - x^k \right\rangle \right. \\ & \left. + \langle \mu, g(x^k) + g'(x^k)(x - x^k) \rangle - \frac{\sigma_k}{2} \|\mu - \mu^k\|^2 \right\}, \end{aligned}$$

where $(x^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}_+^m$ is the current approximation to a primal-dual solution of (1.2), and $\sigma_k > 0$ is the dual stabilization parameter. Adding also equality constraints, it can be seen [28] that the corresponding min-max problem is equivalent to the following QP in the primal-dual space:

$$\begin{aligned} \text{minimize}_{(x, \lambda, \mu)} \quad & \left\{ \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k, \mu^k)(x - x^k), x - x^k \right\rangle \right. \\ & \left. + \frac{\sigma_k}{2} (\|\lambda\|^2 + \|\mu\|^2) \right\} \\ \text{subject to} \quad & h(x^k) + h'(x^k)(x - x^k) - \sigma_k(\lambda - \lambda^k) = 0, \\ & g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k) \leq 0. \end{aligned} \tag{1.8}$$

The dual stabilization parameter is usually based on computing the violation of the KKT optimality conditions (1.2) by the point (x^k, λ^k, μ^k) . For example, for fast local convergence one chooses in (1.8) $\sigma_k = \sigma(x^k, \lambda^k, \mu^k)$, where $\sigma : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \rightarrow \mathbf{R}_+$ is the natural residual of the KKT system (1.2), i.e.,

$$\sigma(x, \lambda, \mu) = \left\| \begin{pmatrix} \frac{\partial L}{\partial x}(x, \lambda, \mu) \\ h(x) \\ \min\{\mu, -g(x)\} \end{pmatrix} \right\|, \tag{1.9}$$

with the minimum applied componentwise.

One immediate observation is that for $\sigma_k > 0$, the constraints in (1.8) have the so-called “elastic mode” feature, and the subproblem is therefore automatically feasible regardless of any CQs or convexity assumptions. For example, fixing any $x \in \mathbf{R}^n$ and taking a suitable λ (uniquely defined for each x by the first constraint in (1.8)) and $\mu > 0$ with all the components large enough, gives points (x, λ, μ) feasible in (1.8). This is the first major difference from the standard SQP.

Another consideration is the following. It has been observed (e.g., in [40, Section 6], and in [21, 22, 24]) that the difficulties with convergence of SQP in degenerate cases are often not because of degeneracy as such, but are due to some undesirable behaviour of the dual sequence. The dual regularization/stabilization term in the objective function of (1.8) can be regarded as an attempt to modify this behaviour. As discussed in Section 2 below, in the sense of local convergence it indeed does the job. The situation is more complicated in the global sense; see Section 3.

2 Local convergence theory

In this section, we first survey the historical accounts on local convergence analyses of sSQP. Then, we state the current state-of-the-art results, and finally briefly describe the (relatively recent) variational tools required to establish those properties.

In [38], local superlinear convergence of sSQP is established under MFCQ (1.6), SOSC (1.7) assumed to hold for all multipliers, the existence of a multiplier $\bar{\mu}$ satisfying the strict complementarity condition $\bar{\mu}_{A(\bar{x})} > 0$, and the assumption that the initial dual iterate is close enough to such a multiplier. Also, [38] gives an analysis in the presence of round-off errors. In [40], the assumption of strict complementarity has been removed. Also, [39] suggests a certain inexact SQP framework which includes sSQP as a particular case. The assumptions, however, still contain MFCQ. In [18], CQs are not used at the expense of employing instead of the weaker SOSC (1.7) the strong SOSC (SSOSC)

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C_+(\bar{x}, \bar{\mu}) \setminus \{0\}, \quad (2.1)$$

where

$$C_+(\bar{x}, \bar{\mu}) = \{\xi \in \mathbf{R}^n \mid h'(\bar{x})\xi = 0, g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0\},$$

and assuming that the dual starting point is close to a multiplier satisfying this SSOSC. (SSOSC (2.1) is stronger than SOSC (1.7) because $C(\bar{x}) \subset C_+(\bar{x}, \bar{\mu})$.) In [12], the result of [18] was recovered from more general principles (some details will be discussed below), and somewhat sharpened. The iterative framework of [12] was further used in [10] to prove local superlinear convergence using SOSC (1.7) only, with no CQs or other assumptions. Moreover, the method was extended to variational problems (see Section 4 below). Quasi-Newton versions of sSQP are analyzed under SOSC in [7]. In [25] it was shown that the SOSC cannot be relaxed when inequality constraints are present, but for equality-constrained problems the weaker condition of noncriticality of the relevant Lagrange multiplier (see the definition below) is sufficient for convergence.

A Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is said to be *critical* if there exists a triple $(\xi, \eta, \zeta) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$, with $\xi \neq 0$, satisfying the system

$$\begin{aligned} \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda}, \bar{\mu})\xi + (h'(\bar{x}))^T \eta + (g'(\bar{x}))^T \zeta &= 0, \quad h'(\bar{x})\xi = 0, \quad g'_{A_+(\bar{x}, \bar{\mu})}(\bar{x})\xi = 0, \\ \zeta_{A_0(\bar{x}, \bar{\mu})} &\geq 0, \quad g'_{A_0(\bar{x}, \bar{\mu})}(\bar{x})\xi \leq 0, \quad \zeta_i \langle g'_i(\bar{x}), \xi \rangle = 0, \quad i \in A_0(\bar{x}, \bar{\mu}), \\ \zeta_{\{1, \dots, m\} \setminus A(\bar{x})} &= 0, \end{aligned} \quad (2.2)$$

and *noncritical* otherwise. We refer the reader to [21, 22, 24, 25, 19, 20] for the role this notion plays in convergence properties of algorithms, stability, error bounds, and other issues. Some comments will also be given below. When there are no inequality constraints, it can be seen from (2.2) that a multiplier $\bar{\lambda} \in \mathcal{M}(\bar{x})$ being critical means that

$$\exists \xi \in \ker h'(\bar{x}) \setminus \{0\} \text{ such that } \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \text{im}(h'(\bar{x}))^T. \quad (2.3)$$

It can be easily seen, essentially observing that $\text{im}(h'(\bar{x}))^T = (\ker h'(\bar{x}))^\perp$, that this noncriticality property for equality-constrained problems is implied by the corresponding version of

SOSC (1.7), but not vice versa. The same conclusion holds for the general case: if (2.2) has a solution with $\xi \neq 0$, multiplying the first equality in (2.2) by this ξ and using the other relations in (2.2), one arrives to a contradiction with SOSC (1.7). It should be emphasized that SOSC is a much stronger assumption than noncriticality. Noncritical multipliers, if they exist, form a relatively open and dense subset of the multiplier set $\mathcal{M}(\bar{x})$, which is of course not so for multipliers satisfying SOSC.

We are now in position to formally state local convergence properties of sSQP [10, 25]. Note that in Theorem 2.1 below, if there are equality constraints only, everything that involves the multiplier μ disappears from the statement. Note also that in the equality-constrained case, finding a stationary point of the sSQP subproblem (1.8) is equivalent to solving the linear system of equations

$$\begin{aligned} \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x - x^k) + (h'(x^k))^T(\lambda - \lambda^k) &= -\frac{\partial L}{\partial x}(x^k, \lambda^k), \\ h'(x^k)(x - x^k) - \sigma_k(\lambda - \lambda^k) &= -h(x^k), \end{aligned} \quad (2.4)$$

in the variables (x, λ) .

Theorem 2.1 *Let $f : \mathbf{R}^n \rightarrow \mathbf{R}$, $h : \mathbf{R}^n \rightarrow \mathbf{R}^l$ and $g : \mathbf{R}^n \rightarrow \mathbf{R}^m$ be twice differentiable in a neighborhood of \bar{x} , with their second derivatives being continuous at \bar{x} . Let $\sigma_k = \sigma(x^k, \lambda^k, \mu^k)$, where σ is given by (1.9). Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a solution of the KKT system (1.2), satisfying SOSC (1.7). If there are equality constraints only, let instead $\bar{\lambda}$ be a noncritical multiplier (i.e., the right-most relation in (2.3) does not hold for any $\xi \neq 0$).*

Then for any $c > 0$ large enough and any starting point $(x^0, \lambda^0, \mu^0) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}_+^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ such that for each $k = 0, 1, \dots$, x^{k+1} is a stationary point of sSQP subproblem (1.8) with associated Lagrange multipliers $(\lambda^{k+1}, \mu^{k+1})$ which satisfies

$$\|(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)\| \leq c \text{dist}((x^k, \lambda^k, \mu^k), \{\bar{x}\} \times \mathcal{M}(\bar{x}));$$

any such sequence converges to $(\bar{x}, \lambda^, \mu^*)$ with some $(\lambda^*, \mu^*) \in \mathcal{M}(\bar{x})$, and the rates of convergence of $\{(x^k, \lambda^k, \mu^k)\}$ to $(\bar{x}, \lambda^*, \mu^*)$ and of $\{\text{dist}((x^k, \lambda^k, \mu^k), \{\bar{x}\} \times \mathcal{M}(\bar{x}))\}$ to zero are superlinear. Moreover, the rates of convergence are quadratic provided the second derivatives of f , h and g are locally Lipschitz-continuous with respect to \bar{x} .*

Some comments are in order. Under SSOSC (2.1), solutions of sSQP subproblems (1.8) can in addition be shown to be unique in some neighbourhood [10]. Furthermore, in the equality-constrained case, locally and under the noncriticality assumption, the linear system (2.4) has the unique solution, i.e., the sSQP subproblem has the unique stationary point [25]. For the equality-constrained case, sSQP is the only currently known method that solves a linear system or a QP per iteration (i.e., an explicitly Newtonian method), and which requires for convergence something weaker than SOSC (in particular, noncriticality of the multiplier) and does not need any CQs. In [25] it is shown that when there are inequality constraints, SOSC cannot be replaced by noncriticality. Whether convergence of the primal (rather than primal-dual) sequence generated by sSQP is also superlinear is an open question [8]. Recall that in general, superlinear convergence of primal-dual sequence does not imply any rate

Table 1: Distance to solution on last 5 iterations

$\ x - \bar{x}\ + \text{dist}(\mu, \mathcal{M}(\bar{x}))$
1.5891e-001
1.0599e-002
7.3688e-005
6.3242e-009
8.4865e-017

for the primal (or dual) sequence separately [4, Exercise 14.8]. For SQP, the primal rate is superlinear [8]. For sSQP, only a kind of “two-step” superlinear estimate for the primal sequence is available [7].

We illustrate the convergence result in Theorem 2.1 with the following example.

Example 2.1 Consider the optimization problem

$$\begin{aligned} \min \quad & x_1 x_2 - x_2^2 / 2 \\ \text{s.t.} \quad & x_2^2 \leq 0, \quad -2x_1 + x_2 \leq 0, \quad x_1 - 2x_2 \leq 0. \end{aligned} \tag{2.5}$$

It can be seen that $\bar{x} = (0, 0)$ is the unique solution of this problem, and that the associated set of Lagrange multipliers is given by

$$\mathcal{M}(\bar{x}) = \{(\mu_1, \mu_2, \mu_3) \in \mathbf{R}^3 \mid \mu_1 \geq 0, \mu_2 = \mu_3 = 0\}.$$

In particular, MFCQ (1.6) does not hold and $\mathcal{M}(\bar{x})$ is unbounded. Furthermore, SOSC (1.7) holds at $(\bar{x}, \bar{\mu})$ for any $\bar{\mu} \in \mathcal{M}(\bar{x})$ with $\bar{\mu}_1 > 0$, but SSOSC (2.1) is not satisfied for any multiplier.

Experiments were performed with Matlab implementation of sSQP, using the built-in subroutine `quadprog` for solving QP subproblems (1.8) and choosing random starting points $x_i^0 \in [-1/2, 1/2]$, $i = 1, 2$, and $\mu_j^0 \in [0, 1]$, $j = 1, 2, 3$. The stopping criterion is $\sigma(x^k, \mu^k) < 10^{-15}$.

In about 10% of the cases, the sequence converged linearly to $(\bar{x}, \bar{\mu})$ with $\bar{\mu}_1 = 0$ (SOSC is not valid at this solution). Such cases appear to correspond to the choices of starting points that are not close enough to a solution satisfying SOSC (so that Theorem 2.1 does not apply). About 3% of the starting points produced unsolvable subproblems at the first iteration (for the same reason as above – starting points not being close enough to a solution). All the remaining runs converged superlinearly to a primal-dual solution satisfying SOSC. Table 1 shows the average values of $\|x^k - \bar{x}\| + \text{dist}(\mu^k, \mathcal{M}(\bar{x}))$ for the last 5 iterations in the cases of convergence to a primal-dual solution satisfying SOSC.

The analysis that leads to Theorem 2.1 relies on the variational Newtonian framework for generalized equations (GEs) with nonisolated solutions, developed in [12] (in our context, in the absence of CQs dual solutions of the KKT system (1.2) are not isolated). To that end, consider the generalized equation (GE)

$$\Phi(u) + N(u) \ni 0, \tag{2.6}$$

where $\Phi : \mathbf{R}^\nu \rightarrow \mathbf{R}^\nu$ is a smooth (single-valued) mapping, and $N(\cdot)$ is a set-valued mapping from \mathbf{R}^ν to the subsets of \mathbf{R}^ν . As is well known, the KKT system (1.2) corresponds to the GE (2.6) with $\nu = n + l + m$, the mapping $\Phi : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \rightarrow \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ given by

$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x, \lambda, \mu), h(x), -g(x) \right), \quad u = (x, \lambda, \mu), \quad (2.7)$$

and with N beining the normal cone to the set

$$Q = \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}_+^m, \quad (2.8)$$

i.e., $N(u) = N_Q(u) = \{v \mid \langle v, w - u \rangle \leq 0 \ \forall w \in Q\} = \{0\} \times \{0\} \times \mathbf{R}_-^m$.

Consider the class of methods for solving (2.6) that, given the current iterate $u^k \in \mathbf{R}^\nu$, generate the next iterate u^{k+1} as a solution of the subproblem of the form

$$\mathcal{A}(u^k, u) + N(u) \ni 0, \quad (2.9)$$

where for $\tilde{u} \in \mathbf{R}^\nu$ the mapping $\mathcal{A}(\tilde{u}, \cdot)$ is some kind of approximation of Φ around \tilde{u} . For example, if

$$\mathcal{A}(\tilde{u}, u) = \{\Phi(\tilde{u}) + \Phi'(\tilde{u})(u - \tilde{u})\}, \quad \tilde{u}, u \in \mathbf{R}^\nu,$$

the iteration subproblem (2.9) becomes that of the Josephy–Newton method for GEs, and when applied to the KKT system (1.2) (i.e., the special case of GE described above), it corresponds to the SQP subproblem (1.3). For each $\tilde{u} \in \mathbf{R}^\nu$, define the set

$$U(\tilde{u}) = \{u \in \mathbf{R}^\nu \mid \mathcal{A}(\tilde{u}, u) + N(u) \ni 0\}, \quad (2.10)$$

so that $U(u^k)$ is the solution set of the iteration subproblem (2.9). As is usual *and natural* in local convergence considerations, one has to specify which of the solutions of (2.9) are allowed to be the next iterate (solutions “far away” must clearly be discarded from local analysis; note that sSQP subproblem (1.8) need not be strongly convex and thus may have such “far away” solutions). In other words, we have to restrict the distance from the current iterate u^k to the next one, i.e., to an element of $U(u^k)$ that can be declared to be u^{k+1} . To that end, for an arbitrary but fixed $c > 0$ define the subset of the solution set of the subproblem (2.9) by

$$U_c(\tilde{u}) = \{u \in U(\tilde{u}) \mid \|u - \tilde{u}\| \leq c \operatorname{dist}(\tilde{u}, \bar{U})\}, \quad (2.11)$$

and consider the iterative scheme

$$u^{k+1} \in U_c(u^k), \quad k = 0, 1, \dots \quad (2.12)$$

Superlinear convergence of this scheme is established under the following three conditions:

- (i) (Upper Lipschitzian behavior of solutions of GE under canonical perturbations)

For every $r \in \mathbf{R}^\nu$ close enough to 0, any solution $u(r)$ of the perturbed GE

$$\Phi(u) + N(u) \ni r \quad (2.13)$$

close enough to \bar{u} satisfies the estimate

$$\operatorname{dist}(u(r), \bar{U}) = O(\|r\|) \quad \text{as } r \rightarrow 0.$$

(ii) (Precision of approximation of Φ in subproblems)

There exists a function $\omega : \mathbf{R}_+ \rightarrow \mathbf{R}_+$ such that $\omega(t) = o(t)$ as $t \rightarrow 0$ and the estimate

$$\sup\{\|w\| \mid w \in \Phi(u) - \mathcal{A}(\tilde{u}, u), \|u - \tilde{u}\| \leq c \operatorname{dist}(\tilde{u}, \bar{U})\} \leq \omega(\operatorname{dist}(\tilde{u}, \bar{U})) \quad (2.14)$$

holds for all $\tilde{u} \in \mathbf{R}^\nu$ close enough to \bar{u} .

(iii) (Solvability of subproblems with the localization condition)

For any $\tilde{u} \in \mathbf{R}^\nu$ close enough to \bar{u} the set $U_c(\tilde{u})$ defined by (2.10), (2.11) is nonempty.

Some comments are in order. For GE corresponding to the KKT system (1.2), the canonically perturbed problem has the form

$$\frac{\partial L}{\partial x}(x, \lambda, \mu) = a, \quad h(x) = b, \quad \mu \geq 0, \quad g(x) \leq c, \quad \langle \mu, g(x) - c \rangle = 0,$$

for $r = (a, b, c) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$. For KKT systems, the upper Lipschitzian behavior of solutions under canonical perturbations (the first assumption above) is equivalent to noncriticality of the Lagrange multiplier (under the smoothness assumptions in this survey) [19]. In particular, it is implied by the SOSC (1.7). The second assumption above naturally holds for Newton-type methods, and in particular for sSQP if the stabilization parameter is properly chosen (for example, based on the KKT natural residual (1.9)). The third assumption on solvability of subproblems and localization condition is where the most work is required [10, 25]. And it is here where noncriticality of the multiplier needs to be strengthened to SOSC if inequality constraints are present.

We finally note that sSQP can also be interpreted within the perturbed Josephy-Newton framework of [23]. However, the main convergence result in this framework requires SMFCQ. If the method is interpreted instead via [12, 10] as outlined above, no CQs are needed to prove local convergence. It is also interesting to mention that a modification of the Newtonian framework of [12] is used in [11] to derive local convergence and rate of convergence results for the augmented Lagrangian algorithm (method of multipliers) under SOSC (1.7) only, significantly improving on the classical results such as in [1] that assume in addition LICQ (1.5) and strict complementarity. Moreover, for the equality-constrained case SOSC can be relaxed to noncriticality [20], as is the case for sSQP (the required analysis is very different though). It is interesting that even though the augmented Lagrangian method is not of Newton type, the Newtonian lines of analysis turned to be very fruitful for this context as well.

3 Globalization issues

As any Newtonian method, sSQP is a local scheme, guaranteed to converge if initialized at a point close enough to a solution with the properties discussed in Section 2. To obtain a complete algorithm, some strategy to globalize convergence is needed (so that arbitrary starting points can be used). This proved to be a rather difficult task. Recall that to globalize SQP at least three different approaches are known: linesearch [4, Chapter 17] or trust-region

[5, Chapter 15.4] for a nonsmooth penalty function, and the filter technique [13, 14, 34]. For example, if a positive definite matrix H_k is employed in the QP (1.3), then the generated direction $x^{k+1} - x^k$ is that of descent for the penalty function

$$\varphi_{c_k}(x) = f(x) + c_k(\|h(x)\|_1 + \|\max\{0, g(x)\}\|_1), \quad (3.1)$$

provided one takes $c_k > \|(\lambda^{k+1}, \mu^{k+1})\|_\infty$. One can then perform linesearch in the obtained direction to guarantee progress towards solving (1.1) via decreasing the penalty function with respect to its value at the previous iterate x^k and then re-defining x^{k+1} accordingly. To find a suitable penalty function for which the direction computed by the sSQP subproblem (1.8) is of descent, proved a challenge. In particular, the penalty function like (3.1), or other “usual” candidates, do not do the job.

Some numerical results on global behaviour of sSQP, without attempting to globalize the method itself, are reported in [30], but this experience is rather limited (just a few test problems are considered). More test problems have been employed in [24] but globalization used there is a heuristic not supported by a proof.

We next survey the few approaches to globalize sSQP that have been proposed so far.

In [36] the globalization technique is based on a linesearch for the so-called primal-dual augmented Lagrangian. This work deals with optimization problems in the format

$$\begin{aligned} & \text{minimize}_x && f(x) \\ & \text{subject to} && h(x) = 0, \quad x \geq 0. \end{aligned} \quad (3.2)$$

(The more general problem (1.1) can be reformulated into this setting adding slack variables.) For the optimization problem (3.2), taking $\mu = f'(x) + (h'(x))^T \lambda$, the natural residual (1.9) can be written as

$$\sigma(x, \lambda) = \left\| \begin{pmatrix} \min\{f'(x) + (h'(x))^T \lambda, x\} \\ h(x) \end{pmatrix} \right\|.$$

Also, the simple nonnegativity constraint is excluded from stabilization. Then, instead of (1.8), the sSQP subproblem in [36] is given by

$$\begin{aligned} & \text{minimize}_{(x, \lambda)} && \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle H_k(x - x^k), x - x^k \rangle + \frac{\sigma_k}{2} \|\lambda\|^2 \\ & \text{subject to} && h(x^k) + h'(x^k)(x - x^k) - \sigma_k(\lambda - \tilde{\lambda}^k) = 0, \quad x \geq 0, \end{aligned} \quad (3.3)$$

where $\tilde{\lambda}^k$ is a reference Lagrange multiplier estimate and H_k is a symmetric matrix such that $H_k + (1/\sigma_k)(h'(x^k))^T h'(x^k)$ is positive definite. It should be noted that there does not seem to be any theory to justify that this “partial” (excluding the constraint $x \geq 0$) stabilization in (3.3) actually inherits local convergence properties of sSQP under some reasonable assumptions. In terms of local convergence, the idea of [36] is to use in addition identification of active inequality constraints, so that the overall algorithm eventually becomes sSQP for the associated equality-constrained problem. From the analysis in [36], if (x^{k+1}, λ^{k+1}) is the (unique) solution of (3.3) then $(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k)$ is a descent direction at (x^k, λ^k) for the primal-dual penalty function

$$\bar{\varphi}_c(x, \lambda; \tilde{\lambda}^k, \sigma_k) = f(x) + \langle \tilde{\lambda}^k, h(x) \rangle + \frac{1}{2\sigma_k} \|h(x)\|^2 + \frac{c}{2\sigma_k} \|h(x) - \sigma_k(\lambda - \tilde{\lambda}^k)\|^2, \quad (3.4)$$

where $c > 0$ is a fixed parameter. This penalty function is minimized using linesearch, thus re-defining (x^{k+1}, λ^{k+1}) . The reference Lagrange multiplier $\tilde{\lambda}^{k+1}$ is updated to λ^{k+1} if either the weighted natural residual for problem (3.2) is small or if the natural residual for the problem of minimizing $\bar{\varphi}_c(x, \lambda; \tilde{\lambda}^k, \sigma_k)$ subject to $x \geq 0$ is small. The dual stabilization parameter σ_k and other algorithmic parameters are updated by certain rules. According to [36, Theorem 4.2], if the generated sequence $\{x^k\}$ is bounded and the sequence $\{H_k\}$ is chosen bounded with $\{H_k + (1/\sigma_k)(h'(x^k))^T h'(x^k)\}$ being also uniformly positive definite, then either there exists an index set \mathcal{K} such that $\lim_{\mathcal{K} \ni k \rightarrow \infty} \sigma(x^k, \lambda^k) = 0$ (accumulation points of this subsequence solve the KKT system) or there exists an index set \mathcal{S} such that $\lim_{\mathcal{S} \ni k \rightarrow \infty} \sigma_k = 0$, $\{\tilde{\lambda}^k\}_{k \in \mathcal{S}}$ is bounded and

$$\lim_{\mathcal{S} \ni k \rightarrow \infty} \left(\left\| \min \left\{ \frac{\partial \bar{\varphi}_c}{\partial x}(x^{k+1}, \lambda^{k+1}; \tilde{\lambda}^k, \sigma_k), x^{k+1} \right\} \right\| + \left\| \frac{\partial \bar{\varphi}_c}{\partial \lambda}(x^{k+1}, \lambda^{k+1}; \tilde{\lambda}^k, \sigma_k) \right\| \right) = 0.$$

Another globalization strategy is proposed in [9]. It is based on the inexact restoration ideas [29], and uses linesearch for a primal-dual nondifferentiable penalty function. This work considers problems in the format

$$\begin{aligned} & \text{minimize}_x && f(x) \\ & \text{subject to} && h(x) = 0, \quad a \leq x \leq b, \end{aligned} \tag{3.5}$$

where a and b are finite bounds. For this problem, the natural residual (1.9) is given by

$$\sigma(x, \lambda) = \left\| \begin{pmatrix} \min\{x - a, \max\{f'(x) + (h'(x))^T \lambda, x - b\}\} \\ h(x) \end{pmatrix} \right\|.$$

The corresponding sSQP subproblem again employs only partial stabilization (for the general equality constraints), and has the form

$$\begin{aligned} & \text{minimize}_{(x, \lambda)} && \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle H_k(x - x^k), x - x^k \rangle + \frac{\sigma_k}{2} \|\lambda\|^2 \\ & \text{subject to} && h(x^k) + h'(x^k)(x - x^k) - \sigma_k(\lambda - \tilde{\lambda}^k) = 0, \quad a \leq x \leq b, \end{aligned} \tag{3.6}$$

where $\tilde{\lambda}^k$ is a reference Lagrange multiplier approximation and H_k is a symmetric positive definite matrix. The penalty function used in [9] is

$$\check{\varphi}_{c_k}(x, \lambda; \tilde{\lambda}^k, \sigma_k) = f(x) + \frac{\sigma_k}{2} \|\lambda\|^2 + c_k \|h(x) - \sigma_k(\lambda - \tilde{\lambda}^k)\|. \tag{3.7}$$

Note that this is a penalty function for the problem of minimizing $f(x) + (\sigma_k/2) \|\lambda\|^2$ subject to $h(x) - \sigma_k(\lambda - \tilde{\lambda}^k) = 0$, and the latter problem is equivalent to minimizing $f(x) + \langle \tilde{\lambda}^k, h(x) \rangle + 1/(2\sigma_k) \|h(x)\|^2$. Thus, this penalty function is also related to the augmented Lagrangian.

The inexact restoration strategy presented in [9] can be interpreted as two-step linesearch for the function (3.7), where in the first step the penalty parameter c_k is increased in order to achieve $\check{\varphi}_{c_k}(x^k, \hat{\lambda}^k; \tilde{\lambda}^k, \sigma_k) < \check{\varphi}_{c_k}(x^k, \lambda^k; \tilde{\lambda}^k, \sigma_k)$ with $\hat{\lambda}^k = \tilde{\lambda}^k + (1/\sigma_k)h(x^k)$, and in the second step linesearch is performed along the direction $(x^{k+1} - x^k, \lambda^{k+1} - \hat{\lambda}^k)$ to re-define (x^{k+1}, λ^{k+1}) so that $\check{\varphi}_{c_k}(x^{k+1}, \lambda^{k+1}; \tilde{\lambda}^k, \sigma_k) \leq \check{\varphi}_{c_k}(x^k, \lambda^k; \tilde{\lambda}^k, \sigma_k)$. If the linesearch direction is

small enough (with respect to the inexact restoration criteria), then (x^{k+1}, λ^{k+1}) is accepted as the new primal-dual iterate and the reference Lagrange multiplier $\tilde{\lambda}^{k+1}$ is updated to λ^{k+1} . The dual stabilization parameter σ_k is updated by a suitable rule. According to [9, Theorem 2], if the sequence of matrices $\{H_k\}$ is chosen uniformly bounded and uniformly positive definite, and if \bar{x} is an accumulation point of the sequence $\{x^k\}$, then \bar{x} is a stationary point of the problem (3.5) if $\{\sigma_k\}$ is bounded away from zero, or it is a stationary point of the problem of minimizing the infeasibility measure $\|h(x)\|^2$ subject to $a \leq x \leq b$ if $\{\sigma_k\}$ converges to zero. The algorithm of [9] solves sSQP subproblems, but in a sense they can be considered as “inner iterations” within the inexact restoration scheme (which drives global convergence of the method). In particular, it is not known whether locally the algorithm indeed behaves as sSQP under some assumptions (i.e., solves only one sSQP subproblem per inexact restoration iteration).

We next comment on some other ideas for sSQP globalization, which led to partial developments of some promises but did not materialize into complete algorithms so far.

A globalization strategy can be attempted using the principle of the augmented Lagrangian method, i.e., decrease the augmented Lagrangian function in the primal space and increase it in the dual. Consider the classical augmented Lagrangian for problem (1.1):

$$\bar{L}(x, \lambda, \mu; \sigma) = f(x) + \frac{\sigma}{2} \left(\left\| \lambda + \frac{1}{\sigma} h(x) \right\|^2 - \|\lambda\|^2 \right) + \frac{\sigma}{2} \left(\left\| \max\{0, \mu + \frac{1}{\sigma} g(x)\} \right\|^2 - \|\mu\|^2 \right),$$

$\sigma > 0$. It can be seen that if $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is a solution of the sSQP subproblem (1.8), then $(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)$ is a descent direction at (x^k, λ^k, μ^k) for the “difference of two augmented Lagrangians” function

$$\psi_{\tilde{\sigma}_k, \tilde{\sigma}_k}(x, \lambda, \mu; x^k, \lambda^k, \mu^k) = \bar{L}(x, \lambda^k, \mu^k; \tilde{\sigma}_k) - \bar{L}(x^k, \lambda, \mu; \tilde{\sigma}_k),$$

for any $\hat{\sigma}_k \in [\tilde{\sigma}_k/2, \tilde{\sigma}_k]$. Moreover, the directional derivative is less than $-\Delta_k$, where

$$\begin{aligned} \Delta_k &= \langle H_k(x^{k+1} - x^k), x^{k+1} - x^k \rangle + \frac{\tilde{\sigma}_k}{2} \|\lambda^{k+1} - \lambda^k\|^2 + \frac{\tilde{\sigma}_k}{2} \|\mu^{k+1} - \mu^k\|^2 \\ &\quad + \frac{\tilde{\sigma}_k}{2} \left\| \lambda^k + \frac{1}{\tilde{\sigma}_k} h(x^k) - \lambda^{k+1} \right\|^2 + \frac{\tilde{\sigma}_k}{2} \left\| \max\{0, \mu^k + \frac{1}{\tilde{\sigma}_k} g(x^k)\} - \mu^{k+1} \right\|^2. \end{aligned}$$

It can be shown that if $\{\Delta_k\}$ tends to zero, the sequence of matrices $\{H_k\}$ is chosen uniformly bounded and uniformly positive definite and $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a limit point of the sequence $\{(x^k, \lambda^k, \mu^k)\}$, then \bar{x} is a stationary point of the problem (1.1) if $\{\sigma_k\}$ is bounded away from zero, or it is a stationary point of the problem of minimizing the infeasibility measure $\|h(x)\|^2 + \|\max\{0, g(x)\}\|^2$ if $\{\sigma_k\}$ converges to zero. However, so far there are no reasonable hypotheses to guarantee $\Delta_k \rightarrow 0$ from a standard linesearch. From another point of view, this strategy is related to finding a solution of an equilibrium problem. If for some $\hat{\sigma}, \tilde{\sigma} > 0$ it holds that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a solution of the optimization problem $\min_{(x, \lambda, \mu)} \psi_{\hat{\sigma}, \tilde{\sigma}}(x, \lambda, \mu; \bar{x}, \bar{\lambda}, \bar{\mu})$, then $(\bar{x}, \bar{\lambda}, \bar{\mu})$ solves the KKT system (1.2). Conversely, if $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a solution of the KKT system (1.2) satisfying SOSC (1.7), then $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a local minimizer of the latter problem.

Another issue concerned with global convergence of sSQP has to do with possible attraction of the iterates to critical Lagrange multipliers, and eventual slow convergence rate

as a consequence. In [21, 22, 24, 27] this phenomenon was exhibited for various Newtonian and Newton-related methods, such as SQP and its quasi-Newton implementations, and the linearly constrained (augmented) Lagrangian methods [35, 31, 15]. Both theoretical considerations and numerical results for SNOPT [16] and MINOS [32] solvers were presented, which put in evidence that when critical multipliers exist, they serve as attractors of the dual sequence generated by the type of methods in question. Moreover, the reason for slow convergence in the degenerate cases is precisely attraction to critical multipliers, as convergence to noncritical ones would have given the primal superlinear rate. Numerical results in [24] show that the effect of attraction (globally, i.e., from “far away” points) to critical multipliers still exists for sSQP too (when evaluating the numbers reported therein, it is important to keep in mind that critical multipliers are typically few; the usual situation is that they form a set of measure zero within the set of all multipliers), but the attraction is much less persistent for sSQP than for the other algorithms. The runs clearly split into two groups. Sometimes the (globalized, heuristically in that reference) process manages to enter the “good” primal-dual region, where the stabilization term starts working properly (has the needed “size”), and then it converges superlinearly with the dual limit being noncritical. However, in a considerable number of cases this does not happen, and then the process still converges slowly to a critical multiplier. Thus, although sSQP does help when compared to the alternatives, by itself it does not seem to be a fully reliable tool for avoiding the effect of attraction and its negative consequences. It would seem that some special modifications would be needed in the “global” phase of the method to reliably avoid convergence to critical multipliers.

Overall, building really satisfactory globalization techniques for sSQP is a challenging matter, which (for general problems) should still be considered an open question at this time.

4 Extensions to variational problems

Denote

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

and let $N_D(x)$ be the dual cone of the tangent (contingent) cone $T_D(x)$ to the set D at $x \in \mathbf{R}^n$, i.e., $N_D(x) = \emptyset$ for $x \notin D$ and otherwise $N_D(x) = (T_D(x))^\circ$ where

$$T_D(x) = \{\xi \in \mathbf{R}^n \mid \exists \{\xi^k\} \rightarrow \xi, t_k \rightarrow 0 \text{ s.t. } x + t_k \xi^k \in D \forall k, t_k > 0\}.$$

Consider the variational problem (VP)

$$F(x) + N_D(x) \ni 0, \tag{4.1}$$

where $F : \mathbf{R}^n \rightarrow \mathbf{R}^n$. In particular, for the optimization problem (1.1) this VP represents the first-order necessary optimality condition

$$x \in D, \quad \langle f'(x), \xi \rangle \geq 0 \quad \forall \xi \in T_D(x),$$

if we take $F(x) = f'(x)$. If the set D is convex, then (4.1) gives the usual variational inequality (VI)

$$x \in D, \quad \langle F(x), y - x \rangle \geq 0 \quad \forall y \in D.$$

Associated to solving VP (4.1) is the KKT system

$$F(x) + (h'(x))^T \lambda + (g'(x))^T \mu = 0, \quad h(x) = 0, \quad \mu \geq 0, \quad g(x) \leq 0, \quad \langle \mu, g(x) \rangle = 0. \quad (4.2)$$

Define the mapping $G : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \rightarrow \mathbf{R}^n$ by

$$G(x, \lambda, \mu) = F(x) + (h'(x))^T \lambda + (g'(x))^T \mu.$$

Let $(x^k, \lambda^k, \mu^k) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}_+^m$ be the current primal-dual approximation to a solution of (4.2), and let $\sigma_k > 0$ be the dual stabilization parameter. Define the affine mapping $\Phi_k : \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \rightarrow \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m$ by

$$\Phi_k(u) = \left(F(x^k) + \frac{\partial G}{\partial x}(x^k, \lambda^k, \mu^k)(x - x^k), \sigma_k \lambda, \sigma_k \mu \right), \quad u = (x, \lambda, \mu),$$

and consider the affine VI of the form

$$u \in Q_k, \quad \langle \Phi_k(u), v - u \rangle \geq 0 \quad \forall v \in Q_k, \quad (4.3)$$

where

$$Q_k = \left\{ u = (x, \lambda, \mu) \in \mathbf{R}^n \times \mathbf{R}^l \times \mathbf{R}^m \left| \begin{array}{l} h(x^k) + h'(x^k)(x - x^k) - \sigma_k(\lambda - \lambda^k) = 0, \\ g(x^k) + g'(x^k)(x - x^k) - \sigma_k(\mu - \mu^k) \leq 0 \end{array} \right. \right\}.$$

As can be easily seen, in the optimization case (1.1) the VI (4.3) is precisely the first-order (primal) necessary optimality condition for the sSQP subproblem (1.8), if one takes $F(x) = f'(x)$. Thus this scheme contains sSQP for optimization as a special case. Note that the method makes good sense also in the variational setting, as solving the fully nonlinear VP (4.1) is replaced by solving a sequence of affine VIs (4.3) (the mapping Φ_k is affine and the set Q_k is polyhedral).

Convergence analysis of this stabilized Newton method for variational problems can be found in [10].

5 Concluding remarks

We presented a survey of literature and some discussion of the stabilized version of the fundamental sequential quadratic programming method for constrained optimization. Further material, in particular comprehensive local convergence analysis, can be found in the forthcoming book [26].

References

- [1] D.P. Bertsekas. Constrained Optimization and Lagrange Multiplier Methods. Academic Press, New York, 1982.
- [2] B.T. Boggs and J.W. Tolle. Sequential quadratic programming. Acta Numerica 4 (1996), 1–51.

- [3] J.F. Bonnans. Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Applied Mathematics and Optimization* 29 (1994), 161–186.
- [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] A.R. Conn, N.I.M. Gould, and Ph.L. Toint. *Trust-Region Methods*. SIAM, Philadelphia, 2000.
- [6] G. Di Pillo and L. Grippo. A new class of augmented Lagrangians in nonlinear programming. *SIAM J. on Control and Optimization* 17 (1979), 618–628.
- [7] D. Fernández. A quasi-Newton strategy for the sSQP method for variational inequality and optimization problems. *Mathematical Programming* 137 (2013), 199–223.
- [8] D. Fernández, A.F. Izmailov, and M.V. Solodov. Sharp primal superlinear convergence results for some Newtonian methods for constrained optimization. *SIAM Journal on Optimization* 20 (2010), 3312–3334.
- [9] D. Fernández, E.A. Pilotta, and G.A. Torres. An inexact restoration strategy for the globalization of the sSQP method. *Computational Optimization and Applications* (2012) DOI 10.1007/s10589-012-9502-y.
- [10] D. Fernández and M. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. *Mathematical Programming* 125 (2010), 47–73.
- [11] D. Fernández and M. Solodov. Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. *SIAM Journal on Optimization* 22 (2012), 384–407.
- [12] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Mathematical Programming* 94 (2002), 91–124.
- [13] R. Fletcher, N. Gould, S. Leyffer, P. Toint, and A. Wächter. Global convergence of trust-region and SQP-filter algorithms for general nonlinear programming. *SIAM Journal on Optimization* 13 (2002), 635–659.
- [14] R. Fletcher, S. Leyffer, and P.L. Toint. On the global convergence of a filter-SQP algorithm. *SIAM Journal on Optimization* 13 (2002), 44–59.
- [15] M.P. Friedlander and M.A. Saunders. A globally convergent linearly constrained Lagrangian method for nonlinear optimization. *SIAM Journal on Optimization* 15 (2005), 863–897.
- [16] P.E. Gill, W. Murray, and M.A. Saunders. SNOPT: An SQP algorithm for large-scale constrained optimization. *SIAM Journal on Optimization* 12 (2002), 979–1006.

- [17] P.E. Gill and E. Wong. Sequential quadratic programming methods. In J. Lee and S. Leyffer, eds., *Mixed Integer Nonlinear Programming, The IMA Volumes in Mathematics and its Applications*, Volume 154, Springer-Verlag, pp. 147–224, 2012.
- [18] W.W. Hager. Stabilized sequential quadratic programming. *Computational Optimization and Applications* 12 (1999), 253–273.
- [19] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. *Mathematical Programming* (2012) DOI 10.1007/s10107-012-0586-z.
- [20] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. Local convergence of the method of multipliers for variational and optimization problems under the sole noncriticality assumption. August 2013. Available at <http://pages.cs.wisc.edu/~solodov/solodov.html>
- [21] A.F. Izmailov and M.V. Solodov. On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. *Mathematical Programming* 117 (2009), 271–304.
- [22] 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* 42 (2009), 231–264.
- [23] A.F. Izmailov and M.V. Solodov. Inexact Josephy–Newton framework for generalized equations and its applications to local analysis of Newtonian methods for constrained optimization *Computational Optimization and Applications* 46 (2010), 347–368.
- [24] A.F. Izmailov and M.V. Solodov. On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. *Mathematical Programming* 126 (2011), 231–257.
- [25] A.F. Izmailov and M.V. Solodov. Stabilized SQP revisited. *Mathematical Programming* 133 (2012), 93–120.
- [26] A.F. Izmailov and M.V. Solodov. *Newton-Type Methods for Optimization and Variational Problems*. Springer Series in Operations Research and Financial Engineering. Forthcoming.
- [27] A.F. Izmailov and E.I. Uskov. Attraction of Newton method to critical Lagrange multipliers: fully quadratic case. 2012.
- [28] D.-H. Li and L. Qi. Stabilized SQP method via linear equations. *Applied Mathematics Technical Report AMR00/5*. University of New South Wales, Sydney, 2000.
- [29] J.M. Martínez and E.A. Pilotta. Inexact restoration methods for nonlinear programming: advances and perspectives. In L.Q. Qi, K.L. Teo, and X.Q. Yang, eds., *Optimization and Control with Applications*, pp. 271–292. Springer, 2005.

- [30] E.M.E. Mostafa, L.N. Vicente, and S.J. Wright. Numerical behavior of a stabilized SQP method for degenerate NLP problems. In C. Blik, C. Jermann, and A. Neumaier, eds., *Global Optimization and Constraint Satisfaction*, Lecture Notes in Computer Science 2861, pp. 123-141. Springer-Verlag, Berlin, 2003.
- [31] B.A. Murtagh and M.A. Saunders. A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints. *Mathematical Programming Study* 16 (1982), 84-117.
- [32] B.A. Murtagh and M.A. Saunders. MINOS 5.0 user's guide. Technical Report SOL 83.20, Stanford University, December 1983.
- [33] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer, New York, 2006. Second Edition.
- [34] A.R. Ribeiro, E.W. Karas, and C.C. Gonzaga. Global convergence of filter methods for nonlinear programming. *SIAM Journal on Optimization* 14 (2008), 646-669.
- [35] S.M. Robinson. A quadratically convergent algorithm for general nonlinear programming problems. *Mathematical Programming* 3 (1972), 145-156.
- [36] D. Robinson and P. Gill. A globally convergent stabilized SQP method. Available at http://www.optimization-online.org/DB_HTML/2012/06/3518.html.
- [37] M.V. Solodov. Constraint qualifications. *Wiley Encyclopedia of Operations Research and Management Science*, J.J. Cochran, ed. John Wiley & Sons, Inc., 2010.
- [38] S.J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Computational Optimization and Applications* 11 (1998), 253-275.
- [39] S.J. Wright. Modifying SQP for degenerate problems *SIAM Journal on Optimization* 13 (2002), 470-497.
- [40] S.J. Wright. Constraint identification and algorithm stabilization for degenerate nonlinear programs. *Mathematical Programming* 95 (2003), 137-160.