Adv Syst Sci Appl 2022; **02**:73–84
Published online at https://ijassa.ipu.ru.

Accelerating Sequential Quadratic Programming for Inequality-Constrained Optimization near Critical Lagrange Multipliers

Alexey F. Izmailov[1]*, Ivan S. Rodin[1]

[1]*Lomonosov Moscow State University, VMK Faculty, IO Department, Moscow, Russia*

Abstract: We consider a sequential quadratic programming algorithm for optimization problems with equality and inequality constraints, equipped with the standard Armijo linesearch procedure for a nonsmooth exact penalty function, intended for globalization of convergence. We are interested in the case when the standard assumptions for local superlinear convergence of the method may not hold. Specifically, we allow for violation of standard constraint qualifications and second-order sufficient optimality conditions, in which case attraction to so-called critical Lagrange multipliers is known to have a negative impact on convergence rate. In these circumstances, some known acceleration techniques can be expected to take effect only provided the true Hessian and the full SQP step are asymptotically accepted, and these are the main issues addressed in this work. The presented constructions extend some previously known ones to the case when inequality constraints are involved.

Keywords: constrained optimization, Karush–Kuhn–Tucker optimality system, critical Lagrange multiplier, 2-regularity, Newton-type methods, sequential quadratic programming, linesearch globalization of convergence, merit function, nonsmooth exact penalty function, true Hessian, unit stepsize, extrapolation

1. INTRODUCTION

In this work we consider the equality- and inequality-constrained optimization problem

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

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraint mappings $h : \mathbb{R}^n \to \mathbb{R}^l$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are sufficiently smooth, and we are concerned with some crucial peculiarities of performance of Newton-type methods near solutions or stationary points of problem (1.1) that are in a sense singular.

Let $L : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}$ be the Lagrangian of problem (1.1), i.e.,

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

where $\langle \cdot,\, \cdot \rangle$ stands for the Euclidian inner product. Then the stationary points and associated Lagrange multipliers of problem (1.1) are characterized by the Karush–Kuhn–Tucker (KKT) optimality system

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

*Corresponding author: izmaf@cs.msu.ru

For a feasible point $\bar{x}$ of problem (1.1), the linear independence constraint qualification (LICQ) consists of saying that the gradients $h_j'(\bar{x})$, $j \in \{1, \ldots, l\}$, $g_i'(\bar{x})$, $i \in A(\bar{x})$, are linearly independent, where $A(\bar{x}) = \{i \in \{1, \ldots, m\} \mid g_i(\bar{x}) = 0\}$ is the set of indices of inequality constraints active at $\bar{x}$. If $\bar{x}$ is a local solution of (1.1), satisfying LICQ, then there exists the unique pair $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$ such that the triple $(\bar{x}, \bar{\lambda}, \bar{\mu})$ solves (1.2).

In this work, we are mostly interested in the case when LICQ does not hold, but $\bar{x}$ is stationary for (1.1) with some (possibly nonunique) associated Lagrange multiplier. In this case, the set of Lagrange multipliers may naturally contain some special instances, called critical multipliers (to be defined formally in Section 2 below), that are known to strongly attract dual sequences of various primal-dual optimization algorithms, and this phenomenon has a strong negative impact on convergence rate; see [14–16] and the summaries of this research in [18, Section 7.1] and [19].

The effect of attraction to critical multipliers can be locally avoided by using some dual stabilization mechanisms, like the one of the stabilized sequential quadratic programming method; see the original proposals in [22, 24], as well as a more recent treatment in [17] and [18, Section 7.2.2]. However, even such modified algorithms typically have large domains of convergence to critical multipliers when they exist [11], and this becomes even more of an issue when globalization of convergence is concerned.

Generally, there can be at least two different approaches to the specified criticality issue. One is indeed to try to avoid convergence to a critical multiplier, and some further tools for this were developed very recently in [5]. The essence of an alternative approach, adopted in this work, can be expressed as follows: since it is difficult to avoid convergence to critical multipliers, this convergence can be accelerated by employing the knowledge of its rather special character. This point will be explained in Section 2, first for generic systems of nonlinear equations and for optimization problems with equality constraints only, using the results from [6–9] and [10], respectively.

One of the outcomes of this analysis is a very simple extrapolation strategy for accelerating convergence of the basic sequential quadratic programming (SQP) method to critical Lagrange multipliers, that can be easily incorporated into globalization schemes based on linesearch. The main goal of this paper is to extend this approach to optimization problems involving inequality constraints as well, and this is accomplished in Section 3. Section 4 presents some numerical results confirming the use of the approach, and Section 5 contains some concluding remarks and discussion of directions for further research.

Some words about our notation and terminology. The Euclidian (l_2), l_1, and l_∞ norms will be denoted by $\|\cdot\|$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$, respectively. For a function $\varphi : \mathbb{R}^n \to \mathbb{R}$, by $\varphi'(x;\, \xi)$ we denote the standard directional derivative of φ at $x \in \mathbb{R}^n$ in a direction $\xi \in \mathbb{R}^n$. The null space and the range space of a linear operator $\Lambda : \mathbb{R}^n \to \mathbb{R}^l$ are denoted by $\ker \Lambda$ and $\operatorname{im} \Lambda$, respectively. For a given $z \in \mathbb{R}^m$ and an index set $I \subset \{1, \ldots, m\}$, the symbol z_I stands for the subvector of z with components corresponding to $i \in I$.

A set $U \subset \mathbb{R}^p$ is called starlike with resect to $\bar{u} \in \mathbb{R}^p$ if for every $u \in U$ and $t \in (0,\, 1]$, it holds that $tu + (1-t)\bar{u} \in U$. Any $v \in \mathbb{R}^p$ is called an excluded direction for a set U starlike with respect to $\bar{u}$ if $\bar{u} + tv \notin U$ for all $t > 0$. A set which is starlike with respect to a given point is called asymptotically dense if the complement of the corresponding set of excluded directions is open and dense.

2. NONLINEAR EQUATIONS AND EQUALITY-CONSTRAINED OPTIMIZATION

Consider a system of nonlinear equations

$$\Phi(u) = 0 \tag{2.3}$$

with a smooth mapping $\Phi : \mathbb{R}^p \to \mathbb{R}^p$. A solution $\bar{u}$ of (2.3) is called (non)singular if the Jacobian $\Phi'(\bar{u})$ is a (non)singular matrix. For a given current iterate $u^k \in \mathbb{R}^p$, the basic

Newton method (NM) defines the next iterate as $u^{k+1} = u^k + v^k$, where v^k is a solution of the linearized (at u^k) equation

$$\Phi(u^k) + \Phi'(u^k)v = 0. \tag{2.4}$$

The classical theory says that when $u^0 \in \mathbb{R}^p$ is taken close enough to a nonsingular solution of (2.3), then the NM defined this way generates the unique sequence of iterates $\{u^k\}$, and this sequence converges to $\bar{u}$ superlinearly; see, e.g., [18, Section 2.1.1].

Here we are mainly interested in the cases of when the solution in question can be singular, perhaps even nonisolated. To that end, we make use of the following second-order regularity concept assuming that Φ is twice differentiable at $\bar{u}$: the mapping Φ is said to be 2-regular at $\bar{u}$ in a direction $v \in \mathbb{R}^p$ if

$$\operatorname{im} \Phi'(\bar{u}) + \Phi''(\bar{u})[v, \ker \Phi'(\bar{u})] = \mathbb{R}^p.$$

If $\Phi'(\bar{u})$ is nonsingular, then Φ is 2-regular at $\bar{u}$ in every direction, including $v = 0$. At the same time, 2-regularity may hold at singular (and even nonisolated) solutions in nonzero directions, including those from $\ker \Phi'(\bar{u})$. The following theorem summarizes the results on local convergence of the basic NM, obtained in [8, Theorem 6.1] and [9, Theorem 2.1].

Theorem 2.1:
Let $\Phi : \mathbb{R}^p \to \mathbb{R}^p$ be twice differentiable near $\bar{u} \in \mathbb{R}^p$, with its second derivative Lipschitz-continuous with respect to $\bar{u}$, that is,

$$\Phi''(u) - \Phi''(\bar{u}) = O(\|u - \bar{u}\|)$$

as $u \to \bar{u}$. Let $\bar{u}$ be a singular solution of equation (2.3)*, and assume that there exists $\bar{v} \in \ker \Phi'(\bar{u})$ such that Φ is 2-regular at $\bar{u}$ in this direction.*

Then there exists a set $U \subset \mathbb{R}^p$ starlike with respect to $\bar{u}$, asymptotically dense at $\bar{u}$, and such that for every starting point $u^0 \in U$, there exists the unique sequence $\{u^k\} \subset \mathbb{R}^p$ such that for all k, the step $v^k = u^{k+1} - u^k$ is the solution of (2.4)*, and $u^k \neq \bar{u}$, $\{u^k\}$ converges to $\bar{u}$,*

$$\lim_{k\to\infty} \frac{\|u^{k+1} - \bar{u}\|}{\|u^k - \bar{u}\|} = \frac{1}{2}, \tag{2.5}$$

and the sequence $\{(u^k - \bar{u})/\|u^k - \bar{u}\|\}$ converges to some $v \in \ker \Phi'(\bar{u})$.

The convergence pattern specified by (2.5) suggests that acceleration of convergence can be achieved by doubling the basic NM step. This, however, should be done with due care, as simply doubling each step may force the resulting sequence to leave the domain of convergence, which is asymptotically dense but need not contain a full neighborhood of the solution in question. To that end, somehow more involved acceleration techniques have been developed in [7,9], such as extrapolation and overrelaxation. The simplest extrapolation technique consists of generating, along with the basic NM sequence $\{u^k\}$, an auxiliary sequence $\{\widehat{u}^k\}$ obtained by doubling the NM step: for each k compute

$$\widehat{u}^{k+1} = u^k + 2v^k. \tag{2.6}$$

According to [9, Theorem 4.1], under the assumptions of Theorem 2.1 above, this auxiliary sequence $\{\widehat{u}^k\}$ converges linearly to $\bar{u}$, with the asymptotic ratio of $1/4$ (instead of $1/2$ for $\{u^k\}$, given by (2.5)), from all points in the domain of convergence of NM sequences $\{u^k\}$.

We now turn our attention to globalization issues, and recall that convergence of the NM can be globalized in standard ways, e.g., by the Armijo linesearch procedure for the residual $\|\Phi(\cdot)\|$ used as a merit function; see [18, Section 5.1.1]. The following is a prototype algorithm of this kind.

Algorithm 2.1:
Fix the parameters $\sigma \in (0,\, 1)$ and $\theta \in (0,\, 1)$. Choose $u^0 \in \mathbb{R}^p$, and set $k = 0$.

1. Compute $v^k \in \mathbb{R}^p$ solving the linear system (2.4). If (2.4) cannot be solved, stop with failure.
2. Set $\alpha = 1$. If the inequality
$$\|\Phi(u^k + \alpha v^k)\| \leq (1 - \sigma\alpha)\|\Phi(u^k)\| \tag{2.7}$$
holds, set $\alpha_k = \alpha$ and go to Step 3. Otherwise, keep replacing α by $\theta\alpha$ until (2.7) is satisfied.
3. Set $u^{k+1} = u^k + \alpha_k v^k$, increase k by 1, and go to Step 1.

According to [18, Theorem 5.3], this algorithm converges globally in a sense that any accumulation point $\bar{u}$ of any sequence $\{u^k\}$ generated by this algorithm satisfies
$$(\Phi'(\bar{u}))^\top \Phi(\bar{u}) = 0.$$
Moreover, according to [18, Theorem 5.4], Step 2 of Algorithm 2.1 accepts the unit stepsize when u^k is close to a nonsingular solution, thus allowing the globalized algorithm to inherit the superlinear convergence rate of the basic NM.

Near singular solutions, the crucial issue of asymptotic acceptance of the unit stepsize is much more involved. The following result was obtained in [6, Theorem 1].

Theorem 2.2:
Under the assumptions of Theorem 2.1, a set U in it can be chosen so that for every starting point $u^0 \in U$, Algorithm 2.1 with $\sigma \in (0,\, 3/4)$ uniquely defines the sequence $\{u^k\}$, and $\alpha_k = 1$ is accepted at Step 2 of this algorithm for all k large enough.

Combining Theorems 2.1 and 2.2, one can conclude that once a sequence generated by Algorithm 2.1 enters the domain U specified in those theorems, this sequence converges linearly to $\bar{u}$ with the asymptotic ratio of $1/2$. Moreover, for the "extrapolated" sequence $\{\widehat{u}^k\}$ generated according to (2.6), the asymptotic ratio is $1/4$. It is important to note that the main iterative sequences $\{u^k\}$ are not affected at all by computing the auxiliary sequences $\{\widehat{u}^k\}$, and hence, doing so does not affect the global convergence properties of the algorithm. It should also be mentioned that generating the auxiliary sequence $\{\widehat{u}^k\}$ costs essentially nothing.

Getting back to optimization, we first consider the case when there are no inequality constraints in (1.1) (i.e., $m = 0$):
$$\text{minimize} \quad f(x) \quad \text{subject to} \quad h(x) = 0. \tag{2.8}$$
In this case, the KKT system (1.2) can be written as (2.3) with $p = n + l$, $\Phi : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}^n \times \mathbb{R}^l$,
$$\Phi(u) = \left(\frac{\partial L}{\partial x}(x,\, \lambda),\, h(x)\right), \tag{2.9}$$
where $u = (x,\, \lambda)$, and the missing argument μ of L is dropped. Algorithm 2.1, as well as its version with extrapolation, and Theorems 2.1 and 2.2 are certainly applicable to this special case of a nonlinear equation (2.3). The problem, however, is that in the optimization context, using the residual of a first-order optimality system as a merit function can hardly be justified: a merit function should be reflecting the intention to find a solution of the optimization problem rather than just any stationary point of it.

One typical choice of an optimization-related merit function for problem (2.8) is the l_1 exact penalty function $\varphi_c : \mathbb{R}^n \to \mathbb{R}$,
$$\varphi_c(x) = f(x) + c\|h(x)\|_1,$$

where $c > 0$ is a penalty parameter; see [1, Section 17], [23, Section 18.4], [18, Section 6.2]. Observe, however, that after this change of the merit function, Theorem 2.2 is no more applicable, and the issue of asymptotically accepting the unit stepsize becomes even more involved, in particular, due to inherent nonsmoothness of φ_c, and to the presence of a penalty parameter c. Another source of complications is that the primal part ξ^k of the NM step $v^k = (\xi^k, \eta^k)$ computed at the current iterate $u^k = (x^k, \lambda^k)$ can be guaranteed to be a direction of descent for φ_c (with an appropriate choice of c) at u^k only provided the Hessian of the Lagrangian

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k) \tag{2.10}$$

is positive definite, which is not at all automatic even close to a solution/multiplier pair satisfying the second-order sufficient optimality condition. To that end, the algorithm for problem (2.8), presented next, has Step 3, where the sufficient descent property of the generated primal direction is verified.

Algorithm 2.2:
Fix the parameters $\bar{c} > 0$, $\widetilde{c} > 0$, $\rho > 0$ and $\sigma, \theta \in (0, 1)$. Choose $u^0 = (x^0, \lambda^0) \in \mathbb{R}^n \times \mathbb{R}^l$, and set $c_{-1} = 0$ and $k = 0$.

1. Compute $v^k = (\xi^k, \eta^k)$ solving the linear system
$$\frac{\partial L}{\partial x}(x^k, \lambda^k) + H_k\xi + (h'(x^k))^\top \eta = 0, \quad h(x^k) + h'(x^k)\xi = 0, \tag{2.11}$$
with H_k given by (2.10). If (2.11) cannot be solved, stop with failure.
2. Set
$$c_k = \max\left\{c_{k-1}, \frac{4(1-\sigma)\|\lambda^k + \eta^k\|_\infty + \|\lambda^k\|_\infty}{3 - 4\sigma} + \bar{c}\right\}. \tag{2.12}$$
If maximum in (2.12) is attained at the second argument, replace c_k by $c_k + \widetilde{c}$.
3. If the inequality
$$\varphi'_{c_k}(x^k; \xi^k) \leq -\rho\|\xi^k\|^2 \tag{2.13}$$
is violated, stop with failure.
4. Set $\alpha = 1$. If the inequality
$$\varphi_{c_k}(x^k + \alpha\xi^k) \leq \varphi_{c_k}(x^k) + \sigma\alpha\varphi'_{c_k}(x^k; \xi^k) \tag{2.14}$$
holds, set $\alpha_k = \alpha$ and go to Step 5. Otherwise, keep replacing α by $\theta\alpha$ until (2.14) is satisfied.
5. Define $u^{k+1} = (x^{k+1}, \lambda^{k+1})$ as $u^{k+1} = u^k + \alpha_k v^k$, increase k by 1, and go to Step 1.

Observe that (2.11) with H_k given by (2.10) is nothing else but the NM equation (2.4) with $v = (\xi, \eta)$, for Φ defined in (2.9). On the other hand, this method can be seen as the SQP algorithm, since (2.11) is the Lagrange optimality system for the equality-constrained quadratic programming subproblem

$$\text{minimize} \quad \langle f'(x^k), \xi\rangle + \frac{1}{2}\langle H_k\xi, \xi\rangle \quad \text{subject to} \quad h(x^k) + h'(x^k)\xi = 0; \tag{2.15}$$

for details see, e.g., [18, Section 4.2].

Asymptotic acceptance of the true Hessian and unit stepsize by Algorithm 2.2 has been investigated very recently in [10]. In order to present the main result of that work, some more terminology will be needed. According to [11, Proposition 1] and [12, Proposition 2], the assumptions of Theorem 2.1 for Φ defined in (2.9) may hold at $\bar{u} = (\bar{x}, \bar{\lambda})$ only if $\bar{\lambda}$ is a critical Lagrange multiplier associated with a stationary point $\bar{x}$ of problem (2.8). Criticality of $\bar{\lambda}$ means that the linear subspace

$$Q(\bar{x}, \bar{\lambda}) = \left\{ \xi \in \ker h'(\bar{x}) \;\middle|\; \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi \in \operatorname{im}(h'(\bar{x}))^\top \right\} \tag{2.16}$$

is nontrivial; see [18, Definition 1.41]. In particular, the assumptions of Theorem 2.1 do hold under the following requirements:

- The multiplier $\bar{\mu}$ is critical of order 1, which means that $\dim Q(\bar{x}, \bar{\lambda}) = 1$, or, in other terms,
$$Q(\bar{x}, \bar{\lambda}) = \operatorname{span}\{\bar{\xi}\} \tag{2.17}$$
with some $\bar{\xi} \in \mathbb{R}^n \setminus \{0\}$.
- It holds that
$$\operatorname{rank} h'(\bar{x}) = l - 1 \tag{2.18}$$
and
$$h''(\bar{x})[\bar{\xi}, \bar{\xi}] \notin \operatorname{im} h'(\bar{x}). \tag{2.19}$$

The following counterpart of Theorem 2.2 for equality-constrained optimization problems appears in [10, Theorem 3.1].

Theorem 2.3:
Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be three times differentiable near $\bar{x} \in \mathbb{R}^n$, with their third derivatives Lipschitz-continuous with respect to $\bar{x}$. Let $\bar{x}$ be a stationary point of problem (2.8)*, with an associated Lagrange multiplier $\bar{\lambda} \in \mathbb{R}^l$, and assume that* (2.17) *with $Q(\bar{x}, \bar{\lambda})$ defined in* (2.16), (2.18)*, and* (2.19) *hold with some $\bar{\xi} \in \mathbb{R}^n \setminus \{0\}$. Assume also that $h'(\bar{x}) \neq 0$, and set $\bar{u} = (\bar{x}, \bar{\lambda})$.*

Then there exist $\rho > 0$ and a set $U \subset \mathbb{R}^n \times \mathbb{R}^l$ starlike with respect to $\bar{u}$, asymptotically dense at $\bar{u}$, and such that for every starting point $u^0 \in U$, Algorithm 2.2 with $\sigma \in (0, 3/4)$ uniquely defines the sequence $\{u^k\}$, this sequence converges to $\bar{u}$, and $\alpha_k = 1$ is accepted at Step 4 of this algorithm for all k large enough.

Getting back to a general optimization problems involving inequality constraints, we recall that according to [18, Definition 7.8], a Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ associated with a stationary point $\bar{x}$ of problem (1.1) is called critical with respect to an index set $A \subset A(\bar{x})$ if $\bar{\mu}_{A(\bar{x})\setminus A} = 0$ and $(\bar{\lambda}, \bar{\mu}_A)$ is a critical Lagrange multiplier associated with the stationary point $\bar{x}$ of the equality-constrained optimization problem

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad h(x) = 0, \; g_A(x) = 0. \tag{2.20}$$

As discussed in [15] and [18, Section 7.1.2], the effect of criticality on performance of primal-dual algorithms shows up precisely this way, through stabilization of specific activity patterns. This is confirmed by the numerical experiments discussed in Section 4 below.

3. OPTIMIZATION WITH INEQUALITY CONSTRAINTS

Unlike Algorithms 2.1 and 2.2, the algorithm for problem (1.1) that we present next is not just a prototype algorithm in the following sense: it is supplied by a mechanism intended to avoid unnecessary terminations with failure, especially far from solutions, and it is exactly what we used in numerical experiments in Section 4. More precisely, the issue of possible infeasibility of the iteration subproblem (3.23) (a counterpart of (2.15) involving linearized inequality constraints) is not addressed in this algorithm, and it still stops with failure in such cases, as this is a general issue concerned with SQP methods, and there are known tools for tackling it; see, e.g., [18, Section 6.2]. However, when the iteration subproblem is feasible but cannot be solved, or when the generated primal direction ξ^k does not pass the sufficient descent test (3.25) (a counterpart of (2.13)), the algorithm does not stop with failure, but rather sequentially adjusts the basic choice of

$$H_k = \frac{\partial^2 L}{\partial x^2}(x^k,\, \lambda^k,\, \mu^k) \tag{3.21}$$

in such a way that eventually the needed ξ^k satisfying (3.25) is always found.

For every value of the penalty parameter $c > 0$, we now need to re-define the penalty function $\varphi_c : \mathbb{R}^n \to \mathbb{R}$ in order to incorporate the inequality constraints:

$$\varphi_c(x) = f(x) + c(\|h(x)\|_1 + \|\max\{0,\, g(x)\}\|_1). \tag{3.22}$$

Algorithm 3.1:
Fix the parameters $\bar{c} > 0$, $\widetilde{c} > 0$, $\rho > 0$ and $\sigma,\, \theta \in (0,\, 1)$. Choose $u^0 = (x^0,\, \lambda^0,\, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l$, and set $c_{-1} = 0$ and $k = 0$.

1. Define H_k according to (3.21).
2. Compute ξ^k as a stationary point of the quadratic programming problem

$$\begin{aligned} &\text{minimize} \quad \langle f'(x^k),\, \xi\rangle + \frac{1}{2}\langle H_k\xi,\, \xi\rangle \\ &\text{subject to} \quad h(x^k) + h'(x^k)\xi = 0,\ g(x^k) + g'(x^k)\xi \leq 0, \end{aligned} \tag{3.23}$$

   and $(\lambda^{k+1},\, \mu^{k+1})$ as an associated Lagrange multiplier. If (3.23) is infeasible, stop with failure. If (3.23) is feasible, but a stationary point of (3.23) cannot be found, go to Step 5. Otherwise, set $\eta^k = \lambda^{k+1} - \lambda^k$, $\zeta^k = \mu^{k+1} - \mu^k$, $v^k = (\xi^k,\, \eta^k,\, \zeta^k)$.
3. Set

$$c_k = \max\left\{c_{k-1},\, \frac{4(1-\sigma)\|(\lambda^{k+1},\, \mu^{k+1})\|_\infty + \|(\lambda^k,\, \mu^k)\|_\infty}{3 - 4\sigma} + \bar{c}\right\}. \tag{3.24}$$

   If max in (3.24) is attained at the second argument, replace c_k by $c_k + \widetilde{c}$. Set

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

4. If the inequality

$$\Delta_k \leq -\rho\|\xi^k\|^2 \tag{3.25}$$

   is satisfied, go to Step 6.
5. Choose $\tau_k > 0$, replace H_k by $H_k + \tau_k I$, and go to Step 2.
6. Set $\alpha = 1$. If the inequality

$$\varphi_{c_k}(x^k + \alpha\xi^k) \leq \varphi_{c_k}(x^k) + \sigma\alpha\Delta_k \tag{3.26}$$

   holds with φ_{c_k} defined according to (3.22), set $\alpha_k = \alpha$ and go to Step 7. Otherwise, keep replacing α by $\theta\alpha$ until (3.26) is satisfied.
7. Define $u^{k+1} = (x^{k+1},\, \lambda^{k+1},\, \mu^{k+1})$ as $u^k + \alpha_k v^k$, increase k by 1, and go to Step 1.

In Step 7 of this algorithm, it is quite typical to use the stepsize parameter in primal updates only; see, e.g., [18, Algorithm 6.7]. However, the theory developed in [10] for the equality-constrained case requires using the stepsize parameter in dual updates as well. This variant of linesearch SQP algorithms is also quite common; see, e.g., [23, Algorithm 18.3]. Theoretical global convergence properties of this kind of algorithms, as well as their rate of convergence properties under "standard" assumptions were recently discussed in [20]; see also references therein.

Here, however, we are concerned with the cases when those "standard" assumptions for fast local convergence may fail to hold, and in particular, with the cases of convergence to multipliers critical with respect to a relevant index set. Motivated by its success in the equality-constrained case, our proposal being tested in this work is to generate an extrapolated auxiliary sequence $\{\widehat{u}^k\}$ according to (2.6), with expectation that its convergence will be faster than that of the main sequence $\{u^k\}$.

4. NUMERICAL EXPERIMENTS

Algorithm 3.1 will be abbreviated as SQP-mH, while its variant supplied with extrapolation, and hence, generating an auxiliary sequence $\{\widehat{u}^k\}$ according to (2.6), will be abbreviated as SQP-mH-EP. Note that (2.6) always employs the direction v^k obtained by the basic choice of H_k specified in (3.21). If for some k such v^k cannot be defined, we put $\widehat{u}^{k+1} = u^{k+1}$.

Instead of modifying the Hessian of the Lagrangian, possible lack of positive definiteness of the true Hessian can be tackled by defining H_k as quasi-Newton approximations of the Hessian. A standard choice adopted here consists of using BFGS approximations complemented by Powell's correction at Step 1 of Algorithm 3.1 instead of (3.21); see [18, Section 4.1] for details. Furthermore, in the resulting algorithm abbreviated as SQP-BFGS, Steps 4 and 5 of Algorithm 3.1 are dropped, and the stepsize parameter is used for primal updates only, i.e., $\lambda^{k+1} = \lambda^k + \eta^k$ and $\mu^{k+1} = \mu^k + \zeta^k$.

The parameter values used in our computations were as follows: $\bar{c} = \widetilde{c} = 1$, $\rho = 10^{-9}$, $\sigma = 0.01$, $\theta = 0.5$. Furthermore, we adopted the following rule for controlling τ_k at Step 5 of Algorithm 3.1: for every given k, when Step 5 is invoked for the first time, we set $\tau_k = 1$, and then multiply it by 2 every next time Algorithm 3.1 invokes Step 5.

Runs of algorithms SQP-BFGS and SQP-mH were terminated with success if a newly generated iterate u^k satisfied

$$\|\Phi(u^k)\| \leq 10^{-6}. \tag{4.27}$$

For SQP-mH-EP, for every $k = 1, 2, \ldots$ we first compute $\widehat{u}^k$, and terminate the run with success if

$$\|\Phi(\widehat{u}^k)\| \leq 10^{-6}; \tag{4.28}$$

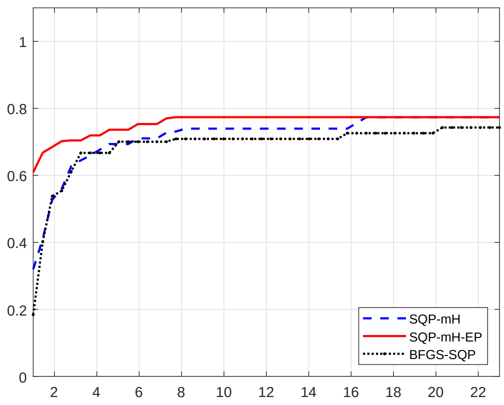
otherwise, we proceed with computing u^k and verifying (4.27). When successful termination did not occur after 200 iterations, or the backtracking procedure at Step 6 of Algorithm 3.1 produced a trial value α such that $\alpha\|v^k\| \leq 10^{-10}$, the process was terminated with failure.

The experiments were performed in Matlab environment, with PATH solver [3, 4] used for quadratic programming subproblems.

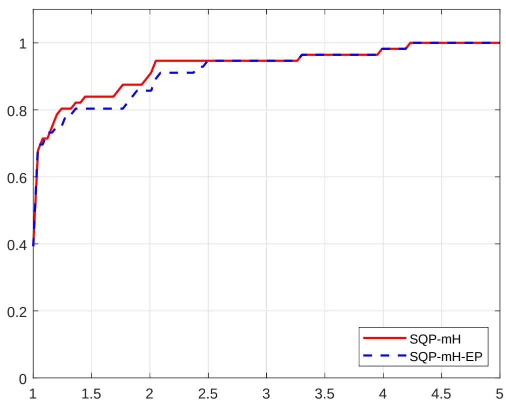
Problem instances for the experiments were taken from DEGEN test collection [13] that contains optimization problems with constraints violating LICQ. We employed all problems from DEGEN involving inequality-constraints (numerical results for purely equality-constrained problems can be found in [10]), except for those where the solutions of interest is not a stationary point (problems 20112, 20220, 20224, 30212 and 30302). This leaves 59 test problems.

For each of those problems, all algorithms being tested were initialized at the same 100 starting points $u^0 = (x^0, \lambda^0, \mu^0)$ generated randomly in the l_∞-ball of radius 100, centered at the primal solution of interest (reported in DEGEN) for the primal part x^0, and at $(0, 0)$ for the dual part (λ^0, μ^0), with the additional nonnegativity restriction on μ^0.

As measures of efficiency, we used the average number of quadratic programming subproblems solved and the average number of evaluations of the objective function, constraint mappings, and their derivatives, per one successful run. These results are presented in the form of performance profiles [2]. For DEGEN test set, we employed a modified construction of performance profiles, intended for the case of multiple runs for each test problems; see [21] for details.



(a) Performance profile

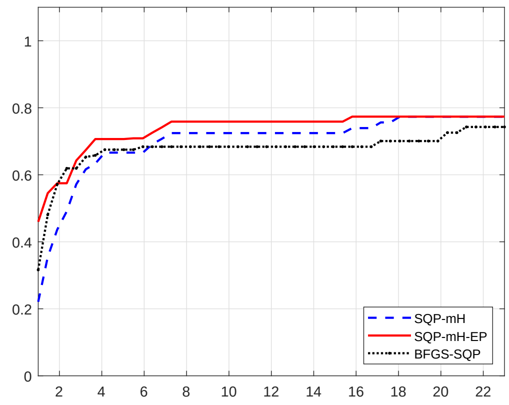


(b) Subproblems per iteration

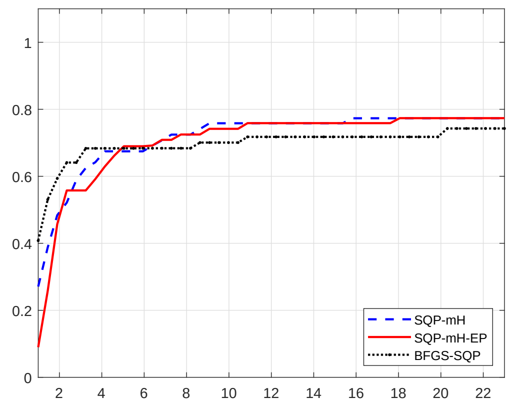
Fig. 4.1. Subproblems solved

The performance profile in Figure 4.1a demonstrates that SQP-mH-EP is as robust as SQP-mH, somehow more robust than SQP-BFGS, and by far more efficient than both by the number of quadratic subproblems solved per a successful run. The effect of using extrapolation is achieved because the true Hessian and the unit stepsize are typically asymptotically accepted, and the set of active constraints of subproblems asymptotically stabilizes.

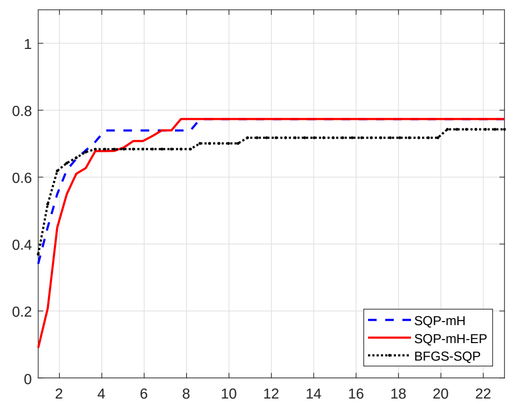
In addition, the plots in Figure 4.1b show which portion of problems required solving no more than a given number of quadratic subproblems per iteration on the average. It can be seen that the behavior of SQP-mH and SQP-mH-EP in this respect is quite similar. In particular, for 40% of test problems, only one quadratic subproblem per iteration had to be solved on the average, while solving two or more subproblems was needed for less than 10% of problems.



(a) Evaluations of f



(b) Evaluations of h and g



(c) Evaluations of derivatives

Fig. 4.2. Performance profiles by evaluations

The performance profiles in terms of evaluations of f, of the constraint mappings, and of their derivatives are presented in Figures 4.2a, 4.2b and 4.2c, respectively. According to

Figure 4.2a, the relative efficiency of SQP-mH-EP by evaluations of f is lower than that by the count of subproblems solved, but the positive effect of using extrapolation is still evident.

The picture in Figure 4.2b by evaluations of h and g, and in Figure 4.2c by evaluations of the derivatives is quite different though from that in Figures 4.1a and 4.2a: SQP-mH-EP demonstrated the best results by these measures of efficiency for about 10% of problems only. This behavior is explained by the fact that even if the true Hessian and the full step are typically asymptotically accepted, and the set of active constraints stabilizes, additional evaluations of h, g, and derivatives are required at extrapolated points for verifying the stopping test (4.28).

5. CONCLUDING REMARKS

We have discussed the use of extrapolation techniques for acceleration of convergence of the SQP method equipped with linesearch, when convergence is to a critical Lagrange multiplier of an optimization problem involving inequality constraints. A specific algorithmic implementation of these constructions has been developed and numerically tested.

It would be interesting to develop a reasonably complete supporting theory, i.e., conditions ensuring asymptotic acceptance of the true Hessian and the unit stepsize, as it is known to be possible for equality-constrained problems, and hopefully can be achieved by some kind of asymptotic reduction to the equality-constrained case. Observe, however, that the key difficulty for such development apparently consists of establishing the relation between the values and directional derivatives of the penalty functions for (1.1) and for (2.20) in the SQP directions.

ACKNOWLEDGEMENTS

This research was supported in part by the Russian Foundation for Basic Research Grants 19-51-12003 NNIO_a and 20-01-00106_a, and by the Volkswagen Foundation.

REFERENCES

1. Bonnans, J.F., Gilbert, J.Ch., Lemaréchal, C. & Sagastizábal, C. (2006) *Numerical optimization. Theoretical and practical aspects. 2nd edition*. Berlin: Springer-Verlag.
2. Dolan, E.D. & More, J.J. (2002) Benchmarking optimization software with performance profiles. Math. Program. 91, 201–213.
3. Ferris, M.C. & Munson, T.S.. PATH. [Online]. Available http://pages.cs.wisc.edu/~ferris/path.html.
4. Ferris, M.C. & Munson, T.S. (1999) Interfaces to PATH 3.0: design, implementation and usage. Comput. Optim. Appl. 12, 207–227.
5. Fischer, A., Izmailov, A.F. & Scheck, W. (2020) Adjusting dual iterates in the presence of critical Lagrange multipliers. SIAM J. Optim. 30, 1555–1581.
6. Fischer, A., Izmailov, A.F. & M.V. Solodov. (2021) Unit stepsize for the Newton method close to critical solutions. Math. Program. 187, 697–721.
7. Griewank, A. (1980) Analysis and modification of Mewton's method at singularities. PhD Thesis. Australian National University, Canberra.
8. Griewank, A. (1980) Starlike domains of convergence for Newton's method at singularities. Numer. Math. 35, 95–111.
9. A. Griewank. (1985) On solving nonlinear equations with simple singularities or nearly singular solutions. SIAM Rev. 27, 537–563.

10. Izmailov, A.F. (2021) Accelerating convergence of a globalized sequential quadratic programming method to critical Lagrange multipliers. Comput. Optim. Appl. 80, 943–978.
11. Izmailov, A.F., Kurennoy, A.S. & Solodov, M.V. (2018) Critical solutions of nonlinear equations: Local attraction for Newton-type methods. Math. Program. 167, 355–379.
12. Izmailov, A.F., Kurennoy, A.S. & Solodov, M.V. (2018) Critical solutions of nonlinear equations: Stability issues. Math. Program. 168, 475–507.
13. Izmailov, A.F. & Solodov, M.V. (2009) DEGEN. [Online]. Available http://w3.impa.br/~optim/DEGEN_collection.zip.
14. Izmailov, A.F. & Solodov, M.V. (2009) On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. Math. Program. 117, 271–304.
15. Izmailov, A.F. & Solodov, M.V. (2009) Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. Comput. Optim. Appl. 42, 231–264.
16. Izmailov, A.F. & Solodov, M.V. (2011) On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. Math. Program. 126, 231–257.
17. Izmailov, A.F. & Solodov, M.V. (2012) Stabilized SQP revisited. Math. Program. 133, 93–120.
18. Izmailov, A.F. & Solodov, M.V. (2014) *Newton-Type Methods for Optimization and Variational Problems.* Cham, Switzerland: Springer Series in Operations Research and Financial Engineering, Springer International Publishing.
19. Izmailov, A.F. & Solodov, M.V. (2015) Critical Lagrange multipliers: what we currently know about them, how they spoil our lives, and what we can do about it. TOP 23, 1–26.
20. Izmailov, A.F. & Solodov, M.V. (2016) Some new facts about sequential quadratic programming methods employing second derivatives. Optim. Meth. Software 31, 1111–1131.
21. Izmailov, A.F., Solodov, M.V. & Uskov, E.I. (2015) Combining stabilized SQP with the augmented Lagrangian algorithm Comput. Optim. Appl. 62, 405–429.
22. Li, D.-H. & Qi, L. (2000) Stabilized SQP method via linear equations. Applied Mathematics Technical Reptort AMR00/5. University of New South Wales, Sydney.
23. Nocedal, J. & Wright, S.J. (2006) *Numerical Optimization. 2nd edition.* New York, Berlin, Heidelberg: Springer-Verlag.
24. Wright, S.J. (1998) Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. 11, 253–275.

