To appear in *Optimization Methods and Software*
Vol. 00, No. 00, Month 20XX, 1–24

Some new facts about sequential quadratic programming methods employing second derivatives

A.F. Izmailov[a] and M.V. Solodov[b*]

[a]*Lomonosov Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia; Peoples' Friendship University of Russia, Miklukho-Maklaya Str. 6, 117198 Moscow, Russia; and Department of Mathematic, Physics and Computer Sciences, Derzhavin Tambov State University, TSU, Internationalnaya 33, 392000 Tambov, Russia;* [b]*IMPA – Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil*

(*received October 2015, revised April 2016*)

For the sequential quadratic programming method (SQP), we show that close to a solution satisfying the same assumptions that are required for its local quadratic convergence (namely, uniqueness of the Lagrange multipliers and the second-order sufficient optimality condition), the direction given by the SQP subproblem using the Hessian of the Lagrangian is a descent direction for the standard l_1-penalty function. We emphasize that this property is not straightforward at all, because the Hessian of the Lagrangian need not be positive definite under these assumptions or, in fact, under any other reasonable set of assumptions. In particular, this descent property was not known previously, under any assumptions (even including the stronger linear independence constraint qualification, strict complementarity, etc.). We also check the property in question by experiments on nonconvex problems from the Hock–Schittkowski test collection for a model algorithm. While to propose any new and complete SQP algorithm is not our goal here, our experiments confirm that the descent condition, and a model method based on it, work as expected. This indicates that the new theoretical findings that we report might be useful for full/practical SQP implementations which employ second derivatives and linesearch for the l_1-penalty function. In particular, our results imply that in SQP methods where using subproblems without Hessian modifications is an option, this option has a solid theoretical justification at least on late iterations.

Keywords: sequential quadratic programming; penalty function; descent direction; quadratic convergence

AMS Subject Classification: 90C30; 90C33; 90C55; 65K05

1. Introduction

We consider the constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \quad g(x) \leq 0, \end{array} \tag{1}$$

*Corresponding author. Email: solodov@impa.br

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraints mappings $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are smooth enough (precise smoothness assumptions will be specified later on, as needed).

One of the efficient approaches to solving (1) is that of sequential quadratic programming (SQP). As suggested by the name, SQP methods are based on sequentially approximating the original problem (1) by quadratic programs (QP) of the form

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

where $x^k \in \mathbb{R}^n$ is the current iterate, and H_k is some symmetric $n \times n$-matrix. We refer to the survey papers [1, 9] for relevant discussions, and to [16, Chapters 4, 6] for a comprehensive convergence analysis of methods of this class.

One of the important issues in the SQP context is the choice of the matrix H_k in the subproblem (2). Let $L : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ be the Lagrangian of problem (1):

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

The basic choice is to take in (2) the Hessian of the Lagrangian:

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k), \tag{3}$$

where $(\lambda^k,\, \mu^k) \in \mathbb{R}^l \times \mathbb{R}^m$ is the current dual iterate. A common alternative is to employ some quasi-Newton approximations of the Hessian. As is well understood, both approaches have advantages and disadvantages, whether practical or theoretical. Without going into any lengthy discussion, it is fair to say that using second-order derivatives (when they are affordable to compute) is still of interest, even if in combination with other techniques, or on late iterations only. Some SQP methods using second derivatives can be consulted in the survey papers [1, 9]. We shall briefly comment on the more recent proposals in [10, 11, 18] (we also mention that we discuss here "full-space" methods, and not "composite-step", as in [12] for example). The method of [18] first solves a convex QP (thus not based on second derivatives) to predict the active set, and then solves the resulting equality-constrained QP with the exact Hessian information. The method in [10, 11] first also solves a convex QP and then uses the obtained information to solve a second (inequality-constrained) QP using the exact Hessian. Thus, those (and various other) methods do employ QP subproblems with the second derivatives information, in one way or another.

The goal of this paper is to draw attention to the new fact established here, which is the descent property of the l_1-penalty function in the direction given by SQP subproblem using second derivatives (when close to solutions satisfying certain assumptions). We next discuss some of the issues that arise.

The first comment is that, as is well known and already mentioned above, the Hessian of the Lagrangian may not be positive definite even close to a solution, under any reasonable assumptions. Then one issue, of course, is solving a nonconvex QP itself. Note, however, that our convergence results require computing merely a stationary point, and not a global solution (although, in the presence of inequality constraints, an additional proximity property is required if a stationary point is not unique). The next difficulty is that it is not automatic that a subproblem's stationary point provides a descent direction for the penalty function, as positive definiteness of H_k in (2) is the key in the standard argument showing that the obtained direction is of descent; more on this in Section 3. Our main result precisely shows that, though not automatic because the matrix is indefinite, the descent property in question actually does hold for a direction given by a stationary point of the QP with the exact Hessian, when approaching a solution satisfying the same assumptitons that are required for SQP superlinear convergence. The conditions in question are the uniqueness of the Lagrange multipliers and the second-order sufficient condition for optimality (the sharpest combination of assumptions that is currently known to guarantee fast SQP convergence; see [16, Chapter 4]). That said, we emphasize that the presented theoretical results are not about improving *anything previously known*. The descent property that we report, which is directly relevant for globalizations using the l_1-penalty function, was not known previously under any assumptions (even including the stronger linear independence constraint qualification, strict complementarity, etc.).

The rest of the paper is organized as follows. In Section 2 we recall the conditions needed for local superlinear convergence of the basic SQP method, and provide a new result demonstrating that under these assumptions, the SQP step yields superlinear decrease of the distance to the primal solution (note that this is not automatic simply from superlinear decrease of the distance to the primal-dual solution). This new result would be required for our subsequent analysis. Section 3 is concerned with the interplay bewteen the possible choices of the penalty parameters and the descent properties of SQP directions for the l_1-penalty function. In particular, this section contains our main theoretical result, demonstrating that no modifications of the Hessian of the Lagrangian are needed for the descent property to hold near a primal-dual solution satisfying the same assumptions as those for the local superlinear convergence. In Section 4, we consider the specificities of the case when there are no inequality constraints (only equality constraints), and further strengthen some of the results for this case. In particular, in the equality-constrained case no constraints qualifications are needed for the descent property, and the augmented Lagrangian (in addition to the Lagrangian) choice for the matrix H_k is possible. Finally, in Section 5 we state a model algorithm, and then report on our computational experiments using nonconvex problems from the Hock–Schittkowski collection [13]. We note that to propose here a complete practical SQP method is not our intention; the purpose is to show that the new property can be used to design a convergent scheme in principle, and can thus be incorporated as an option to potentially improve SQP implementations employing second derivatives.

2. On local convergence properties of SQP

We start with stating the local convergence properties of SQP, which in particular highlight the benefits of using the second-order derivative information. We also prove a new estimate for the distance to the primal solution, which would be needed for our subsequent developments. First, some definitions and terminology are in order.

Recall that stationary points of (1) and associated Lagrange multipliers are characterized by the Karush–Kuhn–Tucker (KKT) system:

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

in the variables $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$. As is well known, under any of the appropriate constraint qualifications [21], KKT conditions are necessary for a given point to be a local solution of (1).

The assumption that there exists the unique Lagrange multiplier $(\bar{\lambda},\, \bar{\mu})$ associated with a given stationary point $\bar{x}$, is often referred to as the strict Mangasarian–Fromovitz constraint qualification (SMFCQ). Note that in general, it is weaker than linear independence of active constraints' gradients (LICQ). The following is the standard second-order sufficient optimality condition (SOSC):

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x},\, \bar{\lambda},\, \bar{\mu})\xi,\, \xi \right\rangle > 0 \quad \forall\, \xi \in C(\bar{x}) \setminus \{0\}, \tag{4}$$

where $A(\bar{x}) = \{i = 1, \ldots, m \mid g_i(\bar{x}) = 0\}$ is the set of inequality constraints active at $\bar{x}$, and

$$C(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0,\ g'_{A(\bar{x})}(\bar{x})\xi \leq 0,\ \langle f'(\bar{x}),\, \xi\rangle \leq 0\} \tag{5}$$

is the usual critical cone of problem (1) at $\bar{x}$. Here and throughout, by y_A we denote the subvector of vector y comprised by components indexed by $i \in A$.

The following theorem on local convergence of SQP using the exact Hessian (3) almost literally repeats [16, Theorem 4.14], except for the property (7). The additional property (7) would be needed for our developments later on; it follows from the proof in [16, Theorem 3.6], which is used in [16, Theorem 4.14]. Theorem 2.1 can also be derived using [14, Theorem 3.1]. Note also that Theorem 2.1 is stronger than those results in the SQP literature which employ LICQ instead of SMFCQ. That said, this difference is not the principal point of our subsequent developments. As mentioned in the introduction, our key directional descent results are new even for LICQ or any other assumptions.

Theorem 2.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a local solution of problem* (1)*, satisfying the SMFCQ and the SOSC* (4) *for the associated Lagrange multiplier $(\bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$.*

Then there exists $\delta > 0$ such that for any $\varepsilon_0 > 0$ small enough and any starting point $(x^0,\, \lambda^0,\, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ satisfying

$$\|(x^0 - \bar{x},\, \lambda^0 - \bar{\lambda},\, \mu^0 - \bar{\mu})\| \leq \varepsilon_0,$$

the following assertions are valid:

(a) *There exists a sequence $\{(x^k,\, \lambda^k,\, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ such that for each $k = 0, 1, \ldots,$ the point x^{k+1} is a stationary point of problem* (2) *with H_k given by* (3)*, and $(\lambda^{k+1},\, \mu^{k+1})$ is an associated Lagrange multiplier, satisfying*

$$\|(x^{k+1} - x^k,\, \lambda^{k+1} - \lambda^k,\, \mu^{k+1} - \mu^k)\| \leq \delta. \tag{6}$$

(b) Any such sequence satisfies

$$\|(x^k - \bar{x},\ \lambda^k - \bar{\lambda},\ \mu^k - \bar{\lambda})\| \le \varepsilon_0 \tag{7}$$

for all k, converges to $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$, and the rate of convergence is superlinear.

Moreover, the rate of convergence is quadratic provided the second derivatives of f, h and g are locally Lipschitz-continuous with respect to $\bar{x}$.

Remark 1. In the equality-constrained case (i.e., when there are no inequality constraints in (1)), the first-order optimality conditions for the SQP subproblem (2) give a system of linear equations. Under the assumptions of Theorem 2.1, this system has unique solution (for $(x^k,\ \lambda^k,\ \mu^k)$ in question). The generated iterative sequence is then uniquely defined and therefore, according to Theorem 2.1, this sequence must satisfy the "localization condition" (6). In other words, in this case (6) is automatic and can be dropped. In the general case, (6) defines appropriate "close-by" stationary points, for which the convergence assertions hold. In principle, if stationary points violating (6) exist and one of them is returned by the QP solver, for such a sequence convergence would not be guaranteed.

We emphasize that in the presence of inequality constraints, localization condition (6) is unavoidable for proving convergence, even under any stronger assumptions. We refer the reader to a detailed discussion in [17, Section 5.1], and in particular to [17, Examples 5.1, 5.2] which exhibit that (6) cannot be removed (in the sense that without it, convergence can be lost) even when LICQ, strict complementarity and SOSC are all satisfied (and thus also strong SOSC), which is the strongest set of assumptions possible.

The following Theorem 2.2 is new, though it is related to [7]. It will be used in the sequel to prove that, when close to a solution, Hessian modifications are not needed to guarantee the directional descent property. The key difference with [7] is that the result therein is *a posteriori*: it *assumes* that a sequence generated by the method converges to the given solution. Here, we establish the superlinear decrease of the distance to the primal solution given by the SQP step from an *arbitrary* point close enough to this solution.

Theorem 2.2. *Under the assumptions of Theorem 2.1, there exists $\delta > 0$ such that for any point $(x^k,\ \lambda^k,\ \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$, there exists a stationary point x^{k+1} of problem (2) with H_k given by (3), with an associated Lagrange multiplier $(\lambda^{k+1},\ \mu^{k+1})$ satisfying (6), and for any such x^{k+1} it holds that*

$$x^{k+1} - \bar{x} = o(\|x^k - \bar{x}\|) \tag{8}$$

as $(x^k,\ \lambda^k,\ \mu^k) \to (\bar{x},\ \bar{\lambda},\ \bar{\mu})$.

Proof. Let $\delta > 0$ be chosen as in Theorem 2.1. Then, according to its result, for any $(x^k,\ \lambda^k,\ \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$, there exists a stationary point x^{k+1} of problem (2) with H_k defined in (3), with an associated Lagrange multiplier $(\lambda^{k+1},\ \mu^{k+1})$ satisfying (6), and for any such triple $(x^{k+1},\ \lambda^{k+1},\ \mu^{k+1})$ it holds that $\{(x^{k+1} - x^k,\ \lambda^{k+1} - \lambda^k,\ \mu^{k+1} - \mu^k)\}$ tends to zero as $(x^k,\ \lambda^k,\ \mu^k) \to (\bar{x},\ \bar{\lambda},\ \bar{\mu})$. Therefore, it remains to establish (8).

We argue by contradiction: suppose that there exists a sequence $\{(x^k,\ \lambda^k,\ \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ convergent to $(\bar{x},\ \bar{\lambda},\ \bar{\mu})$, such that for every k there exists $(p^k,\ \tilde{\lambda}^k,\ \tilde{\mu}^k) \in$

$\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ satisfying the relations

$$\begin{array}{c} f'(x^k) + H_k p + (h'(x^k))^{\mathrm{T}}\lambda + (g'(x^k))^{\mathrm{T}}\mu = 0, \quad h(x^k) + h'(x^k)p = 0, \\ \mu \geq 0, \quad g(x^k) + g'(x^k)p \leq 0, \quad \langle \mu,\, g(x^k) + g'(x^k)p \rangle = 0, \end{array} \tag{9}$$

resulting from the KKT system of the SQP subproblem (2), and such that

$$\{(p^k,\, \tilde{\lambda}^k - \lambda^k,\, \tilde{\mu}^k - \mu^k)\} \to (0,\, 0,\, 0) \tag{10}$$

as $k \to \infty$, and

$$\liminf_{k\to\infty} \frac{\|x^k + p^k - \bar{x}\|}{\|x^k - \bar{x}\|} > 0. \tag{11}$$

For each k, using our smoothness assumptions and the boundedness of $\{(\lambda^k,\, \mu^k)\}$, we obtain that

$$\begin{aligned} & \frac{\partial L}{\partial x}(x^k + p^k,\, \lambda^k,\, \mu^k) + (h'(x^{k+1}))^{\mathrm{T}}(\tilde{\lambda}^k - \lambda^k) + (g'(x^{k+1}))^{\mathrm{T}}(\tilde{\mu}^k - \mu^k) \\ = \; & \frac{\partial L}{\partial x}(x^k,\, \lambda^k,\, \mu^k) + \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)p^k \\ & + (h'(x^k))^{\mathrm{T}}(\tilde{\lambda}^k - \lambda^k) + (g'(x^k))^{\mathrm{T}}(\tilde{\mu}^k - \mu^k) + o(\|p^k\|) \\ = \; & f'(x^k) + H_k p^k + (h'(x^k))^{\mathrm{T}}\tilde{\lambda}^k + (g'(x^k))^{\mathrm{T}}\tilde{\mu}^k + o(\|p^k\|) \\ = \; & o(\|p^k\|) \end{aligned} \tag{12}$$

as $k \to \infty$, where (10) was used in the first equality, (3) was used in the second, and the first relation of (9) was used in the last one.

Similarly, using the second relation of (9), we have that

$$h(x^k + p^k) = h(x^k) + h'(x^k)p^k + o(\|p^k\|) = o(\|p^k\|) \tag{13}$$

as $k \to \infty$.

Furthermore, the last line in (9) gives

$$\min\{\tilde{\mu}^k,\, -g(x^k) - g'(x^k)p^k\} = 0. \tag{14}$$

Since

$$\{g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k)\} \to g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(\bar{x}) < 0,$$

this evidently implies that for each k large enough it holds that $\tilde{\mu}^k_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})} = 0$. Also, since

$$\{g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k + p^k)\} \to g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(\bar{x}) < 0,$$

it holds that $g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k + p^k) < 0$ for all k large enough. Hence, for such k,

$$\min\{\tilde{\mu}^k_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})},\, -g_{\{1,\,\ldots,\,m\}\setminus A(\bar{x})}(x^k + p^k)\} = 0. \tag{15}$$

Finally, using again the equality (14) and the nonexpansiveness of the projection operator onto the set $(-\infty, a]$, $a \in \mathbb{R}$, i.e., the property

$$|\min\{a, b\} - \min\{a, c\}| \leq |b - c| \quad \forall\, a, b, c \in \mathbb{R},$$

we obtain that

$$\begin{aligned} |\min\{\tilde{\mu}^k_{A(\bar{x})}, -g_{A(\bar{x})}(x^k + p^k)\}| &= |\min\{\tilde{\mu}^k_{A(\bar{x})}, -g_{A(\bar{x})}(x^k) - g'_{A(\bar{x})}(x^k)p^k + o(\|p^k\|)\} \\ &\quad - \min\{\tilde{\mu}^k_{A(\bar{x})}, -g_{A(\bar{x})}(x^k) - g'_{A(\bar{x})}(x^k)p^k\}| \\ &= o(\|p^k\|) \end{aligned} \tag{16}$$

as $k \to \infty$, where modulus is applied componentwise.

From (12), (13), (15), and (16), by the primal error bound obtained in [7] under SOSC (see also [16, Proposition 1.46]) we derive the estimate

$$x^k + p^k - \bar{x} = o(\|p^k\|)$$

as $k \to \infty$, which means the existence of a sequence $\{t_k\} \in \mathbb{R}_+$ such that $\{t_k\} \to 0$, and for all k it holds that

$$\|x^k + p^k - \bar{x}\| \leq t_k\|p^k\| \leq t_k(\|x^k + p^k - \bar{x}\| + \|x^k - \bar{x}\|).$$

The latter implies that for all k large enough, we have

$$\frac{1}{2}\|x^k + p^k - \bar{x}\| \leq (1 - t_k)\|x^k + p^k - \bar{x}\| \leq t_k\|x^k - \bar{x}\|,$$

contradicting (11). □

3. On descent directions and penalty parameters

Since the basic SQP scheme (2), as any Newtonian method, is guaranteed to converge only locally, it needs to be coupled with some globalization strategy. One well-established technique consists of linesearch in the computed direction $p^k = x^{k+1} - x^k$ for the l_1-penalty function $\varphi_c : \mathbb{R}^n \to \mathbb{R}$,

$$\varphi_c(x) = f(x) + c(\|(h(x)\|_1 + \|\max\{0, g(x)\}\|_1),$$

where $c > 0$ is a penalty parameter, and the max-operation is applied componentwise. The specified direction p^k, with some associated Lagrange multipliers $(\lambda^{k+1}, \mu^{k+1}) \in \mathbb{R}^l \times \mathbb{R}^m$, satisfies (9).

Our objective is to understand (beyond what is known from previous literature) when the SQP direction p^k obtained using the Hessian of the Lagrangian (3) is of descent for the penalty function.

As is well-known, the directional derivatives of the penalty function are given by the following formula (see, e.g., [16, Proposition 6.1]):

$$\begin{aligned}\varphi_c'(x;\,\xi) \;=\;& \langle f'(x),\,\xi\rangle \\ &+c\left(\sum_{j\in J^0(x)}|\langle h_i'(x),\,\xi\rangle| - \sum_{j\in J^-(x)}\langle h_i'(x),\,\xi\rangle + \sum_{j\in J^+(x)}\langle h_i'(x),\,\xi\rangle\right. \\ &\left. + \sum_{i\in I^0(x)}\max\{0,\,\langle g_i'(x),\,\xi\rangle\} + \sum_{i\in I^+(x)}\langle g_i'(x),\,\xi\rangle\right), \end{aligned} \tag{17}$$

where $x,\,\xi\in\mathbb{R}^n$ are arbitrary, and

$$\begin{aligned} J^-(x) &= \{j=1,\,\ldots,\,l \mid h_j(x)<0\},\\ J^0(x) &= \{j=1,\,\ldots,\,l \mid h_j(x)=0\},\\ J^+(x) &= \{j=1,\,\ldots,\,l \mid h_j(x)>0\},\\ I^0(x) &= \{i=1,\,\ldots,\,m \mid g_i(x)=0\},\\ I^+(x) &= \{i=1,\,\ldots,\,m \mid g_i(x)>0\}.\end{aligned}$$

In particular, for any p^k satisfying (9), it can be seen (e.g., [16, Lemma 6.8]) that

$$\begin{aligned}\varphi_c'(x^k;\,p^k) \;\le\;& \langle f'(x^k),\,p^k\rangle - c(\|(h(x^k)\|_1 + \|\max\{0,\,g(x^k)\})\|_1)\\ \le\;& -\langle H_kp^k,\,p^k\rangle \\ &+(\|(\lambda^{k+1},\,\mu^{k+1})\|_\infty - c)(\|(h(x^k)\|_1 + \|\max\{0,\,g(x^k)\})\|_1). \end{aligned}\tag{18}$$

Accordingly, if H_k is positive definite and $p^k\neq 0$, then taking

$$c_k \ge \|(\lambda^{k+1},\,\mu^{k+1})\|_\infty \tag{19}$$

ensures that

$$\varphi_{c_k}'(x^k;\,p^k)\le \Delta_k<0, \tag{20}$$

where

$$\Delta_k = \langle f'(x^k),\,p^k\rangle - c_k(\|(h(x^k)\|_1 + \|\max\{0,\,g(x^k)\})\|_1). \tag{21}$$

In particular, p^k is a direction of descent for φ_{c_k} at the point x^k. Recall, however, that the Hessian of the Lagrangian (3) cannot be expected to be positive definite. This is the crucial issue we would like to address.

If (20) holds, it can be seen in a standard way that for any fixed $\varepsilon\in(0,\,1)$, the following version of the Armijo inequality is satisfied for all $\alpha>0$ small enough:

$$\varphi_{c_k}(x^k+\alpha p^k)\le \varphi_{c_k}(x^k)+\varepsilon\alpha\Delta_k. \tag{22}$$

Thus, one takes the starting trial value $\alpha=1$, and multiplies it by some parameter $\theta\in(0,\,1)$ until the value $\alpha=\alpha_k$ satisfying (22) is obtained. Then, the next primal iterate is (re)defined as $x^{k+1}=x^k+\alpha_kp^k$. Assuming the sequence of matrices H_k is bounded and these matrices are uniformly positive definite, reasonable global convergence properties of the outlined algorithm are obtained (e.g., [16, Theorem 6.9]).

The key issue is that the requirement of positive definiteness of H_k is in contradiction with the Newtonian choice (3), and thus with the potential for fast local convergence. To some extent, this contradiction can be alleviated using quasi-Newton approximations of the Hessian of the Lagrangian. But this also comes with certain disadvantages, including somewhat incomplete convergence theory (e.g., boundedness of the quasi-Newton matrices, such as BFGS, had not been proven). In this work we consider that second derivatives of the problem data (and thus the Hessian of the Lagrangian) are available and affordable. Then, it is at least natural to explore using this information to the fullest extent possible. One basic observation is that positive definiteness of H_k is sufficient but *not necessary* for the SQP direction p^k to be of descent for φ_{c_k} at x^k (with appropriate c_k). Thus, one may well explore using the Hessian *first*, while being aware that this choice of H_k may require modifications. Of course, this idea is not new; see, e.g., the recent proposal in [5]. As a matter of theory, we do prove that under some reasonable assumptions Hessian modifications are, in fact, not necessary to produce descent when close to a solution (despite the matrix not being positive definite). We also confirm this with some experiments on the well-established Hock–Schittkowski test collection [13]. But if the direction given by the Hessian does not work, we consider a model algorithm which modifies H_k adding to it $\tau_k I$ with some $\tau_k > 0$ (other strategies that ensure eventual positive definiteness of H_k can be used [5]). For the new H_k, a new p^k is computed by solving the SQP subproblem (2); this direction is either accepted or not, etc. Evidently, an appropriate p^k will always be obtained after a finite number of modifications (because H_k becomes positive definite). We emphasize that the above strategy is not advertised here as an overall practical algorithm. We are merely interested in tracking the descent property along a trajectory that we know is convergent to a solution. That said, once the descent property is confirmed, it has a potential to be incorporated into practical SQP algorithms with second-order derivatives.

Before proceeding, let us recall the following well-known fact for equality-constrained problems (see [1] and in [19, p. 542]): for sufficiently large c_k, the SQP direction p^k is of descent for φ_{c_k} at x^k if

$$\langle H_k\xi,\, \xi\rangle > 0 \quad \forall\, \xi \in \ker h'(x^k) \setminus \{0\}. \tag{23}$$

This is, of course, a much weaker requirement than positive definiteness of H_k. One question though is for which choice of c_k the descent property holds true. The next example demonstrates that even a strict inequality in (19) is not sufficient.

Example 3.1. Let $n = 2$, $l = 1$, $m = 0$, $f(x) = (x_1^2 - x_2^2)/2$, $h(x) = x_2$. The unique solution of problem (1) with this data is $\bar{x} = 0$, and it satisfies both LICQ and SOSC (4), with the unique associated Lagrange multiplier $\bar{\lambda} = 0$. Moreover, $h'(x^k)$ and H_k given by (3) do not depend on x^k and λ^k, and always satisfy (23).

Here (1) is a QP, and therefore, the SQP subproblem (2) with H_k from (3) coincides with (1). This implies that $p^k = -x^k$, $\lambda^{k+1} = 0$. Take any $x^k \in \mathbb{R}^2$ such that $|x_1^k| < |x_2^k|$ (such points exist arbitrarily close to $\bar{x}$). Then according to (17)

$$\varphi'_{c_k}(x^k;\, p^k) = -(x_1^k)^2 + (x_2^k)^2 - c_k|x_2^k| > 0$$

for every $c_k \geq 0 = \|\lambda^{k+1}\|_\infty$ small enough.

However, fixing $\bar{c} > 0$ and replacing (19) by

$$c_k \geq \|(\lambda^{k+1}, \mu^{k+1})\|_\infty + \bar{c}, \tag{24}$$

we obtain that in the example above the descent condition (20) would hold for all x^k close enough to $\bar{x}$. In fact, (24) is exactly the condition on penalty parameters that is often used in practice.

We next present our main theoretical result. It establishes that choosing the penalty parameter as in (24) ensures that the SQP direction associated to the Hessian of the Lagrangian (3) is of descent for the corresponding penalty function, when close to a solution with the most natural set of properties: SMFCQ and SOSC (in the inequality-constrained case, the computed direction also has to satisfy the localization condition (6), if the subproblem has more than one stationary point). To the best of our knowledge, there are no comparable results in the literature, including under any stronger assumptions (e.g., LICQ instead of SMFCQ, even with strict complementarity in addition, etc.). Moreover, when there are equality constraints only, in Section 4 below we shall prove an even stronger result. In particular, in that case no constraint qualifications of any kind are needed to ensure the descent property, and the localization condition (6) can be dropped; see Remark 1 and Theorem 4.1 below.

Theorem 3.2. *Under the assumptions of Theorem 2.1, for any $\bar{c} > 0$ there exist $\delta > 0$ and $\gamma > 0$ such that for any $(x^k, \lambda^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, there exists a stationary point x^{k+1} of problem (2) with H_k given by (3), with an associated Lagrange multiplier $(\lambda^{k+1}, \mu^{k+1})$ satisfying (6), and for any such x^{k+1} and any c_k satisfying (24), it holds that*

$$\varphi'_{c_k}(x^k;\, p^k) \leq \Delta_k \leq -\gamma\|p^k\|^2, \tag{25}$$

where $p^k = x^{k+1} - x^k$ and Δ_k is given by (21).

Proof. Let $\delta > 0$ be defined as in Theorem 2.2. Then for any (x^k, λ^k, μ^k) close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ there exists a triple $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ satisfying all the requirements specified in the statement of that theorem, and for any such triple it holds that

$$x^k + p^k - \bar{x} = o(\|x^k - \bar{x}\|)$$

as $(x^k, \lambda^k, \mu^k) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$. Then

$$p^k = -(x^k - \bar{x}) + o(\|x^k - \bar{x}\|),$$

which implies that $\{p^k\} \to 0$, $\|x^k - \bar{x}\| = O(\|p^k\|)$, and thus

$$x^k - \bar{x} = -p^k + o(\|p^k\|) \tag{26}$$

as $(x^k, \lambda^k, \mu^k) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$.

We argue by contradiction. Suppose there exists a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and a sequence of reals $\{c_k\}$, such that for each k a triple $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is as specified in the statement of the theorem,

c_k satisfies (24), $p^k \neq 0$, and

$$\lim_{k \to \infty} \frac{\Delta_k}{\|p^k\|^2} \geq 0 \tag{27}$$

(the limit can be $+\infty$). Passing to a subsequence if necessary, without loss of generality we may assume that the sequence $\{p^k/\|p^k\|\}$ converges to some $\xi \in \mathbb{R}^n$, $\|\xi\| = 1$.

Observe first that for all k the second inequality in (18), (21) and (24) imply the estimate

$$\Delta_k \leq -\langle H_k p^k,\, p^k\rangle - \bar{c}(\|(h(x^k)\|_1 + \|\max\{0,\, g(x^k)\})\|_1). \tag{28}$$

If $h'(\bar{x})\xi \neq 0$, then there exists $\tilde{\gamma} > 0$ such that for all k large enough

$$\|h(x^k)\|_1 = \|h'(x^k)p^k\|_1 \geq \tilde{\gamma}\|p^k\|,$$

where the equality is by the the first constraint in (2). Then (28) implies that

$$\Delta_k \leq -\bar{c}\tilde{\gamma}\|p^k\| + O(\|p^k\|^2) \tag{29}$$

as $k \to \infty$, which contradicts (27). Therefore,

$$h'(\bar{x})\xi = 0. \tag{30}$$

Passing to a further subsequence if necessary, without loss of generality we may assume that the index sets $I_{\geq} = I^0(x^k) \cup I^+(x^k)$ and $I_{<} = \{1,\, \ldots,\, m\} \setminus I_{\geq}$ are constant for all k. Observe that it necessarily holds that $I_{\geq} \subset A(\bar{x})$.

Suppose now that there exists $i \in I_{\geq}$ such that $\langle g_i'(\bar{x}),\, \xi\rangle < 0$. Then there exists $\tilde{\gamma} > 0$ such that for all k large enough

$$\begin{aligned} \max\{0,\, g_i(x^k)\} &= g_i(x^k) \\ &= g_i(\bar{x}) + \langle g_i'(\bar{x}),\, x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|) \\ &= \langle g_i'(\bar{x}),\, -p^k\rangle + o(\|p^k\|) \\ &\geq \tilde{\gamma}\|p^k\|, \end{aligned}$$

where the second equality is by (26). Then (28) again implies (29), contradicting (27). Therefore,

$$g'_{I_{\geq}}(\bar{x})\xi \geq 0. \tag{31}$$

Furthermore, for any $i \in I_{<} \cap A(\bar{x})$ in a similar way (employing (26)) we have that

$$0 \geq g_i(x^k) = g_i(\bar{x}) + \langle g_i'(\bar{x}),\, x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|) = \langle g_i'(\bar{x}),\, -p^k\rangle + o(\|p^k\|)$$

as $k \to \infty$. This evidently implies that

$$\langle g_i'(\bar{x}),\, \xi\rangle \geq 0,$$

and therefore,

$$g'_{I_< \cap A(\bar{x})}(\bar{x})\xi \geq 0. \tag{32}$$

Combining (31) and (32) we conclude that

$$g'_{A(\bar{x})}(\bar{x})\xi \geq 0. \tag{33}$$

For any $i \in I_\geq$, the second constraint in (2) implies that for all k it holds that

$$\langle g'_i(x^k),\, p^k\rangle \leq -g'_i(x^k) \leq 0,$$

which evidently implies that

$$\langle g'_i(\bar{x}),\, \xi\rangle \leq 0.$$

Therefore, taking into account (31),

$$g'_{I_\geq}(\bar{x})\xi = 0. \tag{34}$$

Note that

$$-h(x^k) = h'(x^k)p^k = \|p^k\| \left(h'(x^k)\frac{p^k}{\|p^k\|} \right),$$

so that (30) implies that

$$h(x^k) = o(\|p^k\|), \tag{35}$$

as $k \to \infty$. Also, for each $i \in I_\geq$, we have that

$$0 \leq g_i(x^k) \leq -\langle g'_i(x^k), p^k\rangle = -\|p^k\| \left\langle g'_i(x^k),\, \frac{p^k}{\|p^k\|} \right\rangle,$$

so that (34) implies that

$$\max\{0,\, g_{I_\geq}(x^k)\} = o(\|p^k\|), \tag{36}$$

as $k \to \infty$, while for all k it holds that

$$\max\{0,\, g_{I_<}(x^k)\} = 0. \tag{37}$$

From (21), using (35), (36) and (37), we now derive the estimate

$$\Delta_k = \langle f'(x^k),\, p^k\rangle + o(\|p^k\|)$$

as $k \to \infty$. Thus, if $\langle f'(\bar{x}),\, \xi\rangle < 0$, then there exists $\tilde{\gamma} > 0$ such that

$$\Delta_k = -\tilde{\gamma}\|p^k\| + o(\|p^k\|),$$

again contradicting (27). Therefore,

$$\langle f'(\bar{x}),\, \xi\rangle \geq 0. \tag{38}$$

By (5), (30), (33) and (38) we now conclude that $-\xi \in C(\bar{x})$. Therefore, by SOSC (4) it follows that there exists $\tilde{\gamma} > 0$ such that for all k large enough

$$\langle H_k p^k,\, p^k\rangle \geq \tilde{\gamma}\|p^k\|^2. \tag{39}$$

Then (28) implies that

$$\Delta_k \leq -\tilde{\gamma}\|p^k\|^2,$$

which again contradicts (27). □

We next discuss some options for controlling the penalty parameter.

A simple procedure ensuring that (24) holds for all k, and that the parameters are asymptotically constant when the sequence $\{(\lambda^k,\, \mu^k)\}$ is bounded, can be as follows. Fix $\delta > 0$. For $k = 0$ set $c_0 = \|(\lambda^1,\, \mu^1)\|_\infty + \bar{c} + \delta$. For every $k = 1, 2, \ldots$ check (24) for $c_k = c_{k-1}$. If it holds, accept this c_k; otherwise, set $c_k = \|(\lambda^{k+1},\, \mu^{k+1})\|_\infty + \bar{c} + \delta$.

Note that with such a rule, penalty parameters are nondecreasing. In practice, this can be a drawback. Some large value of c_k produced on early iterations (and for a good reason at the time) may be blocking long steps later on, while according to the rule in question c_k cannot be decreased. On the other hand, more moderate values of this parameter might be acceptable at this stage of the iterative process. Of course, there are more sophisticated rules for controlling the penalty parameter than the simple one described above, including those allowing for its decrease, see, e.g., [2, p. 295]. More directly relevant for our purposes are the following considerations.

The definition of Δ_k in (21) suggests to define c_k directly in such a way that (25) holds for p^k at hand. If x^k is not feasible in problem (1), one can always take

$$c_k \geq \frac{\langle f'(x^k),\, p^k\rangle + \gamma\|p^k\|^2}{\|h(x^k)\|_1 + \|\max\{0,\, g(x^k)\}\|_1}, \tag{40}$$

regardless of which matrix H_k was used to compute p^k. However, the right-hand side of this inequality can be unbounded if x^k approaches a nonoptimal feasible point. This makes (40) at least questionable, and certainly not safe. It may be more promising to combine condition (40) with condition (24). This gives

$$c_k \geq \min\left\{\|(\lambda^{k+1},\, \mu^{k+1})\|_\infty + \bar{c},\; \frac{\langle f'(x^k),\, p^k\rangle + \gamma\|p^k\|^2}{\|h(x^k)\|_1 + \|\max\{0,\, g(x^k)\}\|_1}\right\}, \tag{41}$$

where the second term in the min function is $+\infty$ if x^k is feasible and the nominator therein is positive. Observe that (40), and thus also (41), allow negative values of c_k to produce a descent direction for the penalty function with such a parameter. Of course, the exotic option of taking negative penalty parameters does not seem to make a whole lot of sense. In what follows, we replace (41) by the more natural

$$c_k \geq \max \left\{ 0,\, \min \left\{ \|(\lambda^{k+1},\, \mu^{k+1})\|_\infty + \bar{c},\, \frac{\langle f'(x^k),\, p^k \rangle + \gamma \|p^k\|^2}{\|h(x^k)\|_1 + \|\max\{0,\, g(x^k)\}\|_1} \right\} \right\}. \tag{42}$$

It can be easily checked that Theorem 3.2 (as well as Theorem 5.2 further below) remains valid with (24) replaced by (41) or by (42).

Finally, we consider the following rule for penalty parameters. It goes back to [3], and was recently used in [18] (its efficiency for equality-constrained problems is also claimed in [19, p. 542]). The rule is:

$$c_k \geq \frac{\langle f'(x^k),\, p^k \rangle + s \max\{0,\, \langle H_k p^k,\, p^k \rangle\}}{(1-\nu)(\|h(x^k)\|_1 + \|\max\{0,\, g(x^k)\}\|_1)}, \tag{43}$$

where $\nu \in (0,\, 1)$, and $s \in \{0,\, 1/2\}$ is a parameter characterizing two variants. According to (21), if x^k is infeasible then (43) implies the inequality

$$\Delta_k \leq -\nu c_k (\|h(x^k)\|_1 + \|\max\{0,\, g(x^k)\}\|_1) - s \max\{0,\, \langle H_k p^k,\, p^k \rangle\}. \tag{44}$$

In particular, if $c_k > 0$ then (20) holds, implying that p^k is a direction of descent of φ_{c_k} at x^k.

Under the assumption that the matrices H_k are uniformly positive definite, global convergence proofs for various linesearch SQP algorithms employing (43) with $s = 1/2$ can be found in [3], [18]. For the equality-constrained problems, a result of this kind can be found in [4], under the weaker assumption that the matrices H_k are uniformly positive definite on $\ker h'(x^k)$. Moreover, the latter assumption was removed altogether in [5], at the price of using the Hessian modification strategy when certain tests are violated. In addition, [4] and [5] deal with perturbed versions of the algorithm, in which the iterative KKT systems (9) may be solved with some controlled inexactness.

We next obtain a counterpart of Theorem 3.2 for the rule (43) with $s = 1/2$. I.e., we establish that when close to a solution with the usual properties, the SQP direction associated to the exact Hessian of the Lagrangian is of descent for the penalty function with the parameter satisfying (43) with $s = 1/2$. Recall again that in the absence inequality constraints, the localization condition (6) can be dropped from Theorem 3.3; see Remark 1 and Theorem 4.2 below.

Theorem 3.3. *Under the assumptions of Theorem 2.1, for any $\bar{c} > 0$ there exist $\delta > 0$ and $\gamma > 0$ such that for any $(x^k,\, \lambda^k,\, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ close enough to $(\bar{x},\, \bar{\lambda},\, \bar{\mu})$ and such that x^k is infeasible for problem (1), there exists a stationary point x^{k+1} of problem (2) with H_k defined in (3), with an associated Lagrange multiplier $(\lambda^{k+1},\, \mu^{k+1})$ satisfying (6), and for any such x^{k+1} and any $c_k \geq \bar{c}$ satisfying (43) with $s = 1/2$, the inequality (25) is valid, where $p^k = x^{k+1} - x^k$.*

Proof. The argument almost literally repeats the steps outlined for Theorem 3.2, but with (28) replaced by the estimate

$$\Delta_k \leq -\frac{1}{2} \max\{0,\, \langle H_k p^k,\, p^k \rangle\} - \nu \bar{c} (\|h(x^k)\|_1 + \|\max\{0,\, g(x^k)\}\|_1).$$

The latter follows from (44) with $s = 1/2$, and from the assumption that $c_k \geq \bar{c}$. □

Similarly to (40), the rule (43) can be combined with (24) without any harm for the theory, with the motivation to possibly reduce the value of c_k taken by the algorithm for a given matrix H_k. In our numerical experiments in Section 5, we employ the following combination:

$$c_k \geq \max \left\{ 0, \min \left\{ \|(\lambda^{k+1}, \mu^{k+1})\|_\infty + \bar{c}, \frac{\langle f'(x^k), p^k \rangle + \max\{0, \langle H_k p^k, p^k \rangle / 2\}}{(1-\nu)(\|h(x^k)\|_1 + \|\max\{0, g(x^k)\}\|_1)} \right\} \right\}. \tag{45}$$

Finally, in the more special case of constraints being convex, one may try to combine the considerations above with the strategy in [20]. The latter is aimed at making sure that both primal and dual sequences stay bounded (and in particular, penalty parameters stabilize). Also, inexact solution of subproblems using truncation can be considered along the lines in [15].

4. Equality-constrained problems

In this section we discuss some special features of the problem without inequality constraints, in particular improving some previous results and extending them in some directions.

Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0. \end{array} \tag{46}$$

In this case, the Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ is given by

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

and the SQP subproblem takes the form

$$\begin{array}{ll} \text{minimize} & f(x^k) + \langle f'(x^k), x - x^k \rangle + \dfrac{1}{2} \langle H_k(x - x^k), x - x^k \rangle \\ \text{subject to} & h(x^k) + h'(x^k)(x - x^k) = 0. \end{array} \tag{47}$$

Recall also that for equality constraints, the first inequality in (20) always holds as equality for $p^k = x^{k+1} - x^k$, for any stationary point x^{k+1} of problem (47).

When applied to problem (46), Theorems 3.2 and 3.3 can be sharpened and generalized. Specifically, there is no need to assume any constraint qualification (recall that SMFCQ was assumed previously, which in the current setting is the same as LICQ). Moreover, the basic choice

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) \tag{48}$$

(cf. (3)) can be replaced by a wider class of appropriate matrices, whose structure is motivated by replacing the usual Lagrangian by the augmented Lagrangian. Specifically, instead of (48), one can take

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \tilde{\lambda}^k) + \tilde{c}(h'(x^k))^{\mathrm{T}} h'(x^k), \tag{49}$$

where $\tilde{c} \geq 0$ is a parameter, and $\tilde{\lambda}^k$ is defined by

$$\tilde{\lambda}^0 = \lambda^0, \quad \tilde{\lambda}^k = \lambda^k - \tilde{c}h(x^{k-1}), \ k = 1, 2, \ldots. \tag{50}$$

Note that taking $\tilde{c} = 0$, which is allowed, the basic choice (48) is recovered. It is clear that for $\tilde{c} > 0$ a matrix of the structure in (49) has a better choice of being positive definite, at least because the second term in the right-hand side is always positive semidefinite. We note that this choice of H_k also ensures fast convergence of the associated SQP method. Its local superlinear convergence, under the same weak assumptions as in Theorem 2.1, is established in [16, Theorem 4.25].

Let $\mathbb{S}^n$ stand for the set of symmetric $n \times n$-matrices. We have the following improved counterpart of Theorem 3.2, now for a much wider (than the basic (48)) possible choices of the matrices H_k, including in particular the augmented Lagrangian option (49).

THEOREM 4.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be twice differentiable in a neighborhood of $\bar{x} \in \mathbb{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (46), let $\Lambda \subset \mathbb{R}^l$ be a compact subset of the set of Lagrange multipliers associated with $\bar{x}$, and assume that*

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \lambda)\xi, \xi \right\rangle > 0 \quad \forall\, \lambda \in \Lambda, \ \forall\, \xi \in \ker h'(\bar{x}) \setminus \{0\}. \tag{51}$$

Then for any $\Omega : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l \to \mathbb{S}^n$ which is continuous on $\{\bar{x}\} \times \Lambda \times \Lambda$ and such that

$$\langle \Omega(\bar{x}, \lambda, \lambda)\xi, \xi \rangle = 0 \quad \forall\, \lambda \in \Lambda, \ \forall\, \xi \in \ker h'(\bar{x}), \tag{52}$$

and for any $\bar{c} > 0$, there exists $\gamma > 0$ such that for any triple $(x^k, \lambda^k, \tilde{\lambda}^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$ such that the pair (x^k, λ^k) is close enough to $\{\bar{x}\} \times \Lambda$, and $\|\tilde{\lambda}^k - \lambda^k\|$ is small enough, for any stationary point x^{k+1} of problem (47) with H_k given by

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) + \Omega(x^k, \lambda^k, \tilde{\lambda}^k), \tag{53}$$

for any Lagrange multiplier λ^{k+1} associated with this stationary point, and any c_k satisfying

$$c_k \geq \|\lambda^{k+1}\|_\infty + \bar{c}, \tag{54}$$

it holds that

$$\varphi'_{c_k}(x^k; p^k) \leq -\gamma \|p^k\|^2, \tag{55}$$

where $p^k = x^{k+1} - x^k$.

Proof. We argue by contradiction. Suppose that there exist sequences $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ convergent to $\{\bar{x}\} \times \Lambda$ and $\{\tilde{\lambda}^k\} \subset \mathbb{R}^l$ such that $\|\tilde{\lambda}^k - \lambda^k\| \to 0$, and a sequence of reals $\{c_k\}$, such that for all k all the requirements specified in the statement of the theorem are satisfied, $p^k \neq 0$, and

$$\lim_{k\to\infty} \frac{\varphi'_{c_k}(x^k;\, p^k)}{\|p^k\|^2} \geq 0 \tag{56}$$

(the limit can be $+\infty$). Observe that from (18) and (54) it follows that for all k

$$\varphi'_{c_k}(x^k;\, p^k) \leq -\langle H_k p^k,\, p^k\rangle - \bar{c}\|h(x^k)\|_1. \tag{57}$$

Suppose first that

$$\limsup_{k\to\infty} \frac{\|h'(x^k)p^k\|}{\|p^k\|} > 0. \tag{58}$$

Then, passing to a subsequence, if necessary, we obtain the existence of $\tilde{\gamma} > 0$ such that for all k large enough

$$\|h(x^k)\|_1 = \|h'(x^k)p^k\|_1 \geq \tilde{\gamma}\|p^k\|, \tag{59}$$

where the equality follows from the linearized constraint in (47). Since $\bar{x}$ is feasible, we therefore obtain that $\{p^k\} \to 0$, and taking into account continuity of second derivatives of f and h at $\bar{x}$, continuity of Ω on $\{\bar{x}\} \times \Lambda \times \Lambda$, and relation (53), from (57) we derive that

$$\varphi'_{c_k}(x^k;\, p^k) \leq -\bar{c}\tilde{\gamma}\|p^k\| + O(\|p^k\|^2)$$

as $k \to \infty$, which contradicts (56).

It remains to consider the case when

$$\lim_{k\to\infty} \frac{\|h'(x^k)p^k\|}{\|p^k\|} = 0. \tag{60}$$

Without loss of generality we may assume that the sequence $\{p^k/\|p^k\|\}$ converges to some $\xi \in \ker h'(\bar{x})$, $\|\xi\| = 1$. Employing compactness of Λ, continuity of second derivatives of f and h at $\bar{x}$, continuity of Ω on $\{\bar{x}\} \times \Lambda \times \Lambda$, relation (53), and conditions (51) and (52), we then conclude that there exists $\tilde{\gamma} > 0$ such that (39) holds for all k large enough. Hence, from (57) it follows that

$$\varphi'_{c_k}(x^k;\, p^k) \leq -\tilde{\gamma}\|p^k\|^2,$$

which again contradicts (56). □

If for each $(x,\, \lambda,\, \tilde{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$ we define $\Omega(x,\, \lambda,\, \tilde{\lambda})$ as the symmetric matrix of the quadratic form

$$\xi \to \langle \tilde{\lambda} - \lambda,\, h''(x)[\xi,\, \xi]\rangle + \tilde{c}\|h'(x)\xi\|^2 : \mathbb{R}^n \to \mathbb{R},$$

then (52) holds, and H_k defined according to (49) (the augmented Lagrangian option) satisfies (53). Therefore, Theorem 4.1 demonstrates that under its assumptions step 4 of Algorithm 5.1 would not result in modification of the matrix H_k defined according to (49), (50), provided γ is small enough, (x^k, λ^k) is close enough to $(\bar{x}, \bar{\lambda})$, and provided x^{k-1} is also close enough to $\bar{x}$ when $\tilde{c} > 0$ and $k \geq 1$.

Finally, Theorem 3.3 allows for the following counterpart.

THEOREM 4.2. *Under the assumptions of Theorem 4.1, for any* $\Omega : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l \to \mathbb{S}^n$ *which is continuous on* $\{\bar{x}\} \times \Lambda \times \Lambda$ *and such that* (52) *holds, and for any* $\bar{c} > 0$, *there exists* $\gamma > 0$ *such that for any triple* $(x^k, \lambda^k, \tilde{\lambda}^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l$ *such that the pair* (x^k, λ^k) *is close enough to* $\{\bar{x}\} \times \Lambda$, $h(x^k) \neq 0$, *and* $\|\tilde{\lambda}^k - \lambda^k\|$ *is small enough, for any stationary point* x^{k+1} *of problem* (47) *with* H_k *defined in* (53) *and any* $c_k \geq \bar{c}$ *satisfying*

$$c_k \geq \frac{\langle f'(x^k),\, p^k\rangle + \max\{0,\, \langle H_k p^k,\, p^k\rangle/2\}}{(1-\nu)\|h(x^k)\|_1},$$

the inequality (55) *holds, where* $p^k = x^{k+1} - x^k$.

Proof. Similarly to the proof of Theorem 3.2, suppose the contrary: that there exist sequences $\{(x^k, \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l$ convergent to $\{\bar{x}\} \times \Lambda$ and $\{\tilde{\lambda}^k\} \subset \mathbb{R}^l$ such that $\|\tilde{\lambda}^k - \lambda^k\| \to 0$, and a sequence of reals $\{c_k\}$, such that for all k all the requirements specified in the statement of the theorem are satisfied, $h(x^k) \neq 0$ (which combined with constraints in (47) implies that $p^k \neq 0$), and (56) holds. Note that from (44) with $s = 1/2$, and from the assumption that $c_k \geq \bar{c}$, for all k we have the estimate

$$\varphi'_{c_k}(x^k;\, p^k) \leq -\frac{1}{2}\max\{0,\, \langle H_k p^k,\, p^k\rangle\} - \nu\bar{c}\|h(x^k)\|_1. \tag{61}$$

Suppose first that (58) holds. Then, passing to a subsequence if necessary, we obtain the existence of $\tilde{\gamma} > 0$ such that (59) holds for all k large enough. Since $\bar{x}$ is feasible, this implies that $\{p^k\} \to 0$ and, taking into account continuity of second derivatives of f and h at $\bar{x}$, continuity of Ω on $\{\bar{x}\} \times \Lambda \times \Lambda$, and relation (53), from (61) we derive

$$\varphi'_{c_k}(x^k;\, p^k) \leq -\nu\tilde{\gamma}\|p^k\| + O(\|p^k\|^2)$$

as $k \to \infty$, which contradicts (56).

Suppose now that (60) holds, and without loss of generality suppose that the sequence $\{p^k/\|p^k\|\}$ converges to some $\xi \in \ker h'(\bar{x})$, $\|\xi\| = 1$. Employing compactness of Λ, continuity of second derivatives of f and h at $\bar{x}$, continuity of Ω on $\{\bar{x}\}\times\Lambda\times\Lambda$, relation (53), and conditions (51) and (52), we then conclude that there exists $\tilde{\gamma} > 0$ such that (39) holds for all k large enough, and hence, from (61) we have that

$$\varphi'_{c_k}(x^k;\, p^k) \leq -\frac{1}{2}\tilde{\gamma}\|p^k\|^2.$$

This again contradicts (56). □

5. Computational experiments to confirm the descent property

We first state a model algorithm, which has the purpose of using the direction given by the exact Hessian, without modifications, as often as possible. In other words, this option is tried first. As already stated above, we do not necessarily claim that this is a practical algorithm by itself. In our computational experiments, the goal would be to analyze the descent properties of the exact Hessian directions, along the iterations.

In Algorithm 5.1 below, the notation x^{k+1} is used for the computed stationary point of the SQP subproblem, and later also for the next iterate obtained after linesearch; this cannot lead to any ambiguity, however.

ALGORITHM 5.1. Choose $(x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$ and set $k = 0$. Fix the parameters $\bar{c} > 0$, $\gamma > 0$ and ε, $\theta \in (0, 1)$.

(1) Compute the matrix H_k defined by (3). Choose $\tau_k > 0$.
(2) Compute a stationary point x^{k+1} of problem (2) and an associated Lagrange multiplier $(\lambda^{k+1}, \mu^{k+1})$. Set $p^k = x^{k+1} - x^k$.
(3) If $p^k = 0$, stop. Otherwise, choose c_k satisfying (24).
(4) If (25) is violated, replace H_k by $H_k + \tau_k I$ and go to step 2.
(5) Set $\alpha = 1$.
    (a) If the Armijo inequality (22) holds, set $\alpha_k = \alpha$ and go to step 6.
    (b) Replace α by $\theta\alpha$ and go to step 5a.
(6) Reset $x^{k+1} = x^k + \alpha_k p^k$.
(7) Increase k by 1 and go to step 1.

It is clear that any other "convexification" technique (instead of adding multiples of the identity) can be used in Algorithm 5.1, as long as it guarantees that a sufficiently positive definite matrix is produced eventually; this consideration is similar to [18]. For example, the modified Cholesky factorization of H_k can be used [8, Section 4.2], which can be regarded as a "one-step" convexification.

If the technique of adding the identity multiplied by the parameter τ_k is used, ideally this parameter should be in agreement with the estimate of the largest by the absolute value negative eigenvalue of the Hessian of the Lagrangian at (x^k, λ^k, μ^k). However, computing reliable estimates of this kind can be too costly. For problems with nonlinear constraints, we observed that the value in question often strongly depends on $\|(\lambda^k, \mu^k)\|$. After some experimentations, in our numerical results below we employ $\tau_k = 2\max\{1, \|(\lambda^k, \mu^k)\|\}$.

Since $(p^k, \lambda^{k+1}, \mu^{k+1})$ satisfies (9), if $p^k = 0$ for some k (and the algorithm terminates), then x^k is a stationary point of problem (1), while $(\lambda^{k+1}, \mu^{k+1})$ is an associated Lagrange multiplier. When Algorithm 5.1 generates an infinite sequence, its global convergence properties are characterized by the following theorem. The main ingredients of its proof are quite standard (e.g., [16, Theorem 6.9]); thus we merely indicate some steps, for completeness.

THEOREM 5.2. *Let $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be twice differentiable on $\mathbb{R}^n$, with their second derivatives bounded on $\mathbb{R}^n$. Let a sequence $\{(x^k, \lambda^k, \mu^k)\}$ be generated by Algorithm 5.1, and assume that $c_k = c$ for all k large enough, where c is some constant.*

Then as $k \to \infty$, it holds that either

$$\varphi_{c_k}(x^k) \to -\infty, \tag{62}$$

or

$$\{p^k\} \to 0, \quad \left\{ \frac{\partial L}{\partial x}(x^k,\, \lambda^{k+1},\, \mu^{k+1}) \right\} \to 0, \quad \{h(x^k)\} \to 0,$$

$$\{\max\{0,\, g(x^k)\}\} \to 0, \quad \mu_i^{k+1} g_i(x^k) \to 0,\ i = 1,\, \ldots,\, m.$$

In particular, for every accumulation point $(\bar{x},\, \bar{\lambda},\, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ *of the sequence* $\{(x^k,\, \lambda^k,\, \mu^k)\}$*, it holds that* $\bar{x}$ *is a stationary point of problem* (1) *and* $(\bar{\lambda},\, \bar{\mu})$ *is an associated Lagrange multiplier.*

Proof. Observe first that (24) combined with the fact that $c_k = c$ for all k large enough, imply that the sequence $\{(\lambda^k,\, \mu^k)\}$ is bounded. Then by boundedness of the second derivatives of f, h and g there exists $\Gamma > 0$ such that, for all $k = 0,\, 1,\, \ldots$, it holds that

$$-\Gamma\|\xi\|^2 \leq \left\langle \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k)\xi,\, \xi \right\rangle \quad \forall\, \xi \in \mathbb{R}^n, \quad \left\| \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k) \right\| \leq \Gamma.$$

In particular, this implies that the number of times the term $\tau_k I$ (with fixed $\tau_k > 0$) is added to the Hessian of the Lagrangian within step 4 of the algorithm is finite uniformly in k (since after a uniformly finite number of such additions H_k becomes sufficiently positive definite, so that (25) is satisfied). Hence, the sequence $\{H_k\}$ is bounded.

Observe that according to the mean-value theorem, boundedness of second derivatives of f, h and g implies Lipschitz continuity of the gradient of f and of the gradients of components of h and g, say with some constant $\ell > 0$. Repeating literally the argument in [16, Theorem 6.9], one obtains that the sequence $\{\alpha_k\}$ of stepsizes is bounded away from zero. Then from (22) one deduces that $\Delta_k \to 0$ as $k \to \infty$, and all the assertions also follow the same way as in [16, Theorem 6.9]. □

Recalling Remark 1, we note that while Theorem 5.2 establishes global convergence, the local convergence rate assertion only applies if on late iterations the computed subproblems' solutions satisfy the localization condition (6). That said, and as explained in Remark 1, condition (6) is needed only when subproblems' solutions are not unique. In particular, under our assumptions it is not needed in the equality-constrained case. It would also be not needed in the general case under some stronger but still standard assumptions.

It is clear that instead of assuming in Theorem 5.2 boundedness of the second derivatives of the problem data, one could require that the sequence $\{x^k\}$ stay in a compact set (another common assumption in this setting). Similarly, instead of saying that the penalty parameter is asymptotically unchanged, one could ask for boundedness of the dual sequence (and be a bit more specific about choosing the penalty parameter; for example, using (24)).

Due to Theorem 3.3, convergence properties of Theorem 5.2 remain valid for Algorithm 5.1 with condition (24) replaced by (43) with $s = 1/2$, combined with the requirement $c_k \geq \bar{c}$ for some fixed $\bar{c} > 0$. In particular, global convergence properties of the resulting algorithm are still characterized by Theorem 5.2, the proof of which remains valid without any modifications. One only has to observe that when x^k is feasible and the numerator in the right-hand side of (43) is positive, c_k with the needed properties does not exist, in which case one should also modify the Hessian (in practical implementations this should be done whenever c_k exceeds some large upper limit). In this case from (9) we derive that $h'(x^k)p^k = 0$, $\langle \mu^{k+1}, g'(x^k)p^k \rangle = -\langle \mu^{k+1}, g(x^k) \rangle \geq 0$, and hence,

$$\langle f'(x^k), p^k \rangle = -\langle H_k p^k, p^k \rangle - \langle \lambda^{k+1}, h'(x^k)p^k \rangle - \langle \mu^{k+1}, g'(x^k)p^k \rangle \leq -\langle H_k p^k, p^k \rangle.$$

This implies that if H_k is modified so that the inequality $\langle H_k p^k, p^k \rangle > 0$ would eventually be satisfied, the numerator in the right-hand side of (43) will become negative, in which case any c_k can be considered as satisfying this inequality.

Next, we present computational experiments with Algorithm 5.1, using various rules for penalty parameters discussed in Section 3. We stress again that the purpose of this paper is to report the new properties of the penalty function and of the SQP subproblem using second derivatives. Accordingly, our sole goal in this section is to investigate how often the direction given by the exact Hessian of the Lagrangian can actually be used, and to confirm that usually no Hessian modifications are needed when close to a solution. In particular, no comparisons with other codes are in order, for the purposes stated. In any case, by no means we consider our model algorithm as some final/practical product, though it does work reasonably well (we did perform some comparisons with other methods, just to have some idea for our own understanding).

The experiments were performed in Matlab environment (version 7.10.0.499 (R2010a)). The QP subproblems were solved using the `quadprog` routine of the Matlab Optimization Toolbox, with default parameters and the "LargeScale" option set to "off" (thus, it is an active-set algorithm for medium-scale problems). Linear systems of equations were solved by standard Matlab tools. Our test set includes nonconvex problems from the Hock–Schittkowski collection [13], which are available to us in Matlab. This gives 80 problems. We note that for our experiments we have removed all convex problems from the Hock–Schittkowski collection, because in that case the Hessian of the Lagrangian is always at least positive semidefinite, and thus is much more likely to be positive definite, in which case it was already known that the generated direction would be a descent direction. In the considered test set with 80 problems, 15 have equality constraints only. Since second derivatives are not provided in this collection, we compute approximations of the Hessian of the Lagrangian by finite differences (with the step 10^{-10}). Dual sequence is always initialized at zero.

We refer to our implementation of Algorithm 5.1 as `SQP-modH`. Below we report on the behavior of the three versions of Algorithm 5.1, corresponding to three different rules of setting the penalty parameters. In all the versions, the other parameters are as follows: $\bar{c} = \delta = 1$, $\gamma = \nu = 10^{-9}$, $\varepsilon = 0.1$, $\theta = 0.5$. At step 3 of Algorithm 5.1, the version which we call `SQP-modH 1` computes the value $\bar{c}_k$ in the right-hand side of (24), the version `SQP-modH 2` computes the value in (42), and the version `SQP-modH 3` uses (45). If $k = 0$ or $c_{k-1} < \bar{c}_k$, we set $c_k = \bar{c}_k + \delta$; otherwise we keep $c_k = c_{k-1}$. In order to avoid too large values of the penalty parameter, at step 4 of Algorithm 5.1 we check not only (25), but also the inequality $c_k \leq 10^{10}$, and modify H_k when at least one of those two conditions is violated. In addition, we also modify H_k when in the process of backtracking at step 5 the inequality

$$\alpha \|p^k\| < 10^{-8} \tag{63}$$

Table 1. Runs with Hessian modifications (all the other problems required no such modifications).

Test	SQP-modH 1	SQP-modH 2	SQP-modH 3
hs5	7 / 9 (3)	7 / 9 (3)	7 / 9 (3)
hs7	8 / 10 (3)	8 / 9 (1)	7 / 9 (3)
hs9	6 / 7 (1)	6 / 7 (1)	6 / 7 (1)
hs26	17 / 18 (1)	17 / 18 (1)	17 / 18 (1)
hs33	6 / 10 (6)	6 / 10 (6)	6 / 10 (6)
hs39	8 / 9 (1)	8 / 9 (1)	8 / 9 (1)
hs47	17 / 18 (4)	16 / 17 (4)	16 / 17 (4)
hs54	2 / 4 (2)	2 / 4 (2)	2 / 4 (2)
hs56	8 / 9 (1)	38 / 42 (2)	38 / 42 (2)
hs62	6 / 7 (6)	6 / 7 (6)	6 / 7 (6)
hs93	9 / 10 (2)	–	–
hs111	–	343 / 349 (295)	–
hs117	7 / 8 (2)	7 / 8 (2)	7 / 8 (2)

holds true (the primal step is too small), while the stopping criterion

$$\|\Phi(x^k,\, \lambda^{k+1},\, \mu^{k+1})\| < 10^{-6} \tag{64}$$

is not satisfied for the natural residual of the problem, given by $\Phi : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$,

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

Successful runs are those terminated for some $k \leq 1000$ because (64) holds, or because

$$\|\Phi(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})\| < 10^{-6}$$

is satisfied after step 6 of Algorithm 5.1. All other runs are declared failures. Also, if at some iteration 10 modifications of H_k are not enough to generate a suitable descent direction, the run is terminated with a failure.

Our principal interest is to examine how often the direction given by the exact Hessian is actually used. Table 1 lists all the test problems for which *at least one of the algorithms* (`SQP-modH 1`, `SQP-modH 2`, `SQP-modH 3`) *required at least one Hessian modification*, on a successful run. First note that 73 out of 80 problems were successfully solved by all the algorithms. Hessian modifications were needed on about 16% of the problems. In Table 1, for problems where Hessian modifications occurred, we report on the iteration counts and the numbers of QPs solved (separated by slash), and also on the latest iterations at which a Hessian modification was performed. Failures are shown as "–". Thus, the number of times Hessian modifications were performed for a given problem and algorithm is the difference between the second and first number in every column. We see that even for the problems which required Hessian modifications (which are already not so many), the number of modifications is really small. Moreover, there are very few cases when Hessian modifications were needed on late iterations. And we observed that for equality-constrained problems, in our set of experiments this never happened. This supports the claims of Theorems 3.2, 3.3, 4.1, and 4.2. Specifically, modifications on the last iterations were encountered only for `hs33`, `hs54`, and `hs62`. We examined those problems in more detail, to get a better insight.

For `hs33`, the algorithms converge to a nonoptimal stationary point $\bar{x} = (0, 0, 2)$ with the unique associated Lagrange multiplier. This primal-dual KKT point violates even the second-order necessary optimality condition, not only SOSC (4) required in Theorems 3.2 and 3.3. As a result, for all k large enough the iteration QPs (2) with the true Hessian of the Lagrangian as H_k are unbounded. This leads to H_k being modified.

For `hs54`, Hessian modifications are induced by `quadprog` failures, apparently caused by extremely bad scaling of this problem.

For `hs62`, the Hessian modification occurs for $k = 6$ because (63) is satisfied but (64) is violated. With the modified Hessian, (63) remains satisfied, but now (64) becomes valid as well, and the algorithms successfully terminate. Again, this behavior does not contradict Theorems 3.2 and 3.3.

To conclude, consistent with the presented theoretical results, the SQP directions given by the Hessian of the Lagrangian are usually of descent for the l_1-penalty function, when close to a solution. Thus, they can be incorporated into SQP methods that use second-order derivatives, without the need of modifying the Hessian, at least on late iterations. A word of caution is that our numerical experiments are, of course, limited in scope. Problems in the Hock–Schittkowski collection tend to be fairly well behaved, and it is not necessarily the case that on other types of problems Hessian modifications would be quite as rare as observed in our experiments.

Acknowledgment. The authors thank Pavel Izmailov for his help with numerical experiments. The authors are also grateful for constructive comments by the two anonymous referees.

Funding. This research is supported in part by the Russian Foundation for Basic Research Grant 14-01-00113, by the Russian Science Foundation Grant 15-11-10021, by the Volkswagen Foundation, by CNPq Grants 303724/2015-3 and PVE 401119/2014-9, and by FAPERJ.

References

[1] P. Boggs and J. Tolle. Sequential quadratic programming. Acta Numerica 4 (1996), 1–51.

[2] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal. Numerical optimization. Theoretical and practical aspects. 2nd edition. Springer-Verlag, Berlin, 2006.

[3] J.V. Burke and S.P. Han. A robust sequential quadratic programming method. Math. Program. 43 (1989), 277-303.

[4] R.H. Byrd, F.E. Curtis, and J. Nocedal. An inexact SQP method for equality constrained optimization. SIAM J. Optim. 19 (2008), 351-369.

[5] R.H. Byrd, F.E. Curtis, and J. Nocedal. An inexact Newton method for nonconvex equality constrained optimization. Math. Program. 122 (2010), 273-299.

[6] E. Dolan and J. Moré. Benchmarking optimization software with performance profiles Math. Program. 91 (2002), 201–213.

[7] D. Fernández, A.F. Izmailov, and M.V. Solodov. Sharp primal superlinear convergence results for some Newtonian methods for constrained optimization. SIAM J. Optim. 20 (2010), 3312–3334.

[8] P.E. Gill, W. Murrey, and M.H. Wright. Practical Optimization. Academic Press, London, San Diego, 1981.
[9] N.I.M. Gould, D. Orban, and Ph.L. Toint. Numerical methods for large-scale nonlinear optimization. Acta Numerica (2005), 299–361.
[10] N. Gould and D. Robinson. A second derivative SQP method: global convergence. SIAM J. Optim. 20 (2010), 2023–2048.
[11] N. Gould and D. Robinson. A second derivative SQP method: local convergence and practical issues. SIAM J. Optim. 20 (2010), 2049–2079.
[12] M. Heinkenschloss and L.N. Vicente. Analysis of inexact trust-region SQP algorithms. SIAM J. Optim. (2001), 283–302.
[13] W. Hock and K. Schittkowski. Test examples for nonlinear programming codes. Lecture Notes in Economics and Mathematical Systems. V. 187. Springer-Verlag, Berlin, 1981.
[14] A.F. Izmailov and A.S. Kurennoy. Abstract Newtonian frameworks and their applications. SIAM J. Optim. 23 (2013), 2369–2396.
[15] A.F. Izmailov and M.V. Solodov. A truncated SQP method based on inexact interior-point solutions of subproblems. SIAM J. Optim. 20 (2010), 2584–2613.
[16] A.F. Izmailov and M.V. Solodov. Newton-Type Methods for Optimization and Variational Problems. Springer Series in Operations Research and Financial Engineering, Springer International Publishing, Switzerland, 2014.
[17] A.F. Izmailov and M.V. Solodov. Newton-type methods: A broader view. J. Optim. Theory. Appl. 164 (2015), 577–620.
[18] J.L. Morales, J. Nocedal, and Y. Wu. A sequential quadratic programming algorithm with an additional equality-constrained phase. IMA J. Numer. Anal. 32 (2012), 533-579.
[19] J. Nocedal and S.J. Wright. Numerical Optimization. 2nd edition. Springer-Verlag, New York, Berlin, Heidelberg, 2006.
[20] M.V. Solodov. Global convergence of an SQP method without boundedness assumptions on any of the iterative sequences. Math. Program. 118 (2009), 1–12.
[21] M.V. Solodov. Constraint qualifications. In Wiley Encyclopedia of Operations Research and Management Science, J.J. Cochran, ed. John Wiley & Sons, Inc., 2010.