SOME COMPOSITE-STEP CONSTRAINED OPTIMIZATION METHODS INTERPRETED VIA THE PERTURBED SEQUENTIAL QUADRATIC PROGRAMMING FRAMEWORK

A. F. Izmailov[†], A. S. Kurennoy[‡], and M. V. Solodov[§]

November 11, 2013 (Revised March 3, 2014)

ABSTRACT

We consider the inexact restoration and the composite-step sequential quadratic programming (SQP) methods, and relate them to the so-called perturbed SQP framework. In particular, iterations of the methods in question are interpreted as certain structured perturbations of the basic SQP iterations. This gives a different insight into local behaviour of those algorithms, as well as improved or different local convergence and rate of convergence results.

[†] Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: `izmaf@ccas.ru`
[‡] Moscow State University, MSU, Uchebniy Korpus 2, VMK Faculty, OR Department, Leninskiye Gory, 119991 Moscow, Russia.
Email: `akurennoy@cs.msu.ru`
[§] IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil.
Email: `solodov@impa.br`

1   Introduction

Consider the optimization problem with equality and nonnegativity constraints

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

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the constraints mapping $h : \mathbb{R}^n \to \mathbb{R}^l$ are twice differentiable. As is well known, problems with general inequality constraints can be cast in the form of (1.1) introducing auxiliary slack variables.

Let $L : \mathbb{R}^n \times \mathbb{R}^l \to \mathbb{R}$ be the (partial) Lagrangian of the problem (1.1), including only the equality constraints:

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

Then stationary points and associated Lagrange multipliers of the problem (1.1) are characterized by the Karush–Kuhn–Tucker (KKT) optimality system

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

with respect to $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}^l$ and $\mu \in \mathbb{R}^n$. By $\mathcal{M}(\bar{x})$ we denote the set of Lagrange multipliers associated with a stationary point $\bar{x}$ of the problem (1.1), that is, the pairs (λ, μ) satisfying (1.2) for $x = \bar{x}$.

Recall that the basic sequential quadratic programming (SQP) method for problem (1.1) is the following iterative procedure [1, 11]; see also [20, Chapter 4]. Given a current primal-dual iterate $(x^k, \lambda^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$, the next iterate $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$ is computed as a stationary point and associated Lagrange multipliers of the quadratic programming (QP) subproblem

$$
\begin{aligned}
&\text{minimize} \quad \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x - x^k), x - x^k \right\rangle \\
&\text{subject to} \quad h(x^k) + h'(x^k)(x - x^k) = 0, \ x \geq 0.
\end{aligned}
\tag{1.3}
$$

By the perturbed SQP (pSQP) framework we mean a set of relations that consists of certain structured perturbations of the KKT conditions of the QP subproblem (1.3); the precise description will be given in Section 2. This framework has been developed in [17, 16, 7] and it proved useful for analyzing, in a unified manner, a number of different Newtonian and Newton-related algorithms for constrained optimization (truncated and augmented Lagrangian modifications of SQP itself, sequential quadratically constrained quadratic programming, and linearly constrained Lagrangian methods, to mention some of the applications); see [20, Chapter 4]. In this paper we continue this line of reasoning and show that in addition to the above, local convergence properties of the inexact restoration methods [22, 21, 23, 3, 10, 8, 4] and of composite-step SQP methods [26, 24], [6, Section 15.4], can also be derived from the pSQP theory.

The paper is organized as follows. In Section 2 we formally state the pSQP framework and summarize its convergence properties. Section 3 considers an "exact restoration" scheme,

which is not a practical algorithm but rather serves as a natural first step to the analysis of inexact restoration methods, presented in Section 4. Inexact restoration methods have been receiving much attention in recent years; see [22, 21, 23, 3, 10, 8, 4]. Our considerations are related to the local framework of [3]. We establish local superlinear convergence of inexact restoration by embedding it within the pSQP framework of Section 2. The main difference of our results when compared to [3] is the following. Our analysis requires the strict Mangasarian–Fromovitz constraint qualification and the second-order sufficient optimality condition, and as part of the analysis we establish solvability of the iteration subproblems (i.e., that relevant solutions exist and thus the method is well-defined). By contrast, the analysis in [3] does not use these assumptions (in fact, does not use any regularity conditions at all), but solvability of subproblems is not proven but assumed. Thus our results and those in [3] are different in nature, and can be considered complementary to each other. That said, in Section 5 we show that without assuming any constraint qualifications, the existence of suitable subproblem solutions in the inexact restoration framework is at least questionable.

Section 6 suggests a similar treatment via pSQP framework of composite-step SQP methods [6, Section 15.4] (recall that the composite-step approach serves as the basis for the trust-region globalization strategies of SQP). In comparison with the rate of convergence results in [6, Theorems 15.4.15, 15.4.24], we obtain local superlinear convergence under a weaker constraint qualification, namely, the strict Mangasarian–Fromovitz constraint qualification instead of the linear independence one. Moreover, in [6, Theorems 15.4.15, 15.4.24] it is assumed that the pure (exact) SQP step is taken whenever it is accepted by the trust-region rule, while our analysis demonstrates that superlinear convergence can be preserved even if the SQP step is computed only approximately, assuming that the inexactness is appropriately controlled.

We conclude this section with some notation and definitions that will be used in our analysis. Throughout the paper, for a vector y of any dimension and an index set I, the notation y_I stands for the subvector of y with components y_i, $i \in I$.

For a given stationary point $\bar{x} \in \mathbb{R}^n$ of the problem (1.1) and an associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$, set

$$A = A(\bar{x}) = \{i = 1, \ldots, m \mid \bar{x}_i = 0\}, \quad N = N(\bar{x}) = \{1, \ldots, n\} \setminus A(\bar{x}),$$

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

As defined in [19], $(\bar{\lambda}, \bar{\mu}) \in \mathcal{M}(\bar{x})$ is referred to as a noncritical Lagrange multiplier if there exists no triple $(\xi, \eta, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$, with $\xi \neq 0$, satisfying the system

$$
\begin{aligned}
\frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi + (h'(\bar{x}))^{\mathrm{T}}\eta - \zeta = 0, \quad h'(\bar{x})\xi = 0, \quad \xi_{A_+} = 0, \\
\zeta_{A_0} \geq 0, \quad \xi_{A_0} \geq 0, \quad \zeta_i \xi_i = 0, \ i \in A_0, \\
\zeta_N = 0.
\end{aligned}
\tag{1.4}
$$

We refer the reader to [14, 15, 18, 19, 12, 13] and [20, Chapter 7] for the role this notion plays in convergence properties of algorithms, stability, error bounds, and other issues. Here, we only mention that noncriticality is equivalent to the local Lipschitzian error bound for the

KKT system (1.2) of the problem (1.1) in terms of its natural residual, and to the upper-Lipschitzian behaviour of solutions of the KKT system (1.2) under canonical (right-hand side) perturbations. Also, it can be easily seen that the multiplier is necessarily noncritical if the following second-order sufficient optimality condition (SOSC) holds:

$$\left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\lambda})\xi, \xi \right\rangle > 0 \quad \forall \xi \in C(\bar{x}) \setminus \{0\}, \tag{1.5}$$

where

$$C(\bar{x}) = \{\xi \in \mathbb{R}^n \mid h'(\bar{x})\xi = 0, \ \xi_A \geq 0, \ \langle f'(\bar{x}), \xi \rangle \leq 0\}$$

is the critical cone of the problem (1.1) at $\bar{x}$. It is also clear that noncriticality of a multiplier is a weaker assumption than the SOSC (1.5), i.e., there may well exist (and often do exist) noncritical multipliers that do not satisfy (1.5).

Recall that the strict Mangasarian–Fromovitz constraint qualification (SMFCQ) is said to hold at a stationary point $\bar{x}$ of the problem (1.1) if the Lagrange multiplier $(\bar{\lambda}, \bar{\mu})$ associated to $\bar{x}$ exists and is unique. As we shall not need the algebraic characterization of this condition, we do not state it here (the reader may consult [25], for example). Recall that SMFCQ is implied by the linear independence constraint qualification (LICQ). For the problem (1.1), LICQ can be seen to be equivalent to saying that the vectors $(h'_j(\bar{x}))_N$, $j = 1, \ldots, l$, are linearly independent.

2 The perturbed SQP framework

We can state the iteration subproblem of the basic SQP method (1.3) for the problem (1.1) as

$$\begin{aligned} \text{minimize} \quad & f(x) + \langle f'(x), \xi \rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi, \xi \right\rangle \\ \text{subject to} \quad & h(x) + h'(x)\xi = 0, \ x + \xi \geq 0, \end{aligned} \tag{2.1}$$

so that for a current iterate $(x^k, \lambda^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$, the next iterate is $(x^{k+1}, \lambda^{k+1}, \mu^{k+1})$, where $x^{k+1} = x^k + \xi^k$, ξ^k is a stationary point of the problem (2.1) with $(x, \lambda, \mu) = (x^k, \lambda^k, \mu^k)$, and $(\lambda^{k+1}, \mu^{k+1})$ is an associated Lagrange multiplier.

Then the iteration subproblem of the pSQP framework [16], when specialized for the problem setting of (1.1), has the form of perturbed KKT conditions of the problem (2.1). Specifically,

$$\begin{aligned} \left\| f'(x) + \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi + (h'(x))^{\mathrm{T}}(\lambda + \eta) - \mu - \zeta + \omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)) \right\| &\leq \chi_1(x, \lambda, \mu), \\ \|h(x) + h'(x)\xi + \omega_2((x, \lambda, \mu), \xi)\| &\leq \chi_2(x, \lambda, \mu), \\ \mu + \zeta \geq 0, \quad x + \xi \geq 0, \quad \langle \mu + \zeta, x + \xi \rangle = 0. & \end{aligned} \tag{2.2}$$

Here $\omega_1 : (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n) \to \mathbb{R}^n$ and $\omega_2 : (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n) \times \mathbb{R}^n \to \mathbb{R}^l$ are functions characterizing structural perturbations of the KKT system of the basic SQP subproblem (2.1), while $\chi_1 : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n \to \mathbb{R}_+$ and $\chi_2 : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n \to \mathbb{R}_+$ are some forcing functions controlling additional inexactness allowed when solving the structurally

perturbed SQP subproblems. In particular, if $\omega_1 \equiv 0$ and $\omega_2 \equiv 0$ then (2.2) represents solving the usual SQP subproblems (2.1), perhaps approximately (exactly if further $\chi_1 \equiv 0$ and $\chi_2 \equiv 0$). Otherwise, it is the functions ω_1 and ω_2 that define each specific algorithm within the pSQP framework (they represent "the difference" between the pure SQP iteration and that of the given algorithm). We note that in general, in the last line of (2.2) the inequality constraints in primal variables can be perturbed too (in (2.2) they are not), see [16, 17]; but complementarity relations have to be maintained exactly. We do not need the extra generality of perturbing inequality constraints for the applications in this paper. Note also that for simple bounds as in the current setting, exact complementarity can be maintained employing active-set methods for subproblems, or by using simple "purification procedures", like the one proposed in [16]. For a current iterate $(x^k, \lambda^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$, the next pSQP iterate is $(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) = (x^k + \xi^k, \lambda^k + \eta^k, \mu^k + \zeta^k)$, where (ξ^k, η^k, ζ^k) satisfies (2.2) with $(x, \lambda, \mu) = (x^k, \lambda^k, \mu^k)$.

For *a priori* local convergence analysis of pSQP (i.e., where the existence of the iterative sequence and its convergence are not given but have to be proven), one needs to assume that ω_2 is smooth with respect to the last variable, and for all $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ and $(\xi, \eta, \zeta) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ it holds that

$$\omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)) = \frac{\partial \Psi}{\partial \xi}((x, \lambda, \mu), (\xi, \eta, \zeta)), \tag{2.3}$$

where $\Psi : (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n) \to \mathbb{R}$,

$$\Psi((x, \lambda, \mu), (\xi, \eta, \zeta)) = \psi((x, \lambda, \mu), \xi) + \langle \lambda + \eta, \omega_2((x, \lambda, \mu), \xi) \rangle, \tag{2.4}$$

with some function $\psi : (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n) \times \mathbb{R}^n \to \mathbb{R}$ which is smooth with respect to the last variable. Observe that in this case the system (2.2) with $\chi_1(\cdot) \equiv 0$ and $\chi_2(\cdot) \equiv 0$ coincides with the KKT system of the problem

$$\begin{aligned}
\text{minimize} \quad & f(x) + \langle f'(x), \xi \rangle + \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi, \xi \right\rangle + \psi((x, \lambda, \mu), \xi) \\
\text{subject to} \quad & h(x) + h'(x)\xi + \omega_2((x, \lambda, \mu), \xi) = 0, \; x + \xi \geq 0,
\end{aligned}$$

which is the perturbed version of the SQP subproblem (2.1).

Let the mapping $\Phi : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ be defined by

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

where $u = (x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$, and let $N(u)$ stand for the normal cone to $\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n_+$ at u. Furthermore, define the multifunction Ω from $(\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n) \times (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n)$ to the subsets of $\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ by

$$\Omega(u, v) = \omega(u, v) + \Theta(u),$$

where $v = (\xi, \eta, \zeta)$,

$$\omega(u, v) = (\omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)), \omega_2((x, \lambda, \mu), \xi)),$$

4

and

$$\Theta(u) = \Theta_1(u) \times \Theta_2(u) \times \{0\}$$

with

$$\Theta_1(u) = \{\theta_1 \in \mathbb{R}^n \mid \|\theta_1\| \le \chi_1(x, \lambda, \mu)\}, \quad \Theta_2(u) = \{\theta_2 \in \mathbb{R}^l \mid \|\theta_2\| \le \chi_2(x, \lambda, \mu)\}.$$

With these mappings at hand, the iteration subproblem (2.2) of pSQP can be seen as the generalized equation (GE)

$$\Phi(u^k) + \Phi'(u^k)(u - u^k) + \Omega(u^k, u - u^k) + N(u) \ni 0,$$

which is the iteration subproblem of the inexact Josephy-Newton method for GEs developed in [17]. Then, under the appropriate assumptions on the perturbation terms ψ and ω_2, the following local superlinear convergence result for pSQP can be derived from [17, Theorem 2.1]; see [20, Chapter 4] for details. We note that Theorem 2.1 below is similar to the statement in [16], but with somewhat weaker smoothness requirements on perturbation terms. Its proof is also similar to [16]; see [20].

Theorem 2.1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be twice differentiable in a neighbourhood of $\bar{x} \in \mathbb{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let $\bar{x}$ be a solution of the problem (1.1), satisfying the SMFCQ and the SOSC (1.5) for the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$. Furthermore, let a function $\psi : (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n) \times \mathbb{R}^n \to \mathbb{R}$ and a mapping $\omega_2 : (\mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n) \times \mathbb{R}^n \to \mathbb{R}^l$ possess the following properties:*

(i) ψ *is continuous at $((\bar{x}, \bar{\lambda}, \bar{\mu}), \xi)$ and $\omega_2(\cdot, \xi)$ is continuous at $(\bar{x}, \bar{\lambda}, \bar{\mu})$, for every $\xi \in \mathbb{R}^n$ close enough to 0.*

(ii) ψ *and ω_2 are differentiable with respect to ξ in a neighbourhood of $((\bar{x}, \bar{\lambda}, \bar{\mu}), 0)$ and twice differentiable with respect to ξ at this point.*

(iii) $\frac{\partial \psi}{\partial \xi}$ *and $\frac{\partial \omega_2}{\partial \xi}$ are continuous at $((\bar{x}, \bar{\lambda}, \bar{\mu}), 0)$, and there exists a neighbourhood of 0 in $\mathbb{R}^n$ such that $\frac{\partial \omega_2}{\partial \xi}((x, \lambda, \mu), \cdot)$ is continuous on this neighbourhood for all $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ close enough to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

(iv) *The equalities*

$$\omega_2((\bar{x}, \bar{\lambda}, \bar{\mu}), 0) = 0, \quad \frac{\partial \psi}{\partial \xi}((\bar{x}, \bar{\lambda}, \bar{\mu}), 0) = 0, \quad \frac{\partial \omega_2}{\partial \xi}((\bar{x}, \bar{\lambda}, \bar{\mu}), 0) = 0$$

hold, and for the function Ψ defined by (2.4) it holds that

$$\left\langle \frac{\partial^2 \Psi}{\partial \xi^2}((\bar{x}, \bar{\lambda}, \bar{\mu}), (0, 0, 0))\xi, \xi \right\rangle \ge 0 \quad \forall \xi \in C(\bar{x}).$$

Assume further that the following estimates hold as $(x, \lambda, \mu) \in \mathrm{I\!R}^n \times \mathrm{I\!R}^l \times \mathrm{I\!R}^n$ *tends to* $(\bar{x}, \bar{\lambda}, \bar{\mu})$ *and as* $(\xi, \eta, \zeta) \in \mathrm{I\!R}^n \times \mathrm{I\!R}^l \times \mathrm{I\!R}^n$ *tends to zero:*

$$\omega_2((x, \lambda, \mu), \xi) = o(\|\xi\| + \|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu}\|), \tag{2.5}$$

$$\frac{\partial \Psi}{\partial \xi}((x, \lambda, \mu), (\xi, \eta, \zeta)) = o(\|(\xi, \eta, \zeta)\| + \|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu}\|), \tag{2.6}$$

$$\chi_j(x, \lambda, \mu) = o(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu}\|), \quad j = 1, 2. \tag{2.7}$$

Then there exist $\varepsilon > 0$ *and* $\delta > 0$ *such that for any starting point* $(x^0, \lambda^0, \mu^0) \in \mathrm{I\!R}^n \times \mathrm{I\!R}^l \times \mathrm{I\!R}^n$ *satisfying*

$$\|(x^0 - \bar{x}, \lambda^0 - \bar{\lambda}, \mu^0 - \bar{\mu})\| \le \varepsilon, \tag{2.8}$$

if a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathrm{I\!R}^n \times \mathrm{I\!R}^l \times \mathrm{I\!R}^n$ *satisfies, for each* $k = 0, 1, \ldots$, *the system* (2.2) *with* ω_1 *defined in* (2.3) *and* Ψ *defined in* (2.4), *and also satisfies*

$$\|(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)\| \le \delta, \tag{2.9}$$

then this sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, *and the rate of convergence is superlinear. Moreover, under the stated assumptions,* $\varepsilon > 0$ *and* $\delta > 0$ *can be chosen in such a way that for any* $(x^0, \lambda^0, \mu^0) \in \mathrm{I\!R}^n \times \mathrm{I\!R}^l \times \mathrm{I\!R}^n$ *satisfying* (2.8) *there exists at least one sequence satisfying, for each* $k = 0, 1, \ldots$, *the relations* (2.2) *and* (2.9).

In addition, the rate of convergence is quadratic if the second derivatives of f *and* h *are locally Lipschitz-continuous with respect to* $\bar{x}$, *and if* (2.5)–(2.7) *can be replaced by the estimates*

$$\omega_2((x, \lambda, \mu), \xi) = O(\|\xi\|^2 + \|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu}\|^2),$$

$$\frac{\partial \Psi}{\partial \xi}((x, \lambda, \mu), (\xi, \eta, \zeta)) = O(\|(\xi, \eta, \zeta)\|^2 + \|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu}\|^2),$$

$$\chi_j(x, \lambda, \mu) = O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu}\|^2), \quad j = 1, 2.$$

Theorem 2.1 is, of course, a rather general technical statement. For its applications to specific algorithms, such as truncated and augmented Lagrangian modifications of SQP, sequential quadratically constrained quadratic programming, and linearly constrained Lagrangian methods, see [17, 16]. In the next sections, we shall proceed to inexact restoration and composite-step SQP, the subjects of the present paper. But before that, we state an *a posteriori* primal rate of convergence result for pSQP, established in [7]. Here, convergence of the iterative (primal-dual) sequence is given (or already established via Theorem 2.1), and at issue is the rate of convergence of the primal part of the sequence. Recall that in general, superlinear convergence of primal-dual sequence does not imply any rate for the primal sequence separately [2, Exercise 14.8].

We are now *given* iterative sequences linked by the perturbed KKT conditions of SQP subproblems (1.3) as follows:

$$f'(x^k) + \frac{\partial^2 L}{\partial x^2}(x^k, \lambda^k)(x^{k+1} - x^k) + (h'(x^k))^{\mathrm{T}}\lambda^{k+1} - \mu^{k+1} + \omega_1^k = 0,$$
$$h(x^k) + h'(x^k)(x^{k+1} - x^k) + \omega_2^k = 0, \tag{2.10}$$
$$\mu^{k+1} \ge 0, \quad x^{k+1} \ge 0, \quad \langle \mu^{k+1}, x^{k+1} \rangle = 0,$$

where $\omega_1^k \in \mathbb{R}^n$ and $\omega_2^k \in \mathbb{R}^l$ are the perturbation terms. Again, in general the part invloving inequality constraints (here represented by $x^{k+1} \geq 0$) can be perturbed too, but we do not need this generality for applications to algorithms in this paper. The following statement is an adaptation of the results in [7] to the problem setting of (1.1), i.e., to the set of relations (2.10).

Theorem 2.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^l$ be twice differentiable in a neighbourhood of $\bar{x} \in \mathbb{R}^n$, with their second derivatives being continuous at $\bar{x}$. Let a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ satisfy, for each k, the system (2.10) with some $\omega_1^k \in \mathbb{R}^n$, $\omega_2^k \in \mathbb{R}^l$; let this sequence be convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$ where $\bar{x}$ is a stationary point of the problem (1.1), with an associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$ being noncritical.*
If

$$\pi_{C_+(\bar{x}, \bar{\mu})}(\omega_1^k) = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|),$$

where

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

and

$$\omega_2^k = o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|),$$

as $k \to \infty$, then the rate of convergence of $\{x^k\}$ is superlinear.

3 Exact restoration

It is instructive to consider first an *"exact restoration"* method; it is not a practical algorithm but rather a motivation for inexact restoration methods and a natural first step to their analysis. In particular, it is the exact restoration that defines the associated class of methods within the pSQP framework (i.e., it defines structural pertubations relative to SQP), while inexact restoration can be regarded simply as solving the subproblems of the method approximately in some sense.

Algorithm 3.1 Choose $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ and set $k = 0$.

1. If (x^k, λ^k, μ^k) satisfies the KKT system (1.2), stop.

2. (Feasibility phase.) Compute π^k as a projection of x^k onto the feasible set of the problem (1.1), i.e., a global solution of the subproblem

$$\begin{array}{ll} \text{minimize}_\pi & \|\pi - x\| \\ \text{subject to} & h(\pi) = 0, \ \pi \geq 0 \end{array} \tag{3.1}$$

for $x = x^k$.

3. If $(\pi^k, \lambda^k, \mu^k)$ satisfies the KKT system (1.2), stop.

4. (Optimality phase.) Compute x^{k+1} and (η^k, μ^{k+1}) as a stationary point and an associated Lagrange multiplier of the subproblem

$$\begin{aligned}\text{minimize}_x \quad & L(x, \lambda)\\ \text{subject to} \quad & h'(\pi)(x - \pi) = 0, \; x \geq 0\end{aligned} \tag{3.2}$$

for $\pi = \pi^k$, $\lambda = \lambda^k$.

5. Set $\lambda^{k+1} = \lambda^k + \eta^k$.

6. Increase k by 1 and go to step 1.

For a given $x \in \mathbb{R}^n$, let $\bar{\pi}(x)$ be a projection of x onto the feasible set of the problem (1.1), computed as at the feasibility phase of Algorithm 3.1 solving the problem (3.1). In order to formally apply Theorem 2.1, we need to assume that $\bar{\pi}(\cdot)$ is a fixed single-valued function. As the feasible set is not convex, the projection onto it need not be unique, in general. However, an algorithm used to solve (3.1) follows its internal patterns and computes one specific projection. It is further reasonable to assume that, if at some future iteration a projection of the same point needs to be computed again (however unlikely this might be from a practical viewpoint), the algorithm would return the same result. With this in mind, considering that $\bar{\pi}(\cdot)$ is a single-valued function is justified for all practical purposes. In Theorem 3.1 below this assumption is stated more formally.

Then the next primal iterate x^{k+1} can be seen as $x^k + \xi^k$, where ξ^k is a stationary point of the following counterpart of the problem (3.2):

$$\begin{aligned}\text{minimize}_\xi \quad & L(x + \xi, \lambda)\\ \text{subject to} \quad & h'(\bar{\pi}(x))(x + \xi - \bar{\pi}(x)) = 0, \; x + \xi \geq 0\end{aligned} \tag{3.3}$$

with $x = x^k$, $\lambda = \lambda^k$. For the next dual iterate $(\lambda^{k+1}, \mu^{k+1})$ it holds that $\lambda^{k+1} = \lambda^k + \eta^k$, where (η^k, μ^{k+1}) is a Lagrange multiplier associated with ξ^k.

To place the exact restoration scheme above within the pSQP framework so that the needed assumptions on the perturbation terms be satisfied, we shall replace (3.3) by the (related) subproblem

$$\begin{aligned}\text{minimize}_\xi \quad & L(x + \xi, \lambda) - \langle \lambda, h'(\bar{\pi}(x))\xi \rangle\\ \text{subject to} \quad & h'(\bar{\pi}(x))(x + \xi - \bar{\pi}(x)) = 0, \; x + \xi \geq 0,\end{aligned} \tag{3.4}$$

still with $x = x^k$, $\lambda = \lambda^k$. It can be directly verified that stationary points ξ^{k+1} of the problems (3.3) and (3.4) coincide, and the associated multipliers are of the form (η^k, μ^{k+1}) and $(\lambda^{k+1}, \mu^{k+1})$, with $\lambda^{k+1} = \lambda^k + \eta^k$, respectively. Thus, for the purposes of convergence analysis, we can deal with the modified subproblems (3.4). It turns out that this allows to apply Theorem 2.1 if for $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ and $\xi \in \mathbb{R}^n$ we take

$$\begin{aligned}\psi((x, \lambda, \mu), \xi) \;=\; & \psi((x, \lambda), \xi)\\[4pt] =\; & L(x + \xi, \lambda) - \langle \lambda, h'(\bar{\pi}(x))\xi \rangle - f(x) - \langle f'(x), \xi \rangle - \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi, \xi \right\rangle\\[4pt] =\; & L(x + \xi, \lambda) - \left\langle \frac{\partial L}{\partial x}(x, \lambda), \xi \right\rangle - \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi, \xi \right\rangle\\[4pt] & - \langle \lambda, (h'(\bar{\pi}(x)) - h'(x))\xi \rangle - f(x),\end{aligned} \tag{3.5}$$

$$\omega_2((x, \lambda, \mu), \xi) = \omega_2(x, \xi) = h'(\bar{\pi}(x))(x + \xi - \bar{\pi}(x)) - h(x) - h'(x))\xi. \qquad (3.6)$$

Let $\bar{x} \in \mathrm{I\!R}^n$ be a stationary point of the problem (1.1), and $(\bar{\lambda}, \bar{\mu}) \in \mathrm{I\!R}^l \times \mathrm{I\!R}^n$ be an associated Lagrange multiplier. Evidently, $\|\bar{\pi}(x) - x\| \leq \|x - \bar{x}\|$ for any $x \in \mathrm{I\!R}^n$ and any solution $\bar{\pi}(x)$ of the problem (3.1). In particular, $\bar{\pi}(x) \to \bar{x} = \bar{\pi}(\bar{x})$ as $x \to \bar{x}$. Employing the mean-value theorem, one can directly verify that under the appropriate smoothness assumptions on f and h, the function ψ and the mapping ω_2 defined by (3.5), (3.6) possess all the properties required in Theorem 2.1 with $\chi_j(\cdot) \equiv 0$, $j = 1, 2$. We omit the technical details as they are completely routine, and only mention that for $\Psi((x, \lambda, \mu), (\xi, \eta, \zeta)) = \Psi((x, \lambda), (\xi, \eta))$ defined according to (2.4) it holds that

$$\frac{\partial^2 \Psi}{\partial \xi^2}((\bar{x}, \bar{\lambda}), (0, 0)) = 0,$$

and

$$
\begin{aligned}
\frac{\partial \Psi}{\partial \xi}((x, \lambda), (\xi, \eta)) &= \frac{\partial L}{\partial x}(x + \xi, \lambda) - \frac{\partial L}{\partial x}(x, \lambda) - \frac{\partial^2 L}{\partial x^2}(x, \lambda)\xi + (h'(\bar{\pi}(x)) - h'(x))^{\mathrm{T}}\eta \\
&= o(\|\xi\|) + O(\|\eta\|\|x - \bar{x}\|), \qquad (3.7)
\end{aligned}
$$

$$
\begin{aligned}
\omega_2(x, \xi) &= (h'(\bar{\pi}(x)) - h'(x))\xi + h(\bar{\pi}(x)) - h(x) - h'(\bar{\pi}(x))(\bar{\pi}(x) - x) \\
&= O(\|\xi\|\|x - \bar{x}\| + \|x - \bar{x}\|^2) \qquad (3.8)
\end{aligned}
$$

as $(x, \lambda) \to (\bar{x}, \bar{\lambda})$ and $(\xi, \eta) \to (0, 0)$. Moreover, under stronger smoothness assumptions, the first estimate can be sharpened as follows:

$$\frac{\partial \Psi}{\partial \xi}((x, \lambda), (\xi, \eta)) = O(\|\xi\|^2 + (\|\xi\| + \|\eta\|)\|x - \bar{x}\|). \qquad (3.9)$$

Applying Theorem 2.1, we now obtain conditions for local superlinear convergence of Algorithm 3.1.

Theorem 3.1 *Let $f : \mathrm{I\!R}^n \to \mathrm{I\!R}$ and $h : \mathrm{I\!R}^n \to \mathrm{I\!R}^l$ be twice differentiable in a neighborhood of $\bar{x} \in \mathrm{I\!R}^n$, with their second derivatives continuous at $\bar{x}$. Let $\bar{x}$ be a stationary point of the problem (1.1), satisfying SMFCQ and the SOSC (1.5) for the associated Lagrange multiplier $(\bar{\lambda}, \bar{\mu}) \in \mathrm{I\!R}^l \times \mathrm{I\!R}^n$. Assume that if $x^k = x^j$ for any two iteration indices k and j, then step 2 of Algorithm 3.1 computes $\pi^k = \pi^j$.*

Then there exist $\varepsilon > 0$ and $\delta > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathrm{I\!R}^n \times \mathrm{I\!R}^l \times \mathrm{I\!R}^n$ satisfying

$$\|(x^0 - \bar{x}, \lambda^0 - \bar{\lambda}, \mu^0 - \bar{\mu})\| \leq \varepsilon, \qquad (3.10)$$

if a sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathrm{I\!R}^n \times \mathrm{I\!R}^l \times \mathrm{I\!R}^n$ generated by Algorithm 3.1 satisfies

$$\|(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)\| \leq \delta \qquad (3.11)$$

for each $k = 0, 1, \ldots$, then this sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, $\varepsilon > 0$ and $\delta > 0$ can be chosen in such a way that for any starting point

$(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ *satisfying* (3.10), *there exists at least one sequence conforming, for each* $k = 0, 1, \ldots$, *to Algorithm 3.1 and the condition* (3.11).

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

Also, an *a posteriori* result regarding primal superlinear convergence of Algorithm 3.1 now follows readily from Theorem 2.2. Note that the needed assumption is noncriticality of the multiplier, weaker than the SOSC (1.5).

Theorem 3.2 *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* (1.1), *and let $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^n$ be an associated noncritical Lagrange multiplier. Let an iterative sequence $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ generated by Algorithm 3.1 be convergent to $(\bar{x}, \bar{\lambda}, \bar{\mu})$.*

Then the rate of convergence of $\{x^k\}$ is superlinear.

Proof. According to the discussion above, for each k it holds that $(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) = (x^k + \xi^k, \lambda^k + \eta^k, \mu^k + \zeta^k)$, where (ξ^k, η^k, ζ^k) satisfies the system (2.2) with $\omega_1((x, \lambda, \mu), (\xi, \eta, \zeta)) = \omega_1((x, \lambda), (\xi, \eta))$ defined by (2.3) and the first equality in (3.7), with ω_2 defined by (3.6), and with $\chi_1(\cdot) \equiv 0$ and $\chi_2(\cdot) \equiv 0$. Therefore, the set of relations (2.10) holds for each k, where according to (3.7), (3.8),

$$
\begin{aligned}
\omega_1^k &= \omega_1((x^k, \lambda^k), (x^{k+1} - x^k, \lambda^{k+1} - \lambda^k)) \\
&= o(\|x^{k+1} - x^k\|) + O(\|\lambda^{k+1} - \lambda^k)\|\|x^k - \bar{x}\|) \\
&= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|),
\end{aligned}
$$

$$
\begin{aligned}
\omega_2^k &= \omega_2(x^k, x^{k+1} - x^k) \\
&= O(\|x^{k+1} - x^k\|\|x^k - \bar{x}\| + \|x^k - \bar{x}\|^2) \\
&= o(\|x^{k+1} - x^k\| + \|x^k - \bar{x}\|)
\end{aligned}
$$

as $k \to \infty$. Theorem 2.2 now implies the assertion. ■

4 Inexact restoration

Clearly, solving the subproblems (3.1) and (3.2) in Algorithm 3.1 exactly would be too costly, in most cases simply impossible. The main question of this section is what kind of inexactness can be allowed when solving these subproblems, so that the local convergence and rate of convergence properties of Algorithm 3.1 would remain valid. To that end, we now consider the following framework which we refer to as the *inexact restoration method*.

Algorithm 4.1 Choose functions $\varphi_0, \varphi_1, \varphi_2 : \mathbb{R}_+ \to \mathbb{R}_+$. Choose $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ and set $k = 0$.

1. If (x^k, λ^k, μ^k) satisfies the KKT system (1.2), stop.

2. (Feasibility phase.) Compute $\pi^k \in \mathbb{R}^n$ satisfying

$$\|h(\pi)\| \leq \varphi_0(\|h(x)\|), \quad \pi \geq 0. \tag{4.1}$$

   for $x = x^k$.

3. If $(\pi^k, \lambda^k, \mu^k)$ satisfies the KKT system (1.2), stop.

4. (Optimality phase.) Compute x^{k+1} and (η^k, ζ^k) satisfying

$$\left\| \frac{\partial L}{\partial x}(x, \lambda) + (h'(\pi))^{\mathrm{T}}\eta - \mu - \zeta \right\| \leq \varphi_1\left(\left\| \frac{\partial L}{\partial x}(\pi, \lambda) - \mu \right\| \right), \tag{4.2}$$

$$\|h'(\pi)(x - \pi)\| \leq \varphi_2\left(\left\| \frac{\partial L}{\partial x}(\pi, \lambda) - \mu \right\| \right), \tag{4.3}$$

$$\mu + \zeta \geq 0, \quad x \geq 0, \quad \langle \mu + \zeta, x \rangle = 0 \tag{4.4}$$

   for $\pi = \pi^k$, $\lambda = \lambda^k$, and $\mu = \mu^k$.

5. Set $\lambda^{k+1} = \lambda^k + \eta^k$, $\mu^{k+1} = \mu^k + \zeta^k$.

6. Increase k by 1 and go to step 1.

In the analysis below it will be further assumed that π^k computed at the feasibility phase of Algorithm 4.1 is within a controllable distance from x^k. Specifically, in addition to (4.1), we shall require that

$$\|\pi - x\| \leq K\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\| \tag{4.5}$$

for some $K > 0$ independent of (x, λ, μ). In practice, this can be achieved, e.g., by approximately solving the subproblem (3.1), or by other feasibility restoration strategies. In [3], instead of (4.5) the following stronger condition is employed

$$\|\pi - x\| \leq \tilde{K}\|h(x)\| \tag{4.6}$$

with some $\tilde{K} > 0$, which certainly implies (4.5) with $K = \ell\tilde{K}$ if h is Lipschitz-continuous with respect to $\bar{x}$ with constant $\ell > 0$. Moreover, as a practical implementation of (4.6), in [3] it is suggested to fix $\tilde{K} > 0$ as a parameter of the algorithm, and to compute π^k as an approximate solution of the subproblem

$$\begin{aligned} \text{minimize}_\pi \quad & \|h(\pi)\|^2 \\ \text{subject to} \quad & \|\pi - x\| \leq \tilde{K}\|h(x)\|, \ \pi \geq 0 \end{aligned}$$

for $x = x^k$. The approximate solution is supposed to satisfy the constraints of this subproblem and the condition (4.1). However, the difficulty with this approach is that for a given $\tilde{K}$, the approximate solution of this kind does not necessarily exist, in which case the algorithm in [3] simply declares failure at the feasibility phase. To that end, below we do not employ (4.6)

11

as an actual constraint of the feasibility subproblem of the algorithm, but rather use (4.5) as an ingredient of the analysis.

In [3] it is also assumed that $(x^{k+1}, \eta^k, \zeta^k)$ computed at the optimality phase of Algorithm 4.1 lies within a controllable distance from $(\pi^k, 0, 0)$: in addition to (4.2), (4.3), it is required that

$$\|(x - \pi, \eta, \zeta)\| \leq \tilde{K} \left\| \frac{\partial L}{\partial x}(\pi, \lambda) - \mu \right\|. \tag{4.7}$$

(In [3] Lagrange multipliers for nonnegativity constraints are not used, but otherwise (4.7) essentially corresponds to the constructions in [3].) In practice, this can be achieved by approximately solving the subproblem (3.2), perhaps with some additional constraints; see [3]. However, this additional control can be needed under the relaxed or removed constraint qualifications only. Under the assumptions of Theorem 3.1, it will be enough to assume that $(x^{k+1}, \eta^k, \zeta^k)$ is simply not too far from $(\pi^k, 0, 0)$.

Employing (4.1)–(4.4), by the same transformations as in Section 3, an iteration of Algorithm 4.1 can be seen as solving (2.2) with ω_1 defined by (2.3), where ψ is given by (3.5), with ω_2 defined by (3.6), and with

$$\chi_1(x, \lambda, \mu) = \varphi_1 \left(\left\| \frac{\partial L}{\partial x}(\pi(x, \lambda, \mu), \lambda) - \mu \right\| \right), \tag{4.8}$$

$$\chi_2(x, \lambda, \mu) = \varphi_2 \left(\left\| \frac{\partial L}{\partial x}(\pi(x, \lambda, \mu), \lambda) - \mu \right\| \right), \tag{4.9}$$

where $\pi(x, \lambda, \mu)$ is the point selected at the feasibility phase of the algorithm. Observe that for any $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$, the point $\bar{x}$ satisfies both (4.1) and (4.5) with any $K \geq 1$, and hence, can be selected as $\pi(x, \lambda, \mu)$ at the feasibility phase of the algorithm. Therefore, $\pi(x, \lambda, \mu)$ with the needed properties always exists.

Similarly to the case of exact restoration, we can reasonably assume that if step 2 of Algorithm 4.1 is applied at equal primal-dual points on different iterations then the same result is produced. In particular, $\pi(\cdot)$ is a fixed single-valued function.

By differentiability of h at $\bar{x}$, we conclude that

$$h(x) = h(x) - h(\bar{x}) = O(\|x - \bar{x}\|) \tag{4.10}$$

as $x \to \bar{x}$. Moreover, by twice differentiability of f and h at $\bar{x}$, taking into account (4.5), we obtain that

$$
\begin{aligned}
\left\| \frac{\partial L}{\partial x}(\pi(x, \lambda, \mu), \lambda) - \mu \right\| &\leq \|\mu - \bar{\mu}\| + \left\| \frac{\partial L}{\partial x}(\pi(x, \lambda, \mu), \lambda) - \frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}) \right\| \\
&= \|\mu - \bar{\mu}\| + O(\|(\pi(x, \lambda, \mu) - \bar{x}, \lambda - \bar{\lambda})\|) \\
&= \|\mu - \bar{\mu}\| + O(\|x - \bar{x}\|) + O(\|(\pi(x, \lambda, \mu) - x, \lambda - \bar{\lambda})\|) \\
&= O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|) \tag{4.11}
\end{aligned}
$$

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

Theorem 4.1 *Under the assumptions of Theorem 3.1, suppose that if $(x^k, \lambda^k, \mu^k) = (x^j, \lambda^j, \mu^j)$ for any two iteration indices k and j, then step 2 of Algorithm 4.1 computes $\pi^k = \pi^j$.*

Then for any functions φ_0, φ_1, $\varphi_2 : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\varphi_0(t) = o(t)$, $\varphi_1(t) = o(t)$ and $\varphi_2(t) = o(t)$ as $t \to 0$, and any $K \geq 1$, there exist $\varepsilon > 0$ and $\delta > 0$ such that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ satisfying (3.10), if an iterative sequence $\{(x^k, \pi^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ generated by Algorithm 4.1 satisfies, for each $k = 0, 1, \ldots$, the relations

$$\|\pi^k - x^k\| \leq K \|(x^k - \bar{x}, \lambda^k - \bar{\lambda}, \mu^k - \bar{\mu})\| \tag{4.12}$$

and (3.11), then this sequence converges to $(\bar{x}, \bar{\lambda}, \bar{\mu})$, and the rate of convergence is superlinear. Moreover, $\varepsilon > 0$ and $\delta > 0$ can be chosen in such a way that for any starting point $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ satisfying (3.10), there exists at least one sequence conforming, for each $k = 0, 1, \ldots$, to Algorithm 4.1 and satisfying (3.11) and (4.12).

In addition, the rate of convergence is quadratic if the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$, and if $\varphi_0(t) = O(t^2)$, $\varphi_1(t) = O(t^2)$ and $\varphi_2(t) = O(t^2)$ as $t \to 0$.

Proof. Combining (4.8), (4.9) with (4.10), (4.11), and taking into account the assumptions regarding φ_0, φ_1 and φ_2, we obtain that

$$\chi_j(x, \lambda, \mu) = o(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|), \quad j = 1, 2,$$

as $(x, \lambda, \mu) \to (\bar{x}, \bar{\lambda}, \bar{\mu})$. Moreover, if $\varphi_0(t) = O(t^2)$, $\varphi_1(t) = O(t^2)$ and $\varphi_2(t) = O(t^2)$, then

$$\chi_j(x, \lambda, \mu) = O(\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|^2), \quad j = 1, 2.$$

Observe now that all the considerations in Section 3 remain valid if we replace $\bar{\pi}(x)$ by $\pi(x, \lambda, \mu)$ (with the evident modifications of estimates (3.7)–(3.9), where one should replace $\|x - \bar{x}\|$ by $\|(x - \bar{x}, \lambda - \bar{\lambda}, \mu - \bar{\mu})\|$). This follows from (4.1), (4.5), and (4.10), implying, in particular, that $\pi(x, \lambda, \mu) \to \bar{x} = \pi(\bar{x}, \bar{\lambda}, \bar{\mu})$ and

$$h(\pi(x, \lambda, \mu)) = o(\|h(x)\|) = o(\|x - \bar{x}\|)$$

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

The needed results now follow by applying Theorem 2.1. ∎

5 Inexact restoration and degenerate problems

As already mentioned in Section 1, the analysis in [3] does not assume any constraint qualification (CQ), and the primal-dual solution (even the primal part) does not have to be isolated. In fact, the analysis in [3] is actually quite similar to that in [9] for the abstract Newtonian framework for GE problems with nonisolated solutions. In particular, relations (3)–(7) in [3] give a version of the inexact restoration method subproblem, combined with the stabilizing

localization conditions (4.6) and (4.7). However, in [3] solvability of subproblems is not established – it is assumed that subproblems have solutions satisfying the needed localization conditions; see [3, Theorem 2.3]. The discussion that follows is intended to demonstrate that without CQs, this property of subproblems' solvability with the required localization of solutions is at least questionable.

To apply the general convergence result from [9], we need the conditions (4.1) and the set of relations

$$\left\| \frac{\partial L}{\partial x}(x + \xi, \, \lambda) + (h'(\pi))^{\mathrm{T}} \eta - \mu - \zeta \right\| \leq \varphi_1 \left(\left\| \frac{\partial L}{\partial x}(\pi, \, \lambda) - \mu \right\| \right), \tag{5.1}$$

$$\|h'(\pi)(x + \xi - \pi)\| \leq \varphi_2 \left(\left\| \frac{\partial L}{\partial x}(\pi, \, \lambda) - \mu \right\| \right) \tag{5.2}$$

$$\mu + \zeta \geq 0, \quad x + \xi \geq 0, \quad \langle \mu + \zeta, \, x + \xi \rangle = 0 \tag{5.3}$$

(cf. (4.2)–(4.4)) to have a solution $(\pi, \, \xi, \, \eta, \, \zeta)$ satisfying

$$\|(\xi, \, \eta, \, \zeta)\| = O(\|x - \bar{x}\| + \mathrm{dist}((\lambda, \, \mu), \, \mathcal{M}(\bar{x}))). \tag{5.4}$$

In particular, the above at least must hold, as it is implied by (4.6) and (4.7). Quite remarkably, such a triple always exists – just take $\pi = \bar{x}$, $\xi = \bar{x} - x$, and $(\eta, \, \zeta)$ such that $(\lambda + \eta, \, \mu + \zeta)$ is the projection of $(\lambda, \, \mu)$ onto $\mathcal{M}(\bar{x})$. Therefore, the inexact restoration scheme actually admits one-step termination at an exact primal-dual solution even when localization conditions are imposed. The problem, of course, is that this view of the inexact restoration scheme would not have much to do with reality, as finite termination cannot be expected for any reasonable algorithm. In the case of inexact restoration perhaps even more so, as the method consists of the two subsequent phases, and for finite termination one has to obtain the exact primal solution $\bar{x}$ as π^k at the feasibility phase, which does not even involve the objective function f! Also, it appears that in order to formally apply the iterative framework of [9], one has to follow the line of the analysis giving Theorem 4.1. Specifically, it is more practical to consider any $\pi(x, \, \lambda, \, \mu)$ satisfying (4.1) and (4.5), and to ask whether there exists $(\xi, \, \eta, \, \zeta)$ satisfying (5.1)–(5.4), with $\pi = \pi(x, \, \lambda, \, \mu)$. The next example demonstrates that without any CQ, the answer to this question is in general negative. More precisely, by means of sensitivity theory for optimization problems it can be shown that under some reasonable assumptions, the primal estimate in (5.4) can be achieved, but the dual one is problematic.

Example 5.1 Let $n = 2$, $l = 1$, $f(x) = (x_1^2 + x_2^2)/2$, $h(x) = (x_1 - x_2^2)x_2^2$. Then $\bar{x} = 0$ is a solution of the problem (1.1), $\mathcal{M}(\bar{x}) = \mathbb{R} \times \{0\}$, all the multipliers are noncritical and even satisfy the SOSC.

Let $x = \pi = \mu = (t^2, \, t)$ with $t > 0$ (observe that this $x = \pi$ is feasible), and let $\lambda = 0$. Then

$$\frac{\partial L}{\partial x}(\pi, \, \lambda) - \mu = 0,$$

and (4.2)–(4.4) take the form

$$\xi_1 + \eta t^2 - \zeta_1 = 0, \quad \xi_2 - 2\eta t^3 - \zeta_2 = 0, \quad t^2 + \xi_1 - 2t(t + \xi_2) + t^2 = 0, \tag{5.5}$$

$$t^2 + \zeta_1 \geq 0, \quad t^2 + \xi_1 \geq 0, \quad (t^2 + \zeta_1)(t^2 + \xi_1) = 0,$$
$$t + \zeta_2 \geq 0, \quad t + \xi_2 \geq 0, \quad (t + \zeta_2)(t + \xi_2) = 0. \tag{5.6}$$

If $t + \xi_2 = 0$ then, by (5.6), the last equality in (5.5) cannot hold, and therefore, $t + \xi_2 > 0$, $t + \zeta_2 = 0$.

If $t^2 + \xi_1 = 0$, then from (5.5) we derive that $\xi_2 = -t/2$, $\eta = 1/(4t^2) \to \infty$ as $t \to 0$, and hence, (5.4) cannot hold.

If $t^2 + \xi_1 > 0$ then, by (5.6), $t^2 + \zeta_1 = 0$, and by (5.6), $\eta = 1/(1 + 4t^2) \to 1$ as $t \to 0$, and (5.4) cannot hold again.

6 Composite-step SQP

As is well known (see, e.g., [6]), the exact SQP step defined by the subproblem (2.1) can be decomposed into the following two phases. The normal phase consists of finding π^k as a point satisfying the linearized equality constraints

$$h(x) + h'(x)(\pi - x) = 0 \tag{6.1}$$

for $x = x^k$. This is followed by the tangential phase, where one computes x^{k+1} and $(\lambda^{k+1}, \mu^{k+1})$ as a stationary point and an associated Lagrange multiplier of the QP subproblem

$$\text{minimize}_z \quad \left\langle f'(x) + \frac{\partial^2 L}{\partial x^2}(x, \lambda)(\pi - x), z - \pi \right\rangle + \frac{1}{2}\left\langle \frac{\partial^2 L}{\partial x^2}(x, \lambda)(z - \pi), z - \pi \right\rangle$$
$$\text{subject to} \quad h'(x)(z - \pi) = 0, \; z \geq 0$$

for $x = x^k$, $\pi = \pi^k$, and $\lambda = \lambda^k$.

One can readily see that the result of this two-phase iteration is exactly the same as that of the SQP iteration, so this is merely an interpretation of the latter. However, this decomposition is crucially useful in the context of trust-region globalization strategies for SQP, as it allows to avoid possible infeasibility of subproblems when an additional trust-region constraint is added in (1.3) (i.e., when (1.3) itself is feasible but its solutions may lie outside of the trust-region around the current iterate x^k, imposed for globalization purposes). We refer the reader to [6, Section 15.4] for details. We also note that some extra requirements may appear at the normal phase, such as nonnegativity of π, or the requirement that π is the projection of x onto the set given by the linearized constraints (thus the name "normal phase"; see [6, Section 15.4.4]). These requirements can be needed for designing globally convergent algorithms, but they are not needed for the local convergence analysis via the pSQP framework given below. Note also that the subproblems of both phases need not be solved exactly; moreover, solving them exactly would certainly be wasteful. In fact, satisfying some appropriate mild model reduction conditions does the job for establishing global convergence [6, Section 15.4]. As for local convergence, the conditions and rules for updating the trust-region parameter allow to expect that under some natural assumptions, and perhaps employing second-order corrections, the trust-region radius is locally constant for the tail of the sequence [6, Section 15.4.4]. Then, under the assumptions in question, the trust-region constraint is locally inactive and thus does not interfere with the local convergence

15

analysis of composite-step SQP without trust-region. Implicitly assuming this, we next study which perturbations in both phases of this two-phase interpretation of SQP do not destroy its local superlinear convergence.

We now state the *inexact composite-step SQP method*, allowing for inexactness in both the normal and the tangential phases.

Algorithm 6.1 Choose functions φ_0, φ_1, $\varphi_2 : \mathbb{R}_+ \to \mathbb{R}_+$. Choose $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ and set $k = 0$.

1. If (x^k, λ^k, μ^k) satisfies the KKT system (1.2), stop.

2. (Normal phase.) Compute $\pi^k \in \mathbb{R}^n$ satisfying

$$\|h(x) + h'(x)(\pi - x)\| \leq \varphi_0 \left(\left\| \left(\frac{\partial L}{\partial x}(x, \lambda) - \mu, h(x) \right) \right\| \right) \tag{6.2}$$

   for $x = x^k$, $\lambda = \lambda^k$, and $\mu = \mu^k$.

3. (Tangential phase.) Compute $z = x^{k+1}$ and $(u, v) = (\lambda^{k+1}, \mu^{k+1})$ satisfying

$$\left\| f'(x) + \frac{\partial^2 L}{\partial x^2}(x, \lambda)(z - x) + (h'(x))^{\mathrm{T}} u - v \right\| \leq \varphi_1 \left(\left\| \left(\frac{\partial L}{\partial x}(x, \lambda) - \mu, h(x) \right) \right\| \right), \tag{6.3}$$

$$\|h'(x)(z - \pi)\| \leq \varphi_2 \left(\left\| \left(\frac{\partial L}{\partial x}(x, \lambda) - \mu, h(x) \right) \right\| \right), \tag{6.4}$$

$$v \geq 0, \quad z \geq 0, \quad \langle v, z \rangle = 0 \tag{6.5}$$

   for $x = x^k$, $\pi = \pi^k$, $\lambda = \lambda^k$, and $\mu = \mu^k$.

4. Increase k by 1 and go to step 1.

Observe further that all the derivatives in Algorithm 6.1 are computed at x^k, unlike for exact and inexact restoration schemes, where all the derivatives are computed at π^k produced by the feasibility phase. One consequence of this is that the inexact composite-step SQP method does not give rise to any structural perturbations of the basic SQP, other than those coming from (6.2): as will be seen below, all the other perturbations can be naturally interpreted as truncations of the SQP iteration.

Indeed, for $(x, \lambda, \mu) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ set

$$\omega_2(x, \lambda, \mu) = -h(x) - h'(x)(\pi(x, \lambda, \mu) - x), \tag{6.6}$$

where $\pi(x, \lambda, \mu)$ is the point selected at the normal phase of the algorithm. Observe that π satisfying (6.1) (and even more so (6.2)) always exists if rank $h'(x) = l$.

Comparing (6.3)–(6.5) with the corresponding relations in (2.2) it is now evident that an iteration of Algorithm 6.1 can be seen as solving (2.2) with ω_1 identically equal to zero, and ω_2 defined in (6.6) (which agrees with taking ψ identically equal to zero as well), and with

$$\chi_1(x, \lambda, \mu) = \varphi_1 \left(\left\| \left(\frac{\partial L}{\partial x}(x, \lambda) - \mu, h(x) \right) \right\| \right),$$

$$\chi_2(x,\,\lambda,\,\mu) = \varphi_2\left(\left\|\left(\frac{\partial L}{\partial x}(x,\,\lambda) - \mu,\, h(x)\right)\right\|\right).$$

Moreover, the needed properties of ω_2 readily follow from (6.2) and (6.6), and similarly to Theorem 4.1, we immediately obtain the following local superlinear convergence result for the inexact composite-step SQP method.

Theorem 6.1 *Under the assumptions of Theorem 3.1, suppose that if $(x^k,\,\lambda^k,\,\mu^k) = (x^j,\,\lambda^j,\,\mu^j)$ for any two iteration indices k and j, then step 2 of Algorithm 6.1 computes $\pi^k = \pi^j$.*

Then for any functions $\varphi_0,\,\varphi_1,\,\varphi_2 : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\varphi_0(t) = o(t)$, $\varphi_1(t) = o(t)$ and $\varphi_2(t) = o(t)$ as $t \to 0$, there exist $\varepsilon > 0$ and $\delta > 0$ such that for any starting point $(x^0,\,\lambda^0,\,\mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ satisfying (3.10), if an iterative sequence $\{(x^k,\,\lambda^k,\,\mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ is generated by Algorithm 6.1 and satisfies (3.11) for each $k = 0,\,1,\,\dots$, then this sequence converges to $(\bar{x},\,\bar{\lambda},\,\bar{\mu})$, and the rate of convergence is superlinear. Moreover, $\varepsilon > 0$ and $\delta > 0$ can be chosen in such a way that for any starting point $(x^0,\,\lambda^0,\,\mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^n$ satisfying (3.10), there exists at least one sequence conforming, for each $k = 0,\,1,\,\dots$, to Algorithm 6.1 and satisfying the condition (3.11).

In addition, the rate of convergence is quadratic if the second derivatives of f and h are locally Lipschitz-continuous with respect to $\bar{x}$, and if $\varphi_0(t) = O(t^2)$, $\varphi_1(t) = O(t^2)$ and $\varphi_2(t) = O(t^2)$ as $t \to 0$.

Algorithm 6.1 covers both the Vardi-like approach [26] to composite-step trust-region SQP methods (where φ_0 is supposed to be identically equal to zero; see [6, Section 15.4.1]), as well as Byrd–Omojokun-like methods (where φ_0 is allowed to be nonzero; see [24] and [6, Section 15.4.2]). The corresponding rate-of-convergence results in [6, Theorems 15.4.15, 15.4.24] assume that the pure SQP step is taken whenever possible (i.e., whenever it can be accepted by the trust-region and model reduction rules). Theorem 6.1 demonstrates that in fact there is no need to perform the SQP step exactly in order to preserve superlinear convergence. As for Celis–Dennis–Tapia-like approaches (see [5] and [6, Section 15.4.3]), these are not really composite-step methods: they only involve truncation in the linearized equality constraints part; therefore, they can be embedded into the pSQP framework directly.

References

[1] B.T. Boggs and J.W. 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. Springer–Verlag, Berlin, Germany, 2006. Second Edition.

[3] E.G. Birgin and J.M. Martínez. Local convergence of an inexact-restoration method and numerical experiments. J. Optim. Theory Appl. 127 (2005), 229–247.

[4] L.F. Bueno, A. Friedlander, J.M. Martnez, and F.N.C. Sobral. Inexact restoration method for derivative-free optimization with smooth constraints. SIAM J. Optim. 23 (2013), 1189–1213.

[5] M.R. Celis, J.E. Dennis, and R.A. Tapia R.A. A trust region algorithm for nonlinear equality constrained optimization. In P.T. Boggs, R.H. Byrd, and R.B. Schnabel, eds., Numerical Optimization, pp. 71–82. SIAM, Philadelphia, 1985.

[6] A.R. Conn, N.I.M. Gould, and Ph.L. Toint. Trust-Region Methods. SIAM, Philadelphia, 2000.

[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] D. Fernández, E.A. Pilotta, and G.A. Torres. An inexact restoration strategy for the globalization of the sSQP method. Comput. Optim. Appl. 54 (2013), 595–617.

[9] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. 94 (2002), 91–124.

[10] Fischer A., Friedlander A. A new line search inexact restoration approach for nonlinear programming. Comput. Optim. Appl. 46 (2010), 333–346.

[11] P.E. Gill and E. Wong. Sequential quadratic programming methods. In J. Lee and S. Leyffer, eds., Mixed Integer Nonlinear Programming, The IMA Volumes in Mathematics and its Applications, Volume 154, Springer-Verlag, pp. 147–224, 2012.

[12] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. Math. Program. 142 (2013), 591–604.

[13] A.F. Izmailov, A.S. Kurennoy, and M.V. Solodov. Local convergence of the method of multipliers for variational and optimization problems under the sole noncriticality assumption. August 2013 (Revised January 2014). Available at http://pages.cs.wisc.edu/~solodov/solodov.html

[14] A.F. Izmailov and M.V. Solodov. On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. Math. Program. 117 (2009), 271–304.

[15] A.F. Izmailov and M.V. Solodov. Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. Comput. Optim. Appl. 42 (2009), 231–264.

[16] 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.

[17] A.F. Izmailov and M.V. Solodov. Inexact Josephy–Newton framework for generalized equations and its applications to local analysis of Newtonian methods for constrained optimization. Comput. Optim. Appl. 46 (2010), 347–368.

[18] A.F. Izmailov and M.V. Solodov. On attraction of linearly constrained Lagrangian methods and of stabilized and quasi-Newton SQP methods to critical multipliers. Math. Program. 126 (2011), 231–257.

[19] A.F. Izmailov and M.V. Solodov. Stabilized SQP revisited. Math. Program. 122 (2012), 93–120.

[20] A.F. Izmailov and M.V. Solodov. Newton-Type Methods for Optimization and Variational Problems. Springer Series in Operations Research and Financial Engineering, 2014. `http://www.springer.com/mathematics/applications/book/978-3-319-04246-6`

[21] J.M. Martínez. Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. J. Optim. Theory Appl. 111 (2001), 39–58.

[22] J.M. Martínez and E.A. Pilotta. Inexact restoration algorithms for constrained optimization. J. Optim. Theory Appl. 104 (2000), 135–163.

[23] J.M. Martínez and E.A. Pilotta. Inexact restoration methods for nonlinear programming: advances and perspectives. In Optimization and Control with Applications, Springer, New York, pp. 271–292, 2005.

[24] E.O. Omojokun. Trust region algorithms for optimization with nonlinear equality and inequality constraints. Ph.D. thesis. Dept. of Computer Science, University of Colorado at Boulder, U.S.A., 1989.

[25] M.V. Solodov. Constraint qualifications. Wiley Encyclopedia of Operations Research and Management Science, J.J. Cochran, ed. John Wiley & Sons, Inc., 2010.

[26] A. Vardi. A trust region algorithm for equality constrained minimization: Convergence properties and implementation. SIAM J. Numer. Anal. 22 (1985), 575–591.