

LOCAL CONVERGENCE OF EXACT AND INEXACT AUGMENTED LAGRANGIAN METHODS UNDER THE SECOND-ORDER SUFFICIENT OPTIMALITY CONDITION *

D. FERNÁNDEZ † AND M. V. SOLODOV ‡

Abstract. We establish local convergence and rate of convergence of the classical augmented Lagrangian algorithm under the sole assumption that the dual starting point is close to a multiplier satisfying the second-order sufficient optimality condition. In particular, no constraint qualifications of any kind are needed. Previous literature on the subject required, in addition, the linear independence constraint qualification and either the strict complementarity assumption or a stronger version of the second-order sufficient condition. That said, the classical results allow the initial multiplier estimate to be far from the optimal one, at the expense of proportionally increasing the threshold value for the penalty parameters. Although our primary goal is to avoid constraint qualifications, if the stronger assumptions are introduced then starting points far from the optimal multiplier are allowed within our analysis as well. Using only the second-order sufficient optimality condition, for penalty parameters large enough we prove primal-dual Q -linear convergence rate, which becomes superlinear if the parameters are allowed to go to infinity. Both exact and inexact solutions of sub-problems are considered. In the exact case, we further show that the primal convergence rate is of the same Q -order as the primal-dual rate. Previous assertions for the primal sequence all had to do with the weaker R -rate of convergence and required the stronger assumptions cited above. Finally, we show that under our assumptions one of the popular rules of controlling the penalty parameters ensures their boundedness.

Key words. Augmented Lagrangian, method of multipliers, second-order sufficient optimality condition, linear convergence, superlinear convergence.

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

1. Introduction. Given twice continuously differentiable functions $f : \mathbb{R}^n \rightarrow \mathbb{R}$ and $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$, we consider the optimization problem

$$(1.1) \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_i(x) = 0, \quad i = 1, \dots, l, \\ & g_i(x) \leq 0, \quad i = l + 1, \dots, m, \end{array}$$

where $0 \leq l \leq m$. One of the fundamental approaches to solving (1.1) is the augmented Lagrangian algorithm, known also as the method of multipliers. The method dates back to [19, 26, 27, 28], and had been further developed and studied from various angles in [33, 3, 9, 20, 13, 10, 32, 5, 1, 2, 6, 7, 23], among extensive other literature (see also the monographs [4, 25, 12, 31]). Successful software based on the augmented Lagrangians includes LANCELOT [11] and ALGENCAN [34].

We next describe the basic iteration of the algorithm in question. The augmented Lagrangian function for (1.1) is defined by $\bar{L} : \mathbb{R}^n \times \mathbb{R}^m \times (0, \infty) \rightarrow \mathbb{R}$,

$$(1.2) \quad \bar{L}(x, \mu; \rho) = f(x) + \sum_{i=1}^l (\mu_i g_i(x) + \frac{\rho}{2} g_i(x)^2) + \frac{1}{2\rho} \sum_{i=l+1}^m (\max\{0, \mu_i + \rho g_i(x)\}^2 - \mu_i^2).$$

*Received by the editors October 2010; revised July 2011.

†FaMAF, Universidad Nacional de Córdoba, Medina Allende s/n, 5000 Córdoba, Argentina (dfernand@impa.br).

‡IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil (solodov@impa.br). Research of this author is supported in part by CNPq Grants 300513/2008-9, by PRONEX–Optimization, and by FAPERJ.

Then given the current multiplier estimate $\mu^k \in \mathbb{R}^m$ and the current penalty parameter $\rho_k > 0$, the (exact) augmented Lagrangian method generates the next iterate $(x^{k+1}, \mu^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^m$ as follows:

$$(1.3) \quad \begin{aligned} x^{k+1} & \text{ is a solution of } \min_{x \in \mathbb{R}^n} \bar{L}(x, \mu^k; \rho_k), \\ \mu_i^{k+1} & = \mu_i^k + \rho_k g_i(x^{k+1}), & i = 1, \dots, l, \\ \mu_i^{k+1} & = \max\{0, \mu_i^k + \rho_k g_i(x^{k+1})\}, & i = l+1, \dots, m. \end{aligned}$$

Before explaining our contributions to the analysis of this algorithm, we need to introduce some notation. Stationary points of problem (1.1) and the associated Lagrange multipliers are characterized by the Karush–Kuhn–Tucker (KKT) system

$$(1.4) \quad \begin{aligned} 0 & = \frac{\partial L}{\partial x}(x, \mu), \\ 0 & = g_i(x), & i = 1, \dots, l, \\ 0 \leq \mu_i, \quad g_i(x) \leq 0, \quad \mu_i g_i(x) & = 0, & i = l+1, \dots, m, \end{aligned}$$

where $L : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ is the Lagrangian function of problem (1.1), i.e.,

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

We denote by $\mathcal{M}(\bar{x})$ the set of Lagrange multipliers of problem (1.1) associated with a given stationary point \bar{x} , i.e., $\mu \in \mathcal{M}(\bar{x})$ if and only if (\bar{x}, μ) satisfies the KKT system (1.4). Let

$$\mathcal{I} = \mathcal{I}(\bar{x}) = \{i \in \{1, \dots, m\} \mid g_i(\bar{x}) = 0\}, \quad \mathcal{J} = \mathcal{J}(\bar{x}) = \{1, \dots, m\} \setminus \mathcal{I},$$

be the sets of indices of active and inactive constraints at \bar{x} , respectively. For each $\mu \in \mathcal{M}(\bar{x})$, we introduce the following standard partition of the set \mathcal{I} :

$$\mathcal{I}_1(\mu) = \{1, \dots, l\} \cup \{i \in \{l+1, \dots, m\} \mid \mu_i > 0\}, \quad \mathcal{I}_0(\mu) = \mathcal{I} \setminus \mathcal{I}_1(\mu).$$

The critical cone of problem (1.1) at its stationary point \bar{x} is given by

$$(1.5) \quad \mathcal{C}(\bar{x}) = \left\{ u \in \mathbb{R}^n \mid \begin{aligned} & \langle f'(\bar{x}), u \rangle = 0, \quad \langle g'_i(\bar{x}), u \rangle = 0 \text{ for } i \in \{1, \dots, l\}, \\ & \langle g'_i(\bar{x}), u \rangle \leq 0 \text{ for } i \in \{l+1, \dots, m\} \text{ with } g_i(\bar{x}) = 0 \end{aligned} \right\} \\ = \left\{ u \in \mathbb{R}^n \mid g'_{\mathcal{I}_1(\hat{\mu})}(\bar{x})u = 0, \quad g'_{\mathcal{I}_0(\hat{\mu})}(\bar{x})u \leq 0 \right\},$$

where the second equality is independent of the choice of $\hat{\mu} \in \mathcal{M}(\bar{x})$.

We say that the second-order sufficient optimality condition (SOSC) is satisfied at $(\bar{x}, \bar{\mu})$ with $\bar{\mu} \in \mathcal{M}(\bar{x})$ if

$$(1.6) \quad \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\mu})u, u \right\rangle > 0 \quad \forall u \in \mathcal{C}(\bar{x}) \setminus \{0\}.$$

The usual convergence rate statements for the augmented Lagrangian methods assume the linear independence constraint qualification (i.e., that the gradients $\{g'_i(\bar{x}), i \in \mathcal{I}\}$ are linearly independent, and hence the multiplier $\bar{\mu} \in \mathcal{M}(\bar{x})$ is unique), strict complementarity (i.e., $\bar{\mu}_i > 0$ for all $i \in \mathcal{I}$) and SOSC (1.6). The assertion is that for any initial multiplier estimate μ^0 if $\rho_k \geq \hat{\rho} > 0$ for all k (the farther μ^0 is from the unique optimal multiplier $\bar{\mu}$ the larger should be the penalty parameter threshold $\hat{\rho}$), then the dual sequence generated by (1.3) converges to $\bar{\mu}$ with Q -linear

rate and the primal sequence converges to \bar{x} with R -linear rate. Convergence becomes superlinear if $\rho_k \rightarrow +\infty$. Various versions of such statements can be found, e.g., in [4, Prop. 3.2 and 2.7], [25, Thm. 17.6], [12], [31, Thm. 6.16]. Strict complementarity is not assumed in [13], but a stronger version of second-order sufficiency is employed. Strict complementarity is also not used in [20]. However, the linear independence constraint qualification is required in all the literature.

In this paper, we prove that if the initial multiplier estimate μ^0 is close enough to $\bar{\mu}$ satisfying SOSC (1.6) and the penalty parameters are chosen large enough, then the primal-dual sequence converges Q -linearly to $(\bar{x}, \hat{\mu})$ with some $\hat{\mu} \in \mathcal{M}(\bar{x})$. In particular, no constraint qualifications of any kind are needed. The multipliers set can even be unbounded. Strict complementarity is also not assumed. As usual, convergence becomes superlinear if $\rho_k \rightarrow +\infty$. However, in most implementations it is accepted that the penalty parameters should stay bounded. In this respect, we show that they indeed stay bounded under our assumption of SOSC (1.6), if updated as proposed in [7]. This is also the first result to this effect not assuming any constraint qualifications. Both exact and inexact solutions of subproblems are considered. Furthermore, for the exact solutions of subproblems we show that the primal convergence rate is of the same Q -order as the primal-dual rate. To the best of our knowledge, this is the first result asserting Q -rate of convergence (rather than the weaker R -rate) of the primal sequence generated by the augmented Lagrangian method, under any assumptions. We also show that if the linear independence constraint qualification and the strong second-order sufficient optimality condition are introduced, then initial values of multiplier estimates far from the optimal one can be used in our analysis as well, similar to the classical results.

It is worth commenting that when it comes to the need (or not) for constraint qualifications, it is reasonable to conjecture that in augmented Lagrangian methods they may be not required. Indeed, at least part of the role of constraint qualifications in convergence analyses of optimization methods is to ensure that subproblems are feasible; for example when constraints are linearized as in the standard sequential quadratic programming [8] or in the linearized (augmented) Lagrangian methods [24]. In such cases, some constraint qualification is clearly unavoidable. In the augmented Lagrangian methods, on the other hand, subproblems are unconstrained and there is no obvious reason why assumptions about the constraints should be required in the analysis. Thus removing them from consideration looks appealing and potentially possible. We note, in passing, that the situation is somewhat similar to the stabilized sequential quadratic programming method, where constraint qualifications are also not needed [17]. For this method feasibility of subproblems is also not an issue, although the reason is different: subproblems are always feasible thanks to a certain “elastic mode” feature.

Our approach to convergence of the augmented Lagrangian methods is in the spirit of some recent analyses of Newtonian frameworks for generalized equations in [18] and [21], and their applications to methods for constrained optimization in [17, 16, 22]. The framework of [18] allows nonisolated solutions (and thus nonunique multipliers), and had been employed in [17, 22] to prove convergence of the stabilized sequential quadratic programming method without assuming constraint qualifications. The framework of [21] requires solutions to be isolated (thus the strict Mangasarian–Fromovitz constraint qualification is needed), but reveals that Newtonian lines of analysis might be applicable to some algorithms that are not commonly recognized as being of Newton type, in the sense that their subproblems are not systems of equations

or quadratic programs. For example, such is the case of the linearly constrained augmented Lagrangian method [24], for which improved local convergence results have been obtained in [21, 16] relating the method in question to a certain perturbation of Newtonian iterates. The present paper develops Newtonian analysis of the augmented Lagrangian method by combining the ideas of [18] in dealing with nonisolated solutions in Newton methods with the ideas of [21, 16] in dealing with methods that perhaps “do not look” like Newton methods, but can be re-interpreted in such terms a posteriori by introducing perturbations. The applicability of this line of analysis to the augmented Lagrangian method is actually somewhat surprising, considering that no linear or quadratic approximations appear in (1.3).

In practice, subproblems of the augmented Lagrangian method are often solved approximately, in the sense that the iterates of the unconstrained minimization method used to minimize the augmented Lagrangian are truncated. Specifically, instead of computing an “exact” minimizer in (1.3), a point x^{k+1} satisfying

$$\left\| \frac{\partial \bar{L}}{\partial x}(x^{k+1}, \mu^k; \rho_k) \right\| \leq \epsilon_k$$

is accepted, where $\epsilon_k \geq 0$ is the current approximation tolerance. In theoretical analysis, $\{\epsilon_k\}$ is often an exogenous sequence of scalars converging to zero. In what follows, we define ϵ_k as a specific computable quantity that depends on the violation of KKT conditions (1.4) for problem (1.1) by the point (x^k, μ^k) . This technique is also related to some truncation conditions used in [1]. Thus we prove convergence and rate of convergence for both the inexact/truncated version and the exact method.

The rest of the paper is organized as follows. In Section 2 we state our algorithmic framework and interpret augmented Lagrangian iterates as perturbations of solutions of generalized equations associated to the KKT conditions of the problem. We also sketch the general lines of our convergence analysis. Details are worked out in the next Section 3. Section 4 presents some further related developments. In particular, in Section 4.1 we establish boundedness of the penalty parameters if the latter are generated by one of the popular update rules. Some relations with the classical results are discussed in Section 4.2, where it is shown that under the stronger assumptions initial multiplier estimates far from the optimal one can be used in our analysis as well. Section 4.3 explains that our results extend to the practically important case where linear constraints of the problem are passed as constraints to the subproblems, while the augmented Lagrangian involves general constraints only.

We finish this section with describing our notation. We use $\langle \cdot, \cdot \rangle$ to denote the Euclidean inner product, $\| \cdot \|$ the associated norm, B the closed unit ball, and I the identity matrix (the space is always clear from the context). For any matrix M , $M_{\mathcal{I}}$ denotes the submatrix of M with rows indexed by the set \mathcal{I} . When in matrix notation, vectors are considered columns, and for a vector x we denote by $x_{\mathcal{I}}$ the subvector of x with coordinates indexed by \mathcal{I} . For a set $S \subset \mathbb{R}^q$ and a point $z \in \mathbb{R}^q$, the distance from z to S is defined as $\text{dist}(z, S) = \inf_{s \in S} \|z - s\|$. Then $\Pi_S(z) = \{s \in S \mid \text{dist}(z, S) = \|z - s\|\}$ is the set of all points in S that have minimal distance to z . For a cone $K \subset \mathbb{R}^q$, its polar (negative dual) is $K^\circ = \{\xi \in \mathbb{R}^q \mid \langle z, \xi \rangle \leq 0 \forall z \in K\}$. Recall that for any closed convex cone $K \subset \mathbb{R}^q$ and $z \in \mathbb{R}^q$ it holds that

$$(1.7) \quad \bar{z} = \Pi_K(z) \quad \Leftrightarrow \quad K \ni \bar{z} \perp z - \bar{z} \in K^\circ,$$

where the notation $u \perp v$ means that $\langle u, v \rangle = 0$. We use the notation $\psi(t) = o(t)$ for any function $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}^q$ such that $\lim_{t \rightarrow 0+} t^{-1}\psi(t) = 0$.

2. The algorithmic framework and preliminary considerations. In this section, we formally state the algorithm under consideration and discuss our interpretation of its iterates as solutions of perturbed generalized equations associated to the KKT system (1.4).

To simplify the notation, we define the closed convex cone

$$Q = \{\nu \in \mathbb{R}^m \mid \nu_i \in \mathbb{R}, i = 1, \dots, l; \nu_i \geq 0, i = l + 1, \dots, m\}.$$

Noting that $Q^\circ = \{\xi \in \mathbb{R}^m \mid \xi_i = 0, i = 1, \dots, l; \xi_i \leq 0, i = l + 1, \dots, m\}$, our problem (1.1) takes the compact form

$$(2.1) \quad \begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \in Q^\circ, \end{aligned}$$

with its KKT system (1.4) given by

$$(2.2) \quad \begin{aligned} 0 &= \frac{\partial L}{\partial x}(x, \mu), \\ Q \ni \mu &\perp g(x) \in Q^\circ. \end{aligned}$$

The violation of the KKT conditions (2.2) is measured by the natural residual $\sigma : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}_+$,

$$(2.3) \quad \sigma(x, \mu) = \left\| \begin{bmatrix} f'(x) + (g'(x))^\top \mu \\ \mu - \Pi_Q(\mu + g(x)) \end{bmatrix} \right\|.$$

In particular, $\sigma(x, \mu) = 0$ if and only if (x, μ) solves (2.2). Define the function $G : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n \times \mathbb{R}^m$ and the set-valued mapping \mathcal{N} from $\mathbb{R}^n \times \mathbb{R}^m$ to the subsets of $\mathbb{R}^n \times \mathbb{R}^m$ by

$$(2.4) \quad G(x, \mu) = \begin{bmatrix} \frac{\partial L}{\partial x}(x, \mu) \\ -g(x) \end{bmatrix}, \quad \mathcal{N}(x, \mu) = \{0\} \times \mathcal{N}_Q(\mu),$$

where

$$\mathcal{N}_Q(\mu) = \begin{cases} \{v \in \mathbb{R}^m \mid \langle v, \nu - \mu \rangle \leq 0 \ \forall \nu \in Q\} & \text{if } \mu \in Q, \\ \emptyset, & \text{otherwise,} \end{cases}$$

is the normal cone to Q at $\mu \in \mathbb{R}^m$. Then the KKT system (1.4) (or (2.2)) is further equivalent to the generalized equation (GE)

$$(2.5) \quad 0 \in G(w) + \mathcal{N}(w), \quad w = (x, \mu) \in \mathbb{R}^n \times \mathbb{R}^m.$$

In the present notation, the augmented Lagrangian function (1.2) is

$$(2.6) \quad \bar{L}(x, \mu; \rho) = f(x) + \frac{1}{2\rho} \left(\|\Pi_Q(\mu + \rho g(x))\|^2 - \|\mu\|^2 \right),$$

and its derivatives are given by

$$\frac{\partial \bar{L}}{\partial x}(x, \mu; \rho) = f'(x) + (g'(x))^\top \Pi_Q(\mu + \rho g(x)),$$

$$\frac{\partial \bar{L}}{\partial \mu}(x, \mu; \rho) = \frac{1}{\rho} (\Pi_Q(\mu + \rho g(x)) - \mu).$$

For an arbitrary but fixed $\hat{c} \in (0, +\infty)$, we consider the following iterative procedure. Given a current iterate $(x^k, \mu^k) \in \mathbb{R}^n \times Q$ with $\sigma(x^k, \mu^k) > 0$, a penalty parameter $\rho_k > 0$ and an approximation parameter $\epsilon_k \geq 0$,

$$(2.7) \quad \text{find } x^{k+1} \text{ satisfying } \left\| \frac{\partial \bar{L}}{\partial x}(x^{k+1}, \mu^k; \rho_k) \right\| \leq \epsilon_k,$$

$$(2.8) \quad \left\| \left[\begin{array}{c} x^{k+1} - x^k \\ \Pi_Q(\mu^k + \rho_k g(x^{k+1})) - \mu^k \end{array} \right] \right\| \leq \hat{c} \sigma(x^k, \mu^k);$$

$$(2.9) \quad \text{and set } \mu^{k+1} = \Pi_Q(\mu^k + \rho_k g(x^{k+1})).$$

Some comments are in order. The condition (2.8) is a localization-type condition. Some condition of this kind is unavoidable in local convergence analyses of most algorithms, and the augmented Lagrangian method is no exception [4, 25, 12, 31]. Without extremely strong assumptions, the augmented Lagrangian may have stationary points (as well as minimizers) arbitrarily far from the solution of interest (equivalently, from the current iterate x^k). Such points must be discarded in any meaningful local analysis. This is precisely the role of (2.8). In practical implementations, it is ignored, of course. Another point is that this condition is formally not well-defined for $k = 0$ (the algorithm starts with the initial dual estimate $\mu^0 \in Q$ and generates x^1 , i.e., we do not have x^0). This detail should not introduce any confusion, however, as we can think of (2.7)–(2.9) starting with the index $k = 1$. Alternatively, we could take x^0 arbitrary and fix \hat{c} a posteriori, after x^1 is obtained, choosing it large enough to satisfy (2.8) for $k = 0$.

We next relate the iterative process (2.7)–(2.9) to the study of GE (2.5) under perturbations. By (2.7), there exists $\vartheta \in B$ such that

$$\frac{\partial L}{\partial x}(x^{k+1}, \mu^{k+1}) + \epsilon_k \vartheta = 0,$$

where we used the fact that

$$(2.10) \quad \frac{\partial L}{\partial x}(x^{k+1}, \mu^{k+1}) = \frac{\partial \bar{L}}{\partial x}(x^{k+1}, \mu^k; \rho_k).$$

Also, by the definition of μ^{k+1} , we have that

$$(2.11) \quad \mu^k + \rho_k g(x^{k+1}) - \mu^{k+1} \in \mathcal{N}_Q(\mu^{k+1}).$$

Hence,

$$0 \in \left[\begin{array}{c} \frac{\partial L}{\partial x}(x^{k+1}, \mu^{k+1}) + \epsilon_k \vartheta \\ -g(x^{k+1}) + \frac{1}{\rho_k}(\mu^{k+1} - \mu^k) \end{array} \right] + \mathcal{N}(x^{k+1}, \mu^{k+1}).$$

Defining $w^{k+1} = (x^{k+1}, \mu^{k+1})$, we conclude that

$$(2.12) \quad 0 \in G(w^{k+1}) + p^k + \mathcal{N}(w^{k+1}), \text{ where } p^k = \left(\epsilon_k \vartheta, \frac{\mu^{k+1} - \mu^k}{\rho_k} \right), \vartheta \in B.$$

Thus, the behavior of the sequence generated by the augmented Lagrangian method is connected to the study of solutions of GE (2.5) under perturbations of the form in (2.12). There are some ingredients in our approach that are closely related to [18], where an analysis of a certain class of Newtonian methods for GEs is presented. However, the framework of [18] is not applicable to our case. To see this, it is enough to note that the penalty parameters ρ_k in (2.12) are not functions of the problem variables, unlike the perturbation terms of the Newtonian iteration in [18].

Let Σ_* be the solution set of the GE (2.5) (equivalently, of the KKT system (2.2)) and let $\Sigma(p)$ be the solution set of its right-hand side perturbation, i.e.,

$$(2.13) \quad \Sigma(p) = \{w \in \mathbb{R}^{n+m} \mid 0 \in G(w) + p + \mathcal{N}(w)\}, \quad p \in \mathbb{R}^{n+m}.$$

Let $\bar{w} \in \Sigma_*$ be a specific solution of the GE (2.5). The following upper-Lipschitzian property of $\Sigma(p)$ is one of the main ingredients in local analysis of a number of algorithms for solving GEs and its special cases: there exist $\varepsilon_0, \gamma_0, \tau_0 > 0$ such that

$$(2.14) \quad \Sigma(p) \cap (\bar{w} + \varepsilon_0 B) \subseteq \Sigma_* + \tau_0 \|p\| B \quad \forall p \in \gamma_0 B.$$

According to [18, Theorem 2], in the case of KKT systems the property (2.14) is further equivalent to the first inequality in the following error bound:

$$(2.15) \quad \beta_1 \text{dist}(w, \Sigma_*) \leq \sigma(w) \leq \beta_2 \text{dist}(w, \Sigma_*) \quad \forall w \in \bar{w} + \varepsilon_\sigma B,$$

where $\varepsilon_\sigma > 0$, $\beta_2 \geq \beta_1 > 0$ and $\sigma(\cdot)$ is the natural residual of KKT conditions defined in (2.3). The second inequality in (2.15) holds by the Lipschitz-continuity of the natural residual. Furthermore, the property (2.14) is also equivalent to the assumption that the multiplier $\bar{\mu}$ in $\bar{w} = (\bar{x}, \bar{\mu})$ is noncritical as defined in [22]. We shall not introduce the latter notion here, as for the purposes of this paper it is enough to mention that SOSC (1.6) holding at $(\bar{x}, \bar{\mu})$ implies that $\bar{\mu}$ is noncritical and thus (2.15) holds [22]. Summarizing, under SOSC (1.6) we have the properties stated in (2.14) and (2.15). Note also that since SOSC implies that \bar{x} is locally unique, shrinking ε_σ if necessary, it holds that

$$(2.16) \quad \text{dist}(w, \Sigma_*) = (\|x - \bar{x}\|^2 + \text{dist}(\mu, \mathcal{M}(\bar{x}))^2)^{1/2} \quad \forall w \in \bar{w} + \varepsilon_\sigma B.$$

We now sketch the general line of our convergence analysis. Observe that (2.12) means that $w^{k+1} \in \Sigma(p^k)$. Thus, if $w^{k+1} \in \bar{w} + \varepsilon_0 B$ and $p^k \in \gamma_0 B$, from (2.14) we obtain that

$$(2.17) \quad \begin{aligned} \text{dist}(w^{k+1}, \Sigma_*) &\leq \tau_0 \|p^k\| \\ &\leq \tau_0 \left(\epsilon_k + \frac{1}{\rho_k} \|\mu^{k+1} - \mu^k\| \right). \end{aligned}$$

It should be noted that relations like (2.17) are standard in the literature of augmented Lagrangian methods, but previously they have always been derived assuming the linear independence constraint qualification at \bar{x} , among other things. Our interpretation above of the augmented Lagrangian iterates as solutions of perturbed GEs (2.12) shows that (2.17) is a direct consequence of the upper-Lipschitzian property (2.14), for which SOSC (1.6) is sufficient.

Consider, for the moment, the case of the exact iterations with $\epsilon_k = 0$. If the condition (2.8) holds, then

$$\|w^{k+1} - w^k\| \leq \hat{c}\sigma(w^k)$$

and, by (2.15),

$$\|\mu^{k+1} - \mu^k\| \leq \hat{c}\beta_2 \text{dist}(w^k, \Sigma_*).$$

It then follows from (2.17) that

$$\text{dist}(w^{k+1}, \Sigma_*) \leq \frac{\hat{c}\tau_0\beta_2}{\rho_k} \text{dist}(w^k, \Sigma_*).$$

Hence, convergence (and rate of convergence) essentially would follow if we show that the sequence generated by (2.7)–(2.9) is well-defined and, in particular, that (2.8) holds.

3. Local convergence analysis. The first stage of the analysis is to show that under SOSC (1.6) holding at $(\bar{x}, \bar{\mu})$, the augmented Lagrangian has (exact local) minimizers that satisfy the condition (2.8).

We start with establishing that the augmented Lagrangian has the quadratic growth property around \bar{x} , uniform for all $\hat{\mu} \in \mathcal{M}(\bar{x})$ close to $\bar{\mu}$. This is an extension of [29, Theorem 7.4 (c)], which states this property for the fixed $\bar{\mu}$.

PROPOSITION 3.1. *If $(\bar{x}, \bar{\mu})$ satisfies SOSC (1.6), then there exist the constants $\delta_{\bar{\mu}}, \gamma_{\bar{\mu}}, \eta_{\bar{\mu}}, \rho_{\bar{\mu}} > 0$ such that*

$$(3.1) \quad \bar{L}(x, \hat{\mu}; \rho) \geq f(\bar{x}) + \gamma_{\bar{\mu}} \|x - \bar{x}\|^2,$$

for all $x \in \bar{x} + \delta_{\bar{\mu}}B$, all $\hat{\mu} \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap \mathcal{M}(\bar{x})$, and all $\rho \geq \rho_{\bar{\mu}}$.

Proof. For all $\hat{\mu} \in \mathcal{M}(\bar{x})$ and $\rho > 0$, it holds that $\hat{\mu} = \Pi_Q(\hat{\mu} + \rho g(\bar{x}))$. Hence, $\bar{L}(\bar{x}, \hat{\mu}; \rho) = f(\bar{x})$ and $\frac{\partial \bar{L}}{\partial x}(\bar{x}, \hat{\mu}; \rho) = \frac{\partial L}{\partial x}(\bar{x}, \hat{\mu}) = 0$. Then the second-order expansion of the augmented Lagrangian in the primal variables takes the following form (see [29, Prop. 7.2, (7.3)]):

$$(3.2) \quad \bar{L}(\bar{x} + u, \hat{\mu}; \rho) = f(\bar{x}) + \Phi_0(\hat{\mu}, u) + \rho\Phi_1(\hat{\mu}, u) + o(\|u\|^2),$$

where

$$\begin{aligned} \Phi_0(\hat{\mu}, u) &= \frac{1}{2} \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \hat{\mu})u, u \right\rangle, \\ \Phi_1(\hat{\mu}, u) &= \frac{1}{2} \left(\sum_{i \in \mathcal{I}_1(\hat{\mu})} (\langle g'_i(\bar{x}), u \rangle)^2 + \sum_{i \in \mathcal{I}_0(\hat{\mu})} (\max\{0, \langle g'_i(\bar{x}), u \rangle\})^2 \right). \end{aligned}$$

Note that if $i \in \mathcal{I}_1(\bar{\mu})$, $i \geq l+1$, we have that $\bar{\mu}_i > 0$. Then $\mathcal{I}_1(\bar{\mu}) \subseteq \mathcal{I}_1(\hat{\mu})$ for all $\hat{\mu} \in \mathcal{M}(\bar{x})$ close enough to $\bar{\mu}$, by continuity. For the same (continuity) reason, $\Phi_0(\hat{\mu}, u) > 0$ for all $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, since this property holds at $\bar{\mu}$ by SOSC (1.6) and $\mathcal{C}(\bar{x})$ is closed. Thus, there exists $\eta_{\bar{\mu}} > 0$ such that

$$(3.3) \quad \mathcal{I}_1(\bar{\mu}) \subseteq \mathcal{I}_1(\hat{\mu}) \text{ and } \Phi_0(\hat{\mu}, u) > 0 \forall u \in \mathcal{C}(\bar{x}) \setminus \{0\}, \quad \forall \hat{\mu} \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap \mathcal{M}(\bar{x}).$$

Define $\Psi_0, \Psi_1 : \mathbb{R}^n \rightarrow \mathbb{R}$ by

$$\Psi_0(u) = \min\{\Phi_0(\hat{\mu}, u) \mid \hat{\mu} \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap \mathcal{M}(\bar{x})\} \quad \text{and} \quad \Psi_1(u) = \Phi_1(\bar{\mu}, u).$$

By [30, Theorem 1.17 (c)], Ψ_0 is continuous on \mathbb{R}^n . Let $-t_0 \in \mathbb{R}$ and $t_1 \in \mathbb{R}$ be the minimal values of the continuous functions Ψ_0 and Ψ_1 , respectively, on the compact

set $D = \{u \in S \mid \Psi_0(u) \leq 0\}$, where $S = \{u \in \mathbb{R}^n \mid \|u\| = 1\}$. Obviously, $t_0 \geq 0$ by the definition of D . Note that by the compactness of the set in the definition of Ψ_0 and (3.3), it holds that $\Psi_0(u) > 0$ for all $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, and by the definition (1.5) we have that $\mathcal{C}(\bar{x}) = \{u \in \mathbb{R}^n \mid \Psi_1(u) \leq 0\}$. Thus, $\Psi_1(u) > 0$ for all $u \in D$ (if $\Psi_1(u) \leq 0$ for $u \in S$ then $u \in \mathcal{C}(\bar{x}) \setminus \{0\}$, so that $\Psi_0(u) > 0$ and $u \notin D$). Hence, $t_1 > 0$ by the compactness of D . Choose $\rho_{\bar{\mu}} > t_0/t_1 \geq 0$. As $\Psi_1(u) \geq 0$ for all $u \in \mathbb{R}^n$, by the definition of D we have that $\Psi_0(u) + \rho_{\bar{\mu}}\Psi_1(u) \geq \Psi_0(u) > 0$ for all $u \in S \setminus D$. Also, $\Psi_0(u) + \rho_{\bar{\mu}}\Psi_1(u) > 0$ for all $u \in D$ by the definition of $\rho_{\bar{\mu}}$. Thus $\Psi_0(u) + \rho_{\bar{\mu}}\Psi_1(u) > 0$ for all $u \in S$. By the compactness of S , and by the continuity of Ψ_0 and Ψ_1 , it follows that there exists $\tilde{\varepsilon} > 0$ such that

$$(3.4) \quad \Psi_0(u) + \rho_{\bar{\mu}}\Psi_1(u) \geq \tilde{\varepsilon} > 0 \quad \forall u \in S.$$

Since $\mathcal{I}_1(\mu) \cup \mathcal{I}_0(\mu) = \mathcal{I}$ for any $\mu \in \mathcal{M}(\bar{x})$, it holds that

$$(\mathcal{I}_1(\hat{\mu}) \setminus \mathcal{I}_1(\bar{\mu})) \cup \mathcal{I}_0(\hat{\mu}) = \mathcal{I}_0(\bar{\mu}).$$

Hence, for any $\hat{\mu} \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap \mathcal{M}(\bar{x})$ we have

$$\begin{aligned} 2\Phi_1(\hat{\mu}, u) &= \sum_{i \in \mathcal{I}_1(\bar{\mu})} (\langle g'_i(\bar{x}), u \rangle)^2 + \sum_{i \in \mathcal{I}_1(\hat{\mu}) \setminus \mathcal{I}_1(\bar{\mu})} (\langle g'_i(\bar{x}), u \rangle)^2 + \sum_{i \in \mathcal{I}_0(\hat{\mu})} (\max\{0, \langle g'_i(\bar{x}), u \rangle\})^2 \\ &\geq \sum_{i \in \mathcal{I}_1(\bar{\mu})} (\langle g'_i(\bar{x}), u \rangle)^2 + \sum_{i \in \mathcal{I}_0(\bar{\mu})} (\max\{0, \langle g'_i(\bar{x}), u \rangle\})^2 \\ &= 2\Phi_1(\bar{\mu}, u) = 2\Psi_1(u), \end{aligned}$$

where we used the fact that $t^2 \geq \max\{0, t\}^2$. Combining the latter relation with (3.4), we conclude that for all $u \in S$ it holds that

$$\Phi_0(\hat{\mu}, u) + \rho_{\bar{\mu}}\Phi_1(\hat{\mu}, u) \geq \Psi_0(u) + \rho_{\bar{\mu}}\Psi_1(u) \geq \tilde{\varepsilon}.$$

For each $v \in \mathbb{R}^n \setminus \{0\}$ considering $u = v/\|v\| \in S$, using the definitions of Φ_0 and Φ_1 the latter relation gives

$$\Phi_0(\hat{\mu}, v) + \rho_{\bar{\mu}}\Phi_1(\hat{\mu}, v) \geq \tilde{\varepsilon}\|v\|^2,$$

for all $\hat{\mu} \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap \mathcal{M}(\bar{x})$ and all $v \in \mathbb{R}^n$. Using now (3.2), we guarantee the existence of $\gamma_{\bar{\mu}}, \delta_{\bar{\mu}} > 0$ such that

$$\begin{aligned} \bar{L}(\bar{x} + v, \hat{\mu}; \rho_{\bar{\mu}}) &\geq \min\{\bar{L}(\bar{x} + v, \mu; \rho_{\bar{\mu}}) \mid \mu \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap \mathcal{M}(\bar{x})\} \\ &= \bar{L}(\bar{x} + v, \tilde{\mu}; \rho_{\bar{\mu}}) \\ &\geq f(\bar{x}) + \tilde{\varepsilon}\|v\|^2 + o(\|v\|^2) \\ &\geq f(\bar{x}) + \gamma_{\bar{\mu}}\|v\|^2, \end{aligned}$$

for all $v \in \delta_{\bar{\mu}}B$.

The property (3.1) follows taking $x = \bar{x} + v$, $v \in \delta_{\bar{\mu}}B$, and using the fact that $\bar{L}(x, \mu; \cdot)$ is nondecreasing [29, Proposition 7.1]. \blacksquare

The next result shows, among other things, that under the stated assumptions the augmented Lagrangians have local minimizers in the *interior* of some ball around the point \bar{x} . This conclusion is standard under SOSOC for the case of equality-constrained

problems and in the general case if strict complementarity is assumed (e.g., combining SOSC with [4, Prop. 2.2]). Using the quadratic growth condition and the exterior penalty arguments, it can be seen that the property holds also without strict complementarity (see, e.g., [6, Thm. 2]). We include here a different proof for completeness, and also because some specific estimates derived in the process would be needed later on.

COROLLARY 3.2. *Let $(\bar{x}, \bar{\mu})$ satisfy SOSC (1.6), and let $\delta_{\bar{\mu}}, \gamma_{\bar{\mu}}, \eta_{\bar{\mu}}, \rho_{\bar{\mu}} > 0$ be the constants defined in Proposition 3.1.*

If $\mu \in \mathbb{R}^m$, $x \in \bar{x} + \delta_{\bar{\mu}}B$ and $\rho \geq \rho_{\bar{\mu}}$ are such that $\bar{L}(x, \mu; \rho) \leq f(\bar{x})$, then for all $\hat{\mu} \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap \mathcal{M}(\bar{x})$ it holds that

$$(3.5) \quad \|x - \bar{x}\|^2 \leq \frac{1}{\rho\gamma_{\bar{\mu}}} \langle \Pi_Q(\mu + \rho g(x)) - \mu, \hat{\mu} - \mu \rangle.$$

Furthermore, shrinking $\delta_{\bar{\mu}}, \eta_{\bar{\mu}}$ and increasing $\rho_{\bar{\mu}}$ if necessary, it holds that for all $\mu \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap Q$ and all $\rho \geq \rho_{\bar{\mu}}$ the problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \bar{L}(x, \mu; \rho)$$

has a local minimizer $\tilde{x} \in \text{int}(\bar{x} + \delta_{\bar{\mu}}B)$.

Proof. Using the concavity of $\bar{L}(x, \cdot; \rho)$ [29, Prop. 7.1], we have that

$$(3.6) \quad \begin{aligned} \bar{L}(x, \mu; \rho) &\geq \bar{L}(x, \hat{\mu}; \rho) - \left\langle \frac{\partial \bar{L}}{\partial \mu}(x, \mu; \rho), \hat{\mu} - \mu \right\rangle \\ &= \bar{L}(x, \hat{\mu}; \rho) - \frac{1}{\rho} \langle \Pi_Q(\mu + \rho g(x)) - \mu, \hat{\mu} - \mu \rangle \\ &\geq f(\bar{x}) + \gamma_{\bar{\mu}} \|x - \bar{x}\|^2 - \frac{1}{\rho} \langle \Pi_Q(\mu + \rho g(x)) - \mu, \hat{\mu} - \mu \rangle, \end{aligned}$$

where we used (3.1) for the second inequality. Then if $\bar{L}(x, \mu; \rho) \leq f(\bar{x})$, the property (3.5) follows immediately.

We now prove the last assertion. Fix $\tau \in (0, 1)$. Define $\bar{c} > 0$ and $\ell > 0$ such that

$$\bar{c} = \begin{cases} \frac{1}{3} \min_{i \in \mathcal{J}} \{-g_i(\bar{x})\}, & \text{if } \mathcal{J} \neq \emptyset, \\ 1, & \text{otherwise,} \end{cases}$$

and

$$\|g(x) - g(\bar{x})\| \leq \ell \|x - \bar{x}\| \quad \forall x \in \bar{x} + \delta_{\bar{\mu}}B.$$

Decreasing $\delta_{\bar{\mu}}, \eta_{\bar{\mu}}$ and increasing $\rho_{\bar{\mu}}$, if necessary, we can guarantee that

$$(3.7) \quad 0 < \delta_{\bar{\mu}} \leq \frac{\bar{c}}{\ell}, \quad 0 < \eta_{\bar{\mu}} \leq \frac{\gamma_{\bar{\mu}} \tau^2 \delta_{\bar{\mu}}}{2\ell}, \quad \rho_{\bar{\mu}} \geq \frac{\gamma_{\bar{\mu}} \tau^2}{2\ell^2}.$$

Take any $\mu \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap Q$, any $\rho \geq \rho_{\bar{\mu}}$, and let \tilde{x} be any (global) minimizer of $\bar{L}(\cdot, \mu; \rho)$ over the (compact) set $\bar{x} + \delta_{\bar{\mu}}B$.

Since $g(\tilde{x}) \in Q^\circ$ and the projection operator is nonexpansive, it holds that

$$\|\Pi_Q(\mu + \rho g(\tilde{x}))\| = \|\Pi_Q(\mu + \rho g(\tilde{x})) - \Pi_Q(\rho g(\tilde{x}))\| \leq \|\mu\|.$$

Hence,

$$(3.8) \quad \begin{aligned} \bar{L}(\tilde{x}, \mu; \rho) &\leq \bar{L}(\bar{x}, \mu; \rho) \\ &= f(\bar{x}) + \frac{1}{2\rho} \left(\|\Pi_Q(\mu + \rho g(\bar{x}))\|^2 - \|\mu\|^2 \right) \leq f(\bar{x}). \end{aligned}$$

Condition (3.8) now implies that the bound (3.5) holds for \tilde{x} and $\bar{\mu}$, i.e.,

$$(3.9) \quad \|\tilde{x} - \bar{x}\| \leq \left(\frac{1}{\rho\gamma_{\bar{\mu}}} \langle \Pi_Q(\mu + \rho g(\tilde{x})) - \mu, \bar{\mu} - \mu \rangle \right)^{1/2}.$$

To establish the claim, it remains to show that the right-hand side in (3.9) is strictly smaller than $\delta_{\bar{\mu}}$.

If $i \in \mathcal{J}$ then $\bar{\mu}_i = 0$, and we obtain that

$$\begin{aligned} \frac{1}{\rho}\mu_i + g_i(\tilde{x}) &\leq \frac{1}{\rho}|\mu_i - \bar{\mu}_i| + |g_i(\tilde{x}) - g_i(\bar{x})| + g_i(\bar{x}) \\ &\leq \frac{1}{\rho}\eta_{\bar{\mu}} + \ell\delta_{\bar{\mu}} - 3\bar{c} \\ &\leq 2\ell\delta_{\bar{\mu}} - 3\bar{c} \\ &\leq -\bar{c} < 0, \end{aligned}$$

where the second inequality follows from the definition of \bar{c} , and for the other relations we use (3.7). Hence, $(\mu + \rho g(\tilde{x}))_{\mathcal{J}} < 0$ and we have that

$$(3.10) \quad \|(\Pi_Q(\mu + \rho g(\tilde{x})) - \mu)_{\mathcal{J}}\| = \|\mu_{\mathcal{J}}\| = \|\mu_{\mathcal{J}} - \bar{\mu}_{\mathcal{J}}\|.$$

By the structure of Q and using that $g_i(\bar{x}) = 0$ for $i \in \mathcal{I}$, we have

$$\begin{aligned} |(\Pi_Q(\mu + \rho g(\tilde{x})) - \mu)_i| &= \rho|g_i(\tilde{x}) - g_i(\bar{x})| \text{ for } i \in \{1, \dots, l\}, \\ |(\Pi_Q(\mu + \rho g(\tilde{x})) - \mu)_i| &= |\max\{0, \mu_i + \rho g_i(\tilde{x})\} - \mu_i| \\ &\leq \rho|g_i(\tilde{x}) - g_i(\bar{x})| \text{ for } i \in \mathcal{I} \setminus \{1, \dots, l\}, \end{aligned}$$

where for the last inequality we use the fact that $\mu_i \geq 0$ (because $\mu \in Q$) and $|\max\{a, b\} - \max\{a, c\}| \leq |b - c|$ for any $a, b, c \in \mathbb{R}$. Combining the relations above, we obtain

$$(3.11) \quad \|(\Pi_Q(\mu + \rho g(\tilde{x})) - \mu)_{\mathcal{I}}\| \leq \rho\|g_{\mathcal{I}}(\tilde{x}) - g_{\mathcal{I}}(\bar{x})\|.$$

Now, using (3.10) and (3.11), it follows that

$$\begin{aligned} \frac{1}{\rho} \langle \Pi_Q(\mu + \rho g(\tilde{x})) - \mu, \bar{\mu} - \mu \rangle &\leq \frac{1}{\rho} \|(\Pi_Q(\mu + \rho g(\tilde{x})) - \mu)_{\mathcal{J}}\| \|(\bar{\mu} - \mu)_{\mathcal{J}}\| \\ &\quad + \frac{1}{\rho} \|(\Pi_Q(\mu + \rho g(\tilde{x})) - \mu)_{\mathcal{I}}\| \|(\bar{\mu} - \mu)_{\mathcal{I}}\| \\ &\leq \|\mu_{\mathcal{J}}\| \|(\bar{\mu} - \mu)_{\mathcal{J}}\| + \frac{1}{\rho} \|\mu_{\mathcal{J}} - \bar{\mu}_{\mathcal{J}}\|^2 \\ &\leq \ell\delta_{\bar{\mu}}\eta_{\bar{\mu}} + \frac{1}{\rho}\eta_{\bar{\mu}}^2 \\ &\leq \frac{1}{2}\gamma_{\bar{\mu}}\tau^2\delta_{\bar{\mu}}^2 + \frac{1}{2}\gamma_{\bar{\mu}}\tau^2\delta_{\bar{\mu}}^2 = \gamma_{\bar{\mu}}\tau^2\delta_{\bar{\mu}}^2, \end{aligned}$$

where for the last inequality we use (3.7). Combining this bound with with (3.9), we conclude that

$$\|\tilde{x} - \bar{x}\| \leq \tau \delta_{\bar{\mu}} < \delta_{\bar{\mu}},$$

i.e., $\tilde{x} \in \text{int}(\bar{x} + \delta_{\bar{\mu}}B)$. It follows that \tilde{x} is an unconstrained local minimizer of $\bar{L}(\cdot, \mu; \rho)$.

■

It is now clear that the (exact or inexact) stationarity condition (2.7) can always be achieved. The next step is to prove that the localization property (2.8) is satisfied as well. We first verify this fact for the case of exact solutions of subproblems.

PROPOSITION 3.3. *Let $(\bar{x}, \bar{\mu})$ satisfy SOSC (1.6), and let $\delta_{\bar{\mu}}, \gamma_{\bar{\mu}}, \eta_{\bar{\mu}}, \rho_{\bar{\mu}} > 0$ be the constants defined in Corollary 3.2.*

Then there exist $\varepsilon_1, c_1 > 0$ such that for any $\rho \geq \rho_{\bar{\mu}}$ and $(x, \mu) \in ((\bar{x}, \bar{\mu}) + \varepsilon_1 B) \cap (\mathbb{R}^n \times Q)$ with $\sigma(x, \mu) > 0$, if y is a solution of

$$\underset{x \in \bar{x} + \delta_{\bar{\mu}}B}{\text{minimize}} \bar{L}(x, \mu; \rho)$$

and $\nu = \Pi_Q(\mu + \rho g(y))$, then it holds that

$$\left\| \begin{bmatrix} y - x \\ \nu - \mu \end{bmatrix} \right\| \leq c_1 \sigma(x, \mu).$$

Proof. Suppose the contrary, i.e., that there exists a sequence $\{(x^k, \mu^k, \rho_k)\} \subset \mathbb{R}^n \times Q \times (0, \infty)$ such that $(x^k, \mu^k) \rightarrow (\bar{x}, \bar{\mu})$, $\rho_k \geq \rho_{\bar{\mu}}$ and

$$\zeta_k = \|(y^k - x^k, \nu^k - \mu^k)\| > k\sigma_k, \quad \sigma_k = \sigma(x^k, \mu^k) > 0,$$

where

$$y^k \in \text{argmin}\{\bar{L}(x, \mu^k; \rho_k) \mid x \in \bar{x} + \delta_{\bar{\mu}}B\}, \quad \nu^k = \Pi_Q(\mu^k + \rho_k g(y^k)).$$

By the assumption above, we have that

$$(3.12) \quad \frac{\sigma_k}{\zeta_k} \rightarrow 0.$$

By the definition (2.3) of $\sigma(\cdot)$, this means that

$$(3.13) \quad f'(x^k) + (g'(x^k))^\top \mu^k = o(\zeta_k), \quad \mu^k - \Pi_Q(\mu^k + g(x^k)) = o(\zeta_k).$$

Also, defining $\hat{\mu}^k = \Pi_{\mathcal{M}(\bar{x})}(\mu^k)$, we have that

$$(3.14) \quad x^k - \bar{x} = O(\sigma_k), \quad \mu^k - \hat{\mu}^k = O(\sigma_k), \quad x^k - \bar{x} = o(\zeta_k), \quad \mu^k - \hat{\mu}^k = o(\zeta_k),$$

where the first two relations follow from (2.15) and the last two from (3.12).

Taking a subsequence if necessary, we can assume that

$$(3.15) \quad \frac{1}{\zeta_k} \begin{bmatrix} y^k - x^k \\ \nu^k - \mu^k \end{bmatrix} \rightarrow \begin{bmatrix} u \\ v \end{bmatrix} \neq 0.$$

For k large enough we have that $y^k \in \bar{x} + \delta_{\bar{\mu}}B$, $\rho_k \geq \rho_{\bar{\mu}}$, $\hat{\mu}^k \in (\bar{\mu} + \eta_{\bar{\mu}}B) \cap \mathcal{M}(\bar{x})$ and $\bar{L}(y^k, \mu^k; \rho_k) \leq f(\bar{x})$ (as in (3.8)). Thus, by (3.5), it holds that

$$(3.16) \quad \|y^k - \bar{x}\|^2 \leq \frac{1}{\rho_k \gamma_{\bar{\mu}}} \langle \nu^k - \mu^k, \hat{\mu}^k - \mu^k \rangle.$$

Since $\mu^k \in Q$, we have that

$$\|\nu^k - \mu^k\| = \|\Pi_Q(\mu^k + \rho_k g(y^k)) - \Pi_Q(\mu^k)\| \leq \rho_k \|g(y^k)\|.$$

Then (3.16) implies

$$(3.17) \quad \begin{aligned} \limsup_{k \rightarrow \infty} \|y^k - \bar{x}\|^2 &\leq \limsup_{k \rightarrow \infty} \frac{1}{\rho_k \gamma_{\bar{\mu}}} \|\nu^k - \mu^k\| \|\hat{\mu}^k - \mu^k\| \\ &\leq \limsup_{k \rightarrow \infty} \frac{1}{\gamma_{\bar{\mu}}} \|g(y^k)\| \|\hat{\mu}^k - \mu^k\| = 0, \end{aligned}$$

where we used the fact that $\{y^k\}$ is bounded. From (3.16) we also obtain that

$$\limsup_{k \rightarrow \infty} \frac{\|y^k - \bar{x}\|^2}{\zeta_k^2} \leq \limsup_{k \rightarrow \infty} \frac{1}{\rho_k \gamma_{\bar{\mu}}} \frac{\|\nu^k - \mu^k\|}{\zeta_k} \frac{\|\hat{\mu}^k - \mu^k\|}{\zeta_k} = 0,$$

where we have used the last relation in (3.14), that $\rho_k \geq \rho_{\bar{\mu}}$, and that $\zeta_k^{-1}(\nu^k - \mu^k) \rightarrow v$ by (3.15). Using now the relation above and again (3.14), we conclude that

$$(3.18) \quad u = \lim_{k \rightarrow \infty} \frac{y^k - x^k}{\zeta_k} = \lim_{k \rightarrow \infty} \frac{y^k - \bar{x}}{\zeta_k} - \lim_{k \rightarrow \infty} \frac{x^k - \bar{x}}{\zeta_k} = 0.$$

From (3.16), using that $\zeta_k^{-1}(\nu^k - \mu^k) \rightarrow v$, we also obtain

$$(3.19) \quad \limsup_{k \rightarrow \infty} \frac{\rho_k}{\zeta_k} \|y^k - \bar{x}\|^2 \leq \limsup_{k \rightarrow \infty} \frac{1}{\gamma_{\bar{\mu}}} \frac{\|\nu^k - \mu^k\|}{\zeta_k} \|\hat{\mu}^k - \mu^k\| = 0.$$

Since $y^k \rightarrow \bar{x}$ (by (3.17)), we have that $y^k \in \text{int}(\bar{x} + \delta_{\bar{\mu}} B)$ so that it is an unconstrained local minimizer of $\bar{L}(\cdot, \mu^k; \rho_k)$. We then have that

$$\begin{aligned} 0 &= \frac{\partial \bar{L}}{\partial x}(y^k, \mu^k; \rho_k) \\ &= f'(y^k) + (g'(y^k))^\top \nu^k \\ &= f'(y^k) + (g'(y^k))^\top \nu^k - f'(x^k) - (g'(x^k))^\top \mu^k + o(\zeta_k) \\ &= f'(y^k) - f'(x^k) + (g'(y^k) - g'(x^k))^\top \mu^k + (g'(y^k))^\top (\nu^k - \mu^k) + o(\zeta_k) \\ &= (g'(y^k))^\top (\nu^k - \mu^k) + o(\zeta_k), \end{aligned}$$

where the second equality is by (2.10), the third is by (3.13), and the last follows from $y^k - x^k = o(\zeta_k)$ (by (3.18)). Dividing both sides of the latter equation by ζ_k , taking limits and using that $y^k \rightarrow \bar{x}$ (by (3.17)), we conclude that

$$(3.20) \quad v \in \ker (g'(\bar{x}))^\top.$$

If $i \in \mathcal{J}$ we have that $\bar{\mu}_i = 0$ and $g_i(\bar{x}) < 0$. Then, since $y^k \rightarrow \bar{x}$ (by (3.17)), $x^k \rightarrow \bar{x}$ and $\mu^k \rightarrow \bar{\mu}$, it holds that $\mu_i^k + g_i(x^k) < 0$ and $\mu_i^k + \rho_k g_i(y^k) < 0$ for k large enough. Thus, by (3.13) we have that $\mu_{\mathcal{J}}^k = o(\zeta_k)$, and by the definition of ν^k that $\nu_{\mathcal{J}}^k = 0$. Hence,

$$(3.21) \quad v_{\mathcal{J}} = \lim_{k \rightarrow \infty} \frac{\nu_{\mathcal{J}}^k - \mu_{\mathcal{J}}^k}{\zeta_k} = 0.$$

Combining (3.21) with (3.20) we conclude that

$$v_{\mathcal{I}} \in K = \ker (g'_{\mathcal{I}}(\bar{x}))^{\top}.$$

Since $g_{\mathcal{I}}(y^k) = g_{\mathcal{I}}(\bar{x}) + g'_{\mathcal{I}}(\bar{x})(y^k - \bar{x}) + O(\|y^k - \bar{x}\|^2)$ and $g_{\mathcal{I}}(\bar{x}) = 0$, we have that

$$g_{\mathcal{I}}(y^k) + O(\|y^k - \bar{x}\|^2) = g'_{\mathcal{I}}(\bar{x})(y^k - \bar{x}) \in \operatorname{im} g'_{\mathcal{I}}(\bar{x}) = K^{\perp}.$$

Using now that $\Pi_K(\cdot)$ is a linear operator (since K is a subspace), we obtain that

$$\begin{aligned} 0 &= \lim_{k \rightarrow \infty} \frac{\rho_k}{\zeta_k} \Pi_K(g_{\mathcal{I}}(y^k) + O(\|y^k - \bar{x}\|^2)) \\ &= \lim_{k \rightarrow \infty} \left(\frac{\rho_k}{\zeta_k} \Pi_K(g_{\mathcal{I}}(y^k)) + O\left(\frac{\rho_k \|y^k - \bar{x}\|^2}{\zeta_k}\right) \right) \\ (3.22) \quad &= \lim_{k \rightarrow \infty} \frac{\rho_k}{\zeta_k} \Pi_K(g_{\mathcal{I}}(y^k)), \end{aligned}$$

where (3.19) was employed for the last equality.

Since the index set $\{l+1, \dots, m\}$ is finite, passing onto a subsequence if necessary, we can assume that there exists $\mathcal{I}_2 \subseteq \{l+1, \dots, m\}$ such that $\mu_i^k + \rho_k g_i(y^k) < 0$ for $i \in \mathcal{I}_2$ and all k large enough, while $\mu_i^k + \rho_k g_i(y^k) \geq 0$ for $i \in \{l+1, \dots, m\} \setminus \mathcal{I}_2$. Then $\nu_i^k = 0$ for $i \in \mathcal{I}_2$ and $\nu_i^k = \mu_i^k + \rho_k g_i(y^k) \geq 0$ for $i \in \{l+1, \dots, m\} \setminus \mathcal{I}_2$. Define the closed convex cone

$$Q_2 = \{w_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|} \mid w_i \in \mathbb{R}, i \in \mathcal{I} \setminus \mathcal{I}_2; w_i \geq 0, i \in \mathcal{I} \cap \mathcal{I}_2\}.$$

Since for $i \in \mathcal{I}_2$ we have $\nu_i^k = 0$ while $\mu_i^k \geq 0$ (because $\mu^k \in Q$ and $\mathcal{I}_2 \subseteq \{l+1, \dots, m\}$), it holds that $-(\nu_{\mathcal{I}}^k - \mu_{\mathcal{I}}^k) \in Q_2$. Dividing by $\zeta_k > 0$ and taking limits, we obtain that

$$-v_{\mathcal{I}} \in Q_2.$$

We also have that $\mu_i^k + \rho_k g_i(y^k) - \nu_i^k = 0$ for $i \in \{1, \dots, l\} \subset \mathcal{I}$ by the structure of Q , and for $i \in \{l+1, \dots, m\} \setminus \mathcal{I}_2$ by the definition of \mathcal{I}_2 . By the same definition, $\mu_i^k + \rho_k g_i(y^k) - \nu_i^k = \mu_i^k + \rho_k g_i(y^k) < 0$ for $i \in \mathcal{I}_2$. Hence, $\rho_k g_{\mathcal{I}}(y^k) - (\nu_{\mathcal{I}}^k - \mu_{\mathcal{I}}^k) \in Q_2^{\circ}$. Using the linearity of $\Pi_K(\cdot)$, we then obtain that

$$\begin{aligned} \Pi_K(Q_2^{\circ}) &\ni \lim_{k \rightarrow \infty} \Pi_K \left(\frac{1}{\zeta_k} (\rho_k g_{\mathcal{I}}(y^k) - (\nu_{\mathcal{I}}^k - \mu_{\mathcal{I}}^k)) \right) \\ &= \lim_{k \rightarrow \infty} \left(\frac{\rho_k}{\zeta_k} \Pi_K(g_{\mathcal{I}}(y^k)) - \Pi_K \left(\frac{\nu_{\mathcal{I}}^k - \mu_{\mathcal{I}}^k}{\zeta_k} \right) \right) \\ &= -\Pi_K(v_{\mathcal{I}}) = -v_{\mathcal{I}}, \end{aligned}$$

where the second equality follows from (3.22) and the last holds since $v_{\mathcal{I}} \in K$. Hence, $-v_{\mathcal{I}} = \Pi_K(\xi_{\mathcal{I}})$ for some $\xi_{\mathcal{I}} \in Q_2^{\circ}$. Since $-v_{\mathcal{I}} \in Q_2 \cap K$ and K is a subspace, we then obtain that

$$0 \geq \langle -v_{\mathcal{I}}, \xi_{\mathcal{I}} \rangle = \langle -v_{\mathcal{I}}, \Pi_K(\xi_{\mathcal{I}}) \rangle = \|v_{\mathcal{I}}\|^2,$$

so that $v_{\mathcal{I}} = 0$. Combining this with (3.18) and (3.21), we obtain that $\langle u, v \rangle = 0$, in contradiction with (3.15). This completes the proof. \blacksquare

We have therefore established that the conditions in (2.7) and (2.8) hold if x^{k+1} is the exact local minimizer \hat{x} of $\bar{L}(\cdot, \mu^k; \rho_k)$ in $\bar{x} + \delta_{\bar{\mu}}B$, and we define in (2.8) $\hat{c} = c_1$ given by Proposition 3.3.

Now taking $\hat{c} > c_1$ in (2.8) and an appropriately small $\epsilon_k > 0$ in (2.7), by the continuity of all the involved quantities it follows that (2.7) and (2.8) hold for any x^{k+1} close enough to \hat{x} . We are now in position to state our main convergence result, which also makes precise the choice of the truncation parameter ϵ_k . The proof of the first item in Theorem 3.4 below is an adaptation of the corresponding part in [18, Theorem 1].

THEOREM 3.4. *Let $(\bar{x}, \bar{\mu})$ satisfy SOS (1.6), and let $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ be any function $\psi(t) = o(t)$.*

Then there exist $\bar{\epsilon}, \bar{\rho} > 0$ such that if $(x^0, \mu^0) \in ((\bar{x}, \bar{\mu}) + \bar{\epsilon}B) \cap (\mathbb{R}^n \times Q)$ and the sequence $\{(x^k, \mu^k)\}$ is generated according to (2.7)–(2.9) with $\rho_k \geq \bar{\rho}$ for all k and $\epsilon_k = \psi(\sigma(x^k, \mu^k))$, the following assertions hold:

- (i) *The sequence $\{(x^k, \mu^k)\}$ is well-defined and converges to $(\bar{x}, \hat{\mu})$, with some $\hat{\mu} \in \mathcal{M}(\bar{x})$.*
- (ii) *For any $q \in (0, 1)$ there exists $\bar{\rho}_q$ such that if $\rho_k \geq \bar{\rho}_q$ for all k then the convergence rate of $\{(x^k, \mu^k)\}$ to $(\bar{x}, \hat{\mu})$ is Q -linear with quotient q .*
- (iii) *If $\rho_k \rightarrow +\infty$, the convergence rate is Q -superlinear.*
- (iv) *If $\epsilon_k \leq t_1 \sigma(x^k, \mu^k)^s$ and $\rho_k \geq t_2 / \sigma(x^k, \mu^k)^{s-1}$, where $t_1, t_2 > 0$ and $s > 1$, then the Q -order of superlinear convergence is at least s .*

Proof. If $\sigma(x^k, \mu^k) = 0$ for any iteration index k , the point (x^k, μ^k) satisfies the KKT conditions (1.4) and the algorithm should, naturally, stop. We assume that this does not happen and $\sigma(x^k, \mu^k) > 0$ for all k .

Let $\epsilon_0, \gamma_0, \tau_0$ satisfy (2.14), β_2 satisfy (2.15), ϵ_1, c_1 be given by Proposition 3.3, $\rho_{\bar{\mu}}$ be given by Proposition 3.1, and let \hat{c} in (2.8) satisfy $\hat{c} > c_1$.

For any $\psi(t) = o(t)$, there exists $\epsilon_2 > 0$ such that

$$(3.23) \quad \psi(\sigma(x, \mu)) \leq \frac{1}{4\beta_2\tau_0} \sigma(x, \mu) \quad \forall (x, \mu) \in (\bar{x}, \bar{\mu}) + \epsilon_2 B.$$

Define

$$(3.24) \quad \bar{\rho} = \max\{4\hat{c}\beta_2\tau_0, \rho_{\bar{\mu}}\}, \quad \epsilon' = \min\left\{2\tau_0\gamma_0, \frac{\epsilon_0}{\hat{c}\beta_2 + 1}, \epsilon_1, \epsilon_2\right\}, \quad \bar{\epsilon} = \frac{\epsilon'}{2\hat{c}\beta_2 + 1}.$$

To shorten notation, let $\bar{w} = (\bar{x}, \bar{\mu})$, $w^k = (x^k, \mu^k)$ and $\sigma_k = \sigma(x^k, \mu^k)$.

We shall argue by induction. Suppose we have $w^j \in \mathbb{R}^n \times Q$, $j = 0, \dots, k$, satisfying

$$(3.25) \quad \|w^j - \bar{w}\| \leq \epsilon',$$

$$(3.26) \quad \left\| \frac{\partial L}{\partial x}(w^{j+1}) \right\| \leq \epsilon_j, \quad \|w^{j+1} - w^j\| \leq \hat{c}\sigma_j,$$

where we recall (2.10) for the first relation in (3.26). Note that if $w^0 \in (\bar{w} + \bar{\epsilon}B) \cap (\mathbb{R}^n \times Q)$ then (3.25) holds for $j = 0$ since $\bar{\epsilon} \leq \epsilon' \leq \epsilon_1$, and there exists $w \in \mathbb{R}^n \times Q$ such that $\frac{\partial L}{\partial x}(w) = 0$ by Corollary 3.2, and $\|w - w^0\| \leq \hat{c}\sigma_0$ by Proposition 3.3 (by the remarks preceding the statement of Theorem 3.4 appropriate approximate solutions

for the first iteration satisfy the required conditions as well). Hence, w^1 is well-defined, and the conditions (3.25), (3.26) hold for $j = 0$. We next show that if (3.25), (3.26) hold for $j = 0, \dots, k$, then w^{k+2} is well-defined and the conditions in question hold for $j = k + 1$.

By (2.15),

$$(3.27) \quad \sigma_j \leq \beta_2 \|w^j - \bar{w}\|.$$

Thus, using (3.25) and (3.24), we have that

$$\|w^{j+1} - \bar{w}\| \leq \|w^{j+1} - w^j\| + \|w^j - \bar{w}\| \leq (\hat{c}\beta_2 + 1)\varepsilon' \leq \varepsilon_0.$$

Also, for $p^j = (\epsilon_j \vartheta, (\mu^{j+1} - \mu^j)/\rho_j)$ with $\vartheta \in B$ and $\epsilon_j = \psi(\sigma_j)$, we obtain

$$\|p^j\| \leq \epsilon_j + \frac{1}{\rho_j} \|w^{j+1} - w^j\| \leq \left(\frac{1}{4\beta_2\tau_0} + \frac{\hat{c}}{\bar{\rho}} \right) \sigma_j \leq \left(\frac{1}{4\tau_0} + \frac{1}{4\tau_0} \right) \varepsilon' \leq \gamma_0,$$

where we use (3.23), (3.25) and (3.24). Hence, since $w^{j+1} \in \bar{w} + \varepsilon_0 B$, $p^j \in \gamma_0 B$ and (2.12) holds, from (2.14) we obtain (2.17) for $j = 0, \dots, k$. We then have that

$$\text{dist}(w^{j+1}, \Sigma_*) \leq \tau_0 \left(\epsilon_j + \frac{\hat{c}}{\rho_j} \sigma_j \right) \leq \left(\frac{1}{4\beta_2} + \frac{1}{4\beta_2} \right) \sigma_j \leq \frac{1}{2} \text{dist}(w^j, \Sigma_*),$$

where in the second inequality we use (3.23) and the fact that $\rho_j \geq \bar{\rho} \geq 4\hat{c}\beta_2\tau_0$, and in the last inequality we use (2.15). It then follows that

$$(3.28) \quad \text{dist}(w^{j+1}, \Sigma_*) \leq \frac{1}{2^{j+1}} \text{dist}(w^0, \Sigma_*),$$

for $j = 0, \dots, k$. Combining this relation with (3.26) and (2.15), we obtain

$$(3.29) \quad \begin{aligned} \|w^{k+1} - w^0\| &\leq \sum_{i=0}^k \|w^{i+1} - w^i\| \\ &\leq \hat{c} \sum_{i=0}^k \sigma_i \\ &\leq \hat{c}\beta_2 \sum_{i=0}^k \text{dist}(w^i, \Sigma_*) \\ &\leq \hat{c}\beta_2 \sum_{i=0}^k \frac{1}{2^i} \text{dist}(w^0, \Sigma_*) \\ &\leq 2\hat{c}\beta_2 \text{dist}(w^0, \Sigma_*). \end{aligned}$$

To show that (3.25) holds for $j = k + 1$, note that

$$\|w^{k+1} - \bar{w}\| \leq \|w^{k+1} - w^0\| + \|w^0 - \bar{w}\| \leq (2\hat{c}\beta_2 + 1)\|w^0 - \bar{w}\| \leq (2\hat{c}\beta_2 + 1)\bar{\varepsilon} = \varepsilon',$$

where we use that $\bar{w} \in \Sigma_*$ and (3.29) for the second inequality and (3.24) for the last inequality. Now, since $w^{k+1} \in \mathbb{R}^n \times Q$ and $\varepsilon' \leq \varepsilon_1$, by Corollary 3.2 and Proposition 3.3 there exists $w \in \mathbb{R}^n \times Q$ such that $\frac{\partial L}{\partial x}(w) = 0$ and $\|w - w^{k+1}\| \leq \hat{c}\sigma_{k+1}$. Thus, w^{k+2} in (2.7)–(2.9) is well-defined and (3.26) holds for $j = k + 1$.

The argument above shows that the sequence generated according to (2.7)–(2.9) is well-defined. We next prove that it converges. Note that (3.25) implies that $\{w^k\}$ is bounded. By (3.28), $\text{dist}(w^k, \Sigma_*) \rightarrow 0$ and all accumulation points of $\{w^k\}$ belong to Σ_* . Suppose $\{w^k\}$ has two different accumulation points \hat{w} and \tilde{w} . Using the same argument as that leading to (3.29), for any $i, k \in \mathbb{N}$ with $k < i$ it holds that

$$(3.30) \quad \|w^i - w^k\| \leq 2\hat{c}\beta_2 \text{dist}(w^k, \Sigma_*).$$

Choosing the indices k such that the corresponding subsequence of $\{w^k\}$ converges to \hat{w} and the indices i such that the corresponding subsequence of $\{w^k\}$ converges to \tilde{w} , the left-hand side in (3.30) stays bounded away from zero, while the right-hand side tends to zero. Hence, $\{w^k\}$ cannot have different accumulation points, i.e., it converges (to a point $\hat{w} \in \Sigma_*$).

Next, fixing any k and passing onto the limit as $i \rightarrow \infty$ in the relation $\|w^k - \hat{w}\| \leq \|w^i - w^k\| + \|w^i - \hat{w}\|$, using (3.30) we conclude that

$$(3.31) \quad \|w^k - \hat{w}\| \leq 2\hat{c}\beta_2 \text{dist}(w^k, \Sigma_*).$$

Hence, using also (3.26) and (2.17), we have that

$$(3.32) \quad \begin{aligned} \|w^{k+1} - \hat{w}\| &\leq 2\hat{c}\beta_2 \text{dist}(w^{k+1}, \Sigma_*) \\ &\leq 2\hat{c}\beta_2 \tau_0 \left(\epsilon_k + \frac{1}{\rho_k} \|\mu^{k+1} - \mu^k\| \right) \\ &\leq 2\hat{c}\beta_2 \tau_0 \left(\frac{\epsilon_k}{\sigma_k} + \frac{\hat{c}}{\rho_k} \right) \sigma_k \\ &\leq 2\hat{c}\beta_2^2 \tau_0 \left(\frac{\epsilon_k}{\sigma_k} + \frac{\hat{c}}{\rho_k} \right) \text{dist}(w^k, \Sigma_*) \\ &\leq 2\hat{c}\beta_2^2 \tau_0 \left(\frac{\epsilon_k}{\sigma_k} + \frac{\hat{c}}{\rho_k} \right) \|w^k - \hat{w}\|. \end{aligned}$$

For item (ii) of the assertion, let $q \in (0, 1)$ and define $\bar{\rho}_q = 2\hat{c}^2\beta_2^2\tau_0/q$. Since $\epsilon_k = o(\sigma_k)$ and $\rho_k \geq \bar{\rho}_q$, from (3.32) we obtain that

$$\limsup_{k \rightarrow \infty} \frac{\|w^{k+1} - \hat{w}\|}{\|w^k - \hat{w}\|} \leq \limsup_{k \rightarrow \infty} 2\hat{c}\beta_2^2\tau_0 \left(\frac{\epsilon_k}{\sigma_k} + \frac{\hat{c}}{\bar{\rho}_q} \right) = \frac{1}{\bar{\rho}_q} 2\hat{c}^2\beta_2^2\tau_0 = q.$$

For item (iii), since $\epsilon_k = o(\sigma_k)$ and $\rho_k \rightarrow +\infty$, from (3.32) we have that

$$\lim_{k \rightarrow \infty} \frac{\|w^{k+1} - \hat{w}\|}{\|w^k - \hat{w}\|} \leq \lim_{k \rightarrow \infty} 2\hat{c}\beta_2^2\tau_0 \left(\frac{\epsilon_k}{\sigma_k} + \frac{\hat{c}}{\rho_k} \right) = 0.$$

Finally, for item (iv), using that $\epsilon_k \leq t_1\sigma_k^s$ and $\rho_k \geq t_2/\sigma_k^{s-1}$, from (3.32) we obtain

$$\begin{aligned} \|w^{k+1} - \hat{w}\| &\leq 2\hat{c}\beta_2^2\tau_0 (t_1\sigma_k^{s-1} + \hat{c}\sigma_k^{s-1}/t_2) \|w^k - \hat{w}\| \\ &\leq 2\hat{c}\beta_2^{s+1}\tau_0 (t_1 + \hat{c}/t_2) \|w^k - \hat{w}\|^s, \end{aligned}$$

i.e., the order of convergence is at least $s > 1$. ■

Our next result establishes that if minimization in the subproblems is exact, then the primal sequence converges with the same Q -rate as the primal-dual sequence. This

appears to be the first assertion of primal Q -rate of convergence for the augmented Lagrangian methods, as opposed to the usual (weaker) R -rate.

COROLLARY 3.5. *If in Theorem 3.4 $\epsilon_k = 0$ for all k , then the primal sequence $\{x^k\}$ converges to \bar{x} with the same Q -order as that of the convergence of the primal-dual sequence $\{(x^k, \mu^k)\}$ to $(\bar{x}, \hat{\mu})$.*

Proof. Let $\hat{\mu}^k = \Pi_{\mathcal{M}(\bar{x})}(\mu^k)$. If $\epsilon_k = 0$ then it holds that

$$0 = \frac{\partial \bar{L}}{\partial x}(x^k, \mu^{k-1}; \rho_{k-1}) = f'(x^k) + (g'(x^k))^\top \mu^k.$$

Hence,

$$(3.33) \quad \begin{aligned} f'(\bar{x}) + (g'(\bar{x}))^\top \mu^k &= f'(\bar{x}) - f'(x^k) + (g'(\bar{x}) - g'(x^k))^\top \hat{\mu}^k \\ &\quad + (g'(\bar{x}) - g'(x^k))^\top (\mu^k - \hat{\mu}^k). \end{aligned}$$

Since $\{\rho_k\}$ is bounded away from zero and $(x^k, \mu^k) \rightarrow (\bar{x}, \hat{\mu})$ with $\hat{\mu}_i = 0$ and $g_i(\bar{x}) < 0$ for $i \in \mathcal{J}$, it holds that $\mu_i^{k-1}/\rho_{k-1} + g_i(x^k) < 0$ for all $i \in \mathcal{J}$ and k large enough. Then $\mu_{\mathcal{J}}^k = 0$. Using also that $\mu^k \in Q$, by the Hoffman's error bound for linear systems [15, Lemma 3.2.3] there exists $c_0 > 0$ such that

$$\begin{aligned} \|\mu^k - \hat{\mu}^k\| &\leq c_0 \|f'(\bar{x}) + (g'(\bar{x}))^\top \mu^k\| \\ &\leq c_0 \|f'(\bar{x}) - f'(x^k) + (g'(\bar{x}) - g'(x^k))^\top \hat{\mu}^k\| \\ &\quad + c_0 \|(g'(\bar{x}) - g'(x^k))^\top (\mu^k - \hat{\mu}^k)\| \\ &\leq c_1 \|x^k - \bar{x}\| + c_2 \|x^k - \bar{x}\| \|\mu^k - \hat{\mu}^k\|, \end{aligned}$$

where we used (3.33) and the fact that $\{\hat{\mu}^k\}$ is bounded (because $\{\mu^k\}$ is convergent). Since $\|x^k - \bar{x}\| \rightarrow 0$, we have that $c_2 \|x^k - \bar{x}\| \leq 1/2$ for k large enough, so that the last relation above implies

$$\|\mu^k - \hat{\mu}^k\| \leq 2c_1 \|x^k - \bar{x}\|.$$

Then,

$$\begin{aligned} \|w^k - \hat{w}\| &\leq 2\hat{c}\beta_2 \text{dist}(w^k, \Sigma_*) \\ &= 2\hat{c}\beta_2 (\|x^k - \bar{x}\|^2 + \|\mu^k - \hat{\mu}^k\|^2)^{1/2} \\ &\leq 2\hat{c}\beta_2 \sqrt{1 + 4c_1^2} \|x^k - \bar{x}\|, \end{aligned}$$

where the first inequality is by (3.31) and the equality is by (2.16). Hence,

$$(3.34) \quad \begin{aligned} \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} &\leq \frac{\|w^{k+1} - \hat{w}\|}{\|x^k - \bar{x}\|} \\ &= \frac{\|w^k - \hat{w}\|}{\|x^k - \bar{x}\|} \frac{\|w^{k+1} - \hat{w}\|}{\|w^k - \hat{w}\|} \\ &\leq 2\hat{c}\beta_2 \sqrt{1 + 4c_1^2} \frac{\|w^{k+1} - \hat{w}\|}{\|w^k - \hat{w}\|}. \end{aligned}$$

If convergence of $\{w^k\}$ to \hat{w} is superlinear, then (3.34) shows that the rate of convergence of $\{x^k\}$ to \bar{x} is at least as fast.

Concerning the linear rate of convergence, the conclusion follows as in the proof of item (ii) of Theorem 3.4 re-defining $\bar{\rho}_q = 4\hat{c}^3\beta_2^3\sqrt{1 + 4c_1^2}\tau_0/q$, which makes the upper-limit of right-hand side in (3.34) no larger than $q \in (0, 1)$. \blacksquare

4. Some related developments.

4.1. Boundedness of the penalty parameters. In computational implementations, boundedness of the penalty parameters is important to avoid ill-conditioning in the subproblems of minimizing augmented Lagrangians (although there exist some strategies for dealing with ill-conditioning in this context, they seem to be not quite competitive at this time).

Penalty parameter updates are different in various sources. To fix the setting, we consider the proposal of [7], implemented within the latest version of ALGENCAN [34].

For $\alpha \in (0, 1)$ and $r > 1$ fixed, define

$$(4.1) \quad \rho_{k+1} = \begin{cases} \rho_k & \text{if } \sigma(x^{k+1}, \mu^{k+1}) \leq \alpha \sigma(x^k, \mu^k), \\ r\rho_k & \text{otherwise,} \end{cases}$$

where we recall that $\sigma(\cdot)$ stands for the natural residual of the KKT conditions (2.3). If solutions of subproblems are exact, then the iterations yield stationarity of the Lagrangian at every step (recall (2.10)). In that case, the corresponding part in $\sigma(\cdot)$ is zero in each iteration, and thus the test in (4.1) to decide on the increase of the penalty parameter measures the feasibility-complementarity progress. In this sense, this strategy is similar to what is commonly employed in other augmented Lagrangian methods.

Conditions that guarantee boundedness of the sequence of penalty parameters $\{\rho_k\}$ generated by the scheme (4.1) arise naturally from our analysis above. Since

$$g(x^{k+1}) - \frac{1}{\rho_k}(\mu^{k+1} - \mu^k) \in \mathcal{N}_Q(\mu^{k+1}),$$

it holds that

$$\mu^{k+1} = \Pi_Q \left(\mu^{k+1} + g(x^{k+1}) - \frac{1}{\rho_k}(\mu^{k+1} - \mu^k) \right).$$

Now, using the nonexpansivity of the projection, we have

$$\|\mu^{k+1} - \Pi_Q(\mu^{k+1} + g(x^{k+1}))\| \leq \frac{1}{\rho_k} \|\mu^{k+1} - \mu^k\|.$$

Thus, recalling again the definition (2.3) of $\sigma(\cdot)$, for (x^{k+1}, μ^{k+1}) satisfying (2.7)–(2.9) we obtain that

$$(4.2) \quad \sigma(x^{k+1}, \mu^{k+1}) \leq \left(\epsilon_k^2 + \frac{1}{\rho_k^2} \|\mu^{k+1} - \mu^k\|^2 \right)^{1/2} \leq \epsilon_k + \frac{\hat{c}}{\rho_k} \sigma(x^k, \mu^k).$$

Hence, the boundedness of $\{\rho_k\}$ is ensured if $\rho_k > \hat{c}/\alpha$ and $\epsilon_k \leq (\alpha - \hat{c}/\rho_k)\sigma(x^k, \mu^k)$. The desired result then follows from Theorem 3.4.

COROLLARY 4.1. *If in the setting of Theorem 3.4 the sequence of penalty parameters $\{\rho_k\}$ is generated according to the rule (4.1) with $\rho_0 \geq \bar{\rho}$, then $\{\rho_k\}$ is bounded.*

Proof. By Theorem 3.4 we have that $\epsilon_k = o(\sigma(x^k, \mu^k))$ and the sequence $\{(x^k, \mu^k)\}$ converges to $(\bar{x}, \hat{\mu})$ with $\sigma(\bar{x}, \hat{\mu}) = 0$ (in particular, $\sigma(x^k, \mu^k) \rightarrow 0$).

Suppose that $\rho_k \rightarrow +\infty$. Using (4.2) we then obtain

$$\frac{\sigma(x^{k+1}, \mu^{k+1})}{\sigma(x^k, \mu^k)} \leq \frac{\epsilon_k}{\sigma(x^k, \mu^k)} + \frac{\hat{c}}{\rho_k} < \alpha,$$

for all k large enough. But then (4.1) implies that $\rho_{k+1} = \rho_k$ for all k large enough, in contradiction with the assumption $\rho_k \rightarrow +\infty$. \blacksquare

4.2. Relations with the classical results. In the analysis above the initial multiplier approximation must be close enough to a Lagrange multiplier satisfying SOSC (1.6). The classical results (cited in the Introduction), while requiring significantly stronger assumptions, allow the initial approximation to be far from the (unique in that setting) optimal multiplier $\bar{\mu}$ at the price of increasing proportionally the threshold value for the penalty parameters. Keeping in mind the practical requirement of using moderate values of penalty parameters, this is perhaps not an important difference since the penalty parameters would introduce a bound on how far the initial multiplier can be taken. Nevertheless, it would be nice to complete the picture by clarifying whether or not increasing the penalty threshold one can also take far away multipliers assuming SOSC (1.6) only. At this time, this is an open question. Here, we show that a result similar to the classical can be derived from our analysis under some stronger assumptions.

In the context of generalized equations, we replace the upper-Lipschitzian property (2.14) of $\Sigma(p)$ by the stronger Aubin property of Σ at $(0, \bar{w})$: there exist constants $\varepsilon_0, \gamma_0, \tau_0 > 0$ such that

$$(4.3) \quad \Sigma(p) \cap (\bar{w} + \varepsilon_0 B) \subseteq \Sigma(q) + \tau_0 \|p - q\| B \quad \forall p, q \in \gamma_0 B.$$

This condition guarantees that Σ is nonempty and single-valued near the origin. Moreover, it was shown in [14, Theorem 6] that in the context of the optimization problem (1.1) the Aubin property (4.3) for $\bar{w} = (\bar{x}, \bar{\mu})$ is equivalent to the combination of the linear independence constraint qualification (LICQ) at \bar{x} (i.e., the set $\{g'_i(\bar{x}), i \in \mathcal{I}\}$ is linearly independent) with the strong second-order sufficient optimality condition (SSOSC) at $(\bar{x}, \bar{\mu})$, i.e.,

$$(4.4) \quad \left\langle \frac{\partial^2 L}{\partial x^2}(\bar{x}, \bar{\mu}) u, u \right\rangle > 0 \quad \forall u \in \mathcal{C}^+(\bar{x}) \setminus \{0\},$$

where

$$\mathcal{C}^+(\bar{x}) = \{u \in \mathbb{R}^n \mid \langle g'_i(\bar{x}), u \rangle = 0 \text{ for } i \in \mathcal{I}_1(\bar{\mu})\}.$$

Note that under the strict complementarity condition $\bar{\mu}_i > 0$ for all $i \in \mathcal{I}$, used in the classical results, it holds that $\mathcal{C}^+(\bar{x}) = \mathcal{C}(\bar{x})$ and thus SSOSC (4.4) and SOSC (1.6) are the same in that case.

We next show that under the assumptions of LICQ and SSOSC, and without strict complementarity, our analysis can handle initial multiplier estimates far from the optimal $\bar{\mu}$, similar to the classical results.

PROPOSITION 4.2. *Let \bar{x} satisfy LICQ and $(\bar{x}, \bar{\mu})$ satisfy SSOSC (4.4). Then there exist $\hat{\eta}, \hat{\rho}, \hat{\delta}, \hat{\tau} > 0$ such that for each (μ, ρ, ϑ) in the set*

$$D = \left\{ (\mu, \rho, \vartheta) \in \mathbb{R}^{m+1+n} \mid \left(\|\vartheta\|^2 + \frac{1}{\rho^2} \|\mu - \bar{\mu}\|^2 \right)^{1/2} \leq \hat{\eta}, \rho \geq \hat{\rho} \right\},$$

there exists the unique vector $\hat{x} \in \bar{x} + \hat{\delta} B$ satisfying

$$\frac{\partial \bar{L}}{\partial x}(\hat{x}, \mu; \rho) = \vartheta.$$

Moreover, it holds that

$$\left\| \begin{bmatrix} \hat{x} - \bar{x} \\ \hat{\mu} - \bar{\mu} \end{bmatrix} \right\| \leq \hat{\tau} \left(\|\vartheta\|^2 + \frac{1}{\rho^2} \|\mu - \bar{\mu}\|^2 \right)^{1/2},$$

where $\hat{\mu} = \Pi_Q(\mu + \rho g(\hat{x}))$.

Proof. By the assumptions, we have that Σ satisfies the Aubin property (4.3).

Define

$$(4.5) \quad \hat{\tau} = 2\tau_0, \quad \hat{\delta} = \varepsilon_0, \quad \tilde{\gamma} = \min\{\gamma_0, \varepsilon_0/\tau_0\}, \quad \hat{\eta} = \tilde{\gamma}/2, \quad \hat{\rho} = 2\varepsilon_0/\tilde{\gamma},$$

and take any $(\mu, \rho, \vartheta) \in D$. Let $v = (\mu - \bar{\mu})/\rho$ and $p^0 = (-\vartheta, -v)$. Since $\|p^0\| \leq \hat{\eta} \leq \gamma_0$, by (4.3) there exists $(x^1, \mu^1) = w^1 \in \Sigma(p^0)$ such that $\|w^1 - \bar{w}\| \leq \tau_0 \|p^0\| \leq \varepsilon_0$. Similarly, we can define the sequences $\{p^j\}$ and $\{w^j = (x^j, \mu^j)\}$ such that

$$p^j = \left(-\vartheta, \frac{1}{\rho}(\mu^j - \bar{\mu}) - v\right), \quad w^j \in \Sigma(p^{j-1}) \quad \text{and} \quad \|w^j - \bar{w}\| \leq \varepsilon_0,$$

for all $j \geq 1$. This construction can be done because

$$\begin{aligned} \|p^j\|^2 &= \|\vartheta\|^2 + \left\|\frac{1}{\rho}(\mu^j - \bar{\mu}) - v\right\|^2 \\ &\leq 2 \left(\frac{1}{\rho^2} \|\mu^j - \bar{\mu}\|^2 + \|\vartheta\|^2 + \|v\|^2 \right) \\ &\leq 2 \left(\frac{1}{\rho^2} \varepsilon_0^2 + \frac{1}{4} \tilde{\gamma}^2 \right) \leq \tilde{\gamma}^2, \end{aligned}$$

where the last two relations come from (4.5) and the fact that $(\mu, \rho, \vartheta) \in D$. Now, using (4.3), for any $k > j$ we obtain

$$\begin{aligned} \|w^k - w^j\| &\leq \tau_0 \|p^k - p^j\| \\ &= \frac{\tau_0}{\rho} \|\mu^{k-1} - \mu^{j-1}\| \leq \frac{\tau_0}{\rho} \|w^{k-1} - w^{j-1}\| \\ &\leq \left(\frac{\tau_0}{\rho}\right)^{j-1} \|w^{k-j+1} - w^1\| \\ &\leq \left(\frac{\tau_0}{\rho}\right)^{j-1} \tau_0 \|p^{k-j} - p^0\| \\ &= \left(\frac{\tau_0}{\rho}\right)^j \|\mu^{k-j} - \bar{\mu}\| \leq \frac{1}{2^j} \varepsilon_0, \end{aligned}$$

where for the last inequality we use the facts that $\|w^{k-j} - \bar{w}\| \leq \varepsilon_0$ and $\rho \geq \hat{\rho} \geq 2\tau_0$. Thus, since $\{w^j\}$ is a Cauchy sequence, there exists \hat{w} such that $w^j \rightarrow \hat{w} = (\hat{x}, \hat{\mu})$.

Hence, $p^j \rightarrow \hat{p} = \left(-\vartheta, \frac{1}{\rho}(\hat{\mu} - \bar{\mu}) - v\right)$ and $\hat{w} \in \Sigma(\hat{p})$. Then,

$$(4.6) \quad \vartheta = f'(\hat{x}) + (g'(\hat{x}))^\top \hat{\mu},$$

$$(4.7) \quad g(\hat{x}) + v - \frac{1}{\rho}(\hat{\mu} - \bar{\mu}) \in \mathcal{N}_Q(\hat{\mu}).$$

By (4.7) and using that $v = (\mu - \bar{\mu})/\rho$ we have $\mu + \rho g(\hat{x}) - \hat{\mu} \in \mathcal{N}_Q(\hat{\mu})$, which is equivalent to $\hat{\mu} = \Pi_Q(\mu + \rho g(\hat{x}))$. Thus, from (4.6) we obtain

$$\frac{\partial \bar{L}}{\partial x}(\hat{x}, \mu; \rho) = f'(\hat{x}) + (g'(\hat{x}))^\top \hat{\mu} = \vartheta.$$

The uniqueness of \hat{x} comes from the fact that Σ is single-valued.

To prove the last relation, by (4.3) and using that $\bar{w} \in \Sigma(0)$, we obtain

$$\begin{aligned} \|\hat{x} - \bar{x}\|^2 + \|\hat{\mu} - \bar{\mu}\|^2 &= \|\hat{w} - \bar{w}\|^2 \leq \tau_0^2 \|\hat{\rho}\|^2 \\ &\leq 2\tau_0^2 \left(\frac{1}{\rho^2} \|\hat{\mu} - \bar{\mu}\|^2 + \|\vartheta\|^2 + \|v\|^2 \right) \\ &\leq 2\tau_0^2 (\|\vartheta\|^2 + \|v\|^2) + \frac{1}{2} \|\hat{\mu} - \bar{\mu}\|^2, \end{aligned}$$

where we used that $\rho \geq \hat{\rho} \geq 2\tau_0$ for the last inequality. Then,

$$\|\hat{x} - \bar{x}\|^2 + \|\hat{\mu} - \bar{\mu}\|^2 \leq 4\tau_0^2 (\|\vartheta\|^2 + \|v\|^2).$$

Thus, the desired result follows noticing that $\hat{\tau} = 2\tau_0$. \blacksquare

Note that in the case of equality-constrained problems the above reproduces the classical result (see [4, Proposition 2.14]), while in the presence of inequality constraints there is an improvement, as the strict complementarity condition is not needed.

Thus, under the assumptions of LICQ and SSOSC, we can take an initial multiplier μ^0 far from $\bar{\mu}$, at the expense of increasing the penalty parameter. Moreover, our Theorem 3.4 can be used from the next (i.e., second) iteration. To that end, considering $\hat{\eta} \leq \bar{\varepsilon}/\hat{\tau}$ and $\hat{\rho} \geq \bar{\rho}$, if we take $(\mu^0, \rho_0, \vartheta) \in D$ we obtain

$$(x^1, \mu^1) = (\hat{x}, \Pi_Q(\mu^0 + \rho_0 g(\hat{x}))) \in ((\bar{x}, \bar{\mu}) + \bar{\varepsilon}B) \cap (\mathbb{R}^n \times Q).$$

We would like to emphasize that the definition of the set D depends on the unique multiplier $\bar{\mu}$. For the degenerate case, the definition of an appropriate set D allowing initial multiplier estimates far from the multiplier set joint with the corresponding existence result is an open question.

4.3. Extension to the case with upper and lower level constraints. When there are some simple (bound or, more generally, linear) constraints, these are often treated by augmented Lagrangian methods directly [10, 1]. I.e, such constraints are not included into the augmented Lagrangian; instead the partial augmented Lagrangian composed with general constraints is minimized subject to the simple constraints. To that end, consider the problem (1.1) with additional linear constraints:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_i(x) = 0, \quad i = 1, \dots, l, \\ & g_i(x) \leq 0, \quad i = l + 1, \dots, m, \\ & Ax \leq a, \end{aligned}$$

where A is an $q \times n$ matrix and $a \in \mathbb{R}^q$. The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \rightarrow \mathbb{R}$ of this problem is given by $L(x, \mu, \lambda) = f(x) + \langle \mu, g(x) \rangle + \langle \lambda, Ax - a \rangle$.

The partial augmented Lagrangian with respect to general constraints is then defined by (2.6) as before, and the subproblem consists of solving

$$(4.8) \quad \min \bar{L}(x, \mu^k; \rho_k) \quad \text{s.t.} \quad Ax \leq a.$$

The multiplier update formula for μ^{k+1} is the usual (1.3), and the new multipliers λ^{k+1} associated to the linear constraints are those resulting from solving the subproblem (4.8).

Using considerations analogous to (2.10) and (2.11) before, the optimality conditions for (4.8) can be written as

$$0 \in \begin{bmatrix} \frac{\partial L}{\partial x}(x^{k+1}, \mu^{k+1}, \lambda^{k+1}) \\ -g(x^{k+1}) + \frac{1}{\rho_k}(\mu^{k+1} - \mu^k) \\ Ax^{k+1} - a \end{bmatrix} + \mathcal{N}(x^{k+1}, \mu^{k+1}, \lambda^{k+1}),$$

where

$$\mathcal{N}(x, \mu, \lambda) = \{0\} \times \mathcal{N}_Q(\mu) \times \mathcal{N}_{\mathbb{R}_+^q}(\lambda).$$

It is clear that the structure at hand is again a perturbed GE representing the KKT conditions for the problem in consideration, with the same type of perturbations as before for general constraints and the associated multipliers, and with zero perturbations associated to the linear constraints. The preceding analysis easily extends.

5. Concluding remarks. We have established convergence and rate of convergence of the augmented Lagrangian method under the sole assumption that the dual starting point is close enough to a multiplier satisfying the usual second-order sufficient optimality condition. No constraint qualifications and no strict complementarity were assumed. Some possible directions of future research concern extending the analysis to various modifications of the classical algorithm. These include using different penalty parameters for different constraints, non-quadratic penalty terms, etc. Also, extensions of the method of multipliers to variational problems could be considered.

REFERENCES

- [1] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. On augmented Lagrangian methods with general lower-level constraints. *SIAM J. Optim.*, 18:1286–1309, 2007.
- [2] R. Andreani, E.G. Birgin, J.M. Martínez, and M.L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Math. Program.*, 111:5–32, 2008.
- [3] D.P. Bertsekas. Multiplier methods: A survey. *Automatica*, 12:133–145, 1976.
- [4] D.P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Academic Press, New York, 1982.
- [5] E.G. Birgin, R. Castillo, and J.M. Martínez. Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems. *Comp. Optim. Appl.*, 31:31–56, 2005.
- [6] E.G. Birgin, C.A. Floudas, and J.M. Martínez. Global minimization using an augmented Lagrangian method with variable lower-level constraints. *Math. Program.*, 125:139–162, 2010.
- [7] E.G. Birgin, D. Fernández, and J.M. Martínez. On the boundedness of penalty parameters in an augmented Lagrangian method with lower level constraints. *Optim. Methods Soft.*, 2011. DOI: 10.1080/10556788.2011.556634.
- [8] B.T. Boggs and J.W. Tolle. Sequential quadratic programming. *Acta Numerica*, 4:1–51, 1996.
- [9] A.R. Conn, N. Gould, and P.L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM J. Numer. Anal.*, 28:545–572, 1991.
- [10] A.R. Conn, N. Gould, A. Sartenaer, and P.L. Toint. Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. *SIAM J. Optim.*, 6:674–703, 1996.
- [11] A.R. Conn, N.I.M. Gould, and Ph.L. Toint. LANCELOT: a Fortran package for largescale nonlinear optimization (Release A). Number 17 in Springer Series in Computational Mathematics. Springer Verlag, Heidelberg, Berlin, New York, 1992.
- [12] A.R. Conn, N.I.M. Gould, and Ph.L. Toint. *Trust-Region Methods*. SIAM, Philadelphia, 2000.
- [13] L. Contesse-Becker. Extended convergence results for the method of multipliers for non-strictly binding inequality constraints. *J. Optim. Theory Appl.*, 79:273–310, 1993.

- [14] A. L. Dontchev and R. T. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM J. Optim.*, 6:1087–1105, 1996.
- [15] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag, New York, 2003.
- [16] D. Fernández, A.F. Izmailov, and M.V. Solodov. Sharp primal superlinear convergence results for some Newtonian methods for constrained optimization. *SIAM J. Optim.* 20:3312–3334, 2010.
- [17] D. Fernández and M. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. *Math. Program.*, 125:47–73, 2010.
- [18] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Math. Program.*, 94:91–124, 2002.
- [19] M.R. Hestenes. Multiplier and gradient methods. *J. Opt. Theory and Appl.*, 4, 303–320, 1969.
- [20] K. Ito and K. Kunisch. The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces. *Math. Program.*, 46:341–360, 1990.
- [21] 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. *Comp. Optim. Appl.* 46:347–368, 2010.
- [22] A.F. Izmailov and M.V. Solodov. Stabilized SQP revisited. *Math. Program.*, 2010. DOI: 10.1007/s10107-010-0413-3.
- [23] H.-Z. Luo, X.-L. Sun, and D. Li. On the convergence of augmented Lagrangian methods for constrained global optimization. *SIAM J. Optim.*, 18:1209–1230, 2007.
- [24] B.A. Murtagh and M.A. Saunders. A projected Lagrangian algorithm and its implementation for sparse nonlinear constraints. *Math. Progr. Study*, 16:84–117, 1982.
- [25] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer–Verlag, New-York, 1999.
- [26] M.J.D. Powell. A method for nonlinear constraints in minimization problems. In R. Fletcher, ed., *Optimization*, pp. 283–298, Academic Press, London and New York, 1969.
- [27] R.T. Rockafellar. A dual approach for solving nonlinear programming problems by unconstrained optimization. *Math. Progr.*, 5:354–373, 1973.
- [28] R.T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Control Optim.*, 12, 268–285, 1974.
- [29] R.T. Rockafellar. Lagrange multipliers and optimality. *SIAM Rev.*, 35:183–238, 1993.
- [30] R.T. Rockafellar and J.-B. Wets. *Variational Analysis*. Springer-Verlag, New York, 1997.
- [31] A. Ruszczyński. *Nonlinear Optimization*. Princeton University Press, 2006.
- [32] T. Pennanen. Local convergence of the proximal point algorithm and multiplier methods without monotonicity. *Math. Oper. Res.*, 27:170–191, 2002.
- [33] B.T. Polyak and N.V. Tret’yakov. The method of penalty estimates for conditional extremum problems. *USSR Computational Mathematics and Mathematical Physics*, 13:42–58, 1974.
- [34] <http://www.ime.usp.br/~egbirgin/tango/>