Improved Local Convergence Results for Augmented Lagrangian Methods in C^2-Cone Reducible Constrained Optimization*

Christian Kanzow† Daniel Steck†

November 3, 2017

Abstract. This paper deals with a class of cone-reducible constrained optimization problems which encompasses nonlinear programming, semidefinite programming, second-order cone programming, and any combination thereof. Using the second-order sufficient condition and a strict version of the Robinson constraint qualification, we provide a (semi-)local error bound which generalizes known results from the literature. Moreover, under the same assumptions, we prove that an augmented Lagrangian method is locally convergent with rate proportional to $1/\rho_k$, where ρ_k is the penalty parameter, and that $\{\rho_k\}$ remains bounded.

Keywords. Augmented Lagrangian method, local convergence, rate of convergence, C^2-cone reducible sets, error bound, semidefinite programming, second-order cone programming.

1 Introduction

Let X, H be finite-dimensional Euclidean spaces and consider the constrained minimization problem

$$\min \ f(x) \quad \text{s.t.} \quad G(x) \in K, \tag{P}$$

where $f : X \to \mathbb{R}$, $G : X \to H$, and $K \subseteq H$ is a nonempty closed convex set. For the sake of simplicity, we assume that f and G are twice continuously differentiable.

In the present paper, we analyze problems of the type (P) in the situation where the set K is so-called C^2-cone reducible (see Definition 2.1). This encompasses nonlinear programming, semidefinite programming, second-order cone programming, and any combination thereof [26]. In these situations, we typically have $X = \mathbb{R}^n$, $n \in \mathbb{N}$, and the space H is a Cartesian product including factors of the form $\mathbb{R}^{m_j}$, $m_j \in \mathbb{N}$, together with a suitable matrix space in the case of semidefinite programming. More details will be given in Section 5.

The main purpose of this paper is to provide some new results on the local convergence of augmented Lagrangian-type methods for C^2-cone reducible problems. Recall that the

*This research was supported by the German Research Foundation (DFG) within the priority program "Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization" (SPP 1962) under grant number KA 1296/24-1.

†University of Würzburg, Institute of Mathematics, Campus Hubland Nord, Emil-Fischer-Str. 30, 97074 Würzburg, Germany; {kanzow,daniel.steck}@mathematik.uni-wuerzburg.de

augmented Lagrangian method (ALM) is one of the classical approaches for constrained optimization [3, 4, 9, 22]. As a result, there is a substantial amount of literature dealing with local convergence aspects of such methods, in particular for standard nonlinear programming. The classical results in this direction assume the second-order sufficient condition (SOSC) together with the linear independence constraint qualification and yield primal R-linear and dual Q-linear convergence [3, 10, 11, 14, 22]. Note that many of these papers use additional assumptions such as a stronger SOSC or strict complementarity. The sharpest local convergence result for nonlinear programming which we are aware of is given in [13], where it is shown that the ALM converges locally with primal-dual rate proportional to $1/\rho_k$ (ρ_k being the penalty parameter) under the assumptions that the initial multiplier is sufficiently close to an optimal one and that SOSC holds.

For more sophisticated problem classes such as semidefinite or second-order cone programming, the analysis of ALMs is much more complicated. In [27], it was shown that a version of the ALM for semidefinite programming converges locally with primal-dual rate proportional to $1/\rho_k$ under the so-called *strong SOSC* and *constraint nondegeneracy* assumptions. For second-order cone programming, similar results were obtained in [19, 20] using again the constraint nondegeneracy assumption together with either strict complementarity or strong SOSC. Let us also stress that all these papers require that the initial multiplier is sufficiently close to an optimal one.

In the present paper, we analyze the local convergence of the augmented Lagrangian method for generic C^2-cone reducible problems, thereby subsuming semidefinite and second-order cone programming. Using the recently developed sensitivity theory from [12], we show that, under SOSC and a strict version of the Robinson constraint qualification (see Section 2), the KKT system of the problem (P) admits a primal-dual error bound which does not require any (a-priori) proximity of the multipliers. We then use this tool to prove that the ALM converges with primal-dual rate of convergence proportional to $1/\rho_k$. Our analysis is more general than that in [19, 20, 27] since our assumptions are weaker than strong SOSC and constraint nondegeneracy. Moreover, we do not require any assumption on the initial multiplier used in the algorithm.

Let us mention that the analysis in this paper is essentially a special instance of that conducted by the authors in [15] for variational problems in Banach spaces. In contrast to that paper, however, the present one is aimed specifically at C^2-cone reducible problems and takes into account the specific structure of such problems.

This paper is organized as follows: we start with some preliminary material in Section 2, where we also introduce the key assumptions for our analysis. The augmented Lagrangian method is formally presented in Section 3, and we provide a local convergence analysis in Section 4. The applications to semidefinite and second-order cone programming are discussed in Section 5. Finally, Section 6 contains some concluding remarks.

Notation: We write $\langle \cdot, \cdot \rangle$ for the scalar product in X or H, and $\|\cdot\|$ with an appropriate subscript (e.g. $\|\cdot\|_X$) to denote norms. The linear hull of a vector x in X or H is denoted by $\operatorname{span}(x)$, and its orthogonal complement by $x^\perp$. If $K \subseteq H$ is a nonempty closed convex set, then P_K denotes the projection onto K, and $d_K = \operatorname{dist}(\cdot, K)$ the distance function to K. Finally, a prime $'$ always indicates the derivative of a function with respect to x.

2 Preliminaries

For a set S in some space Z and a point $x \in S$, we denote by

$$\mathcal{T}_S(x) := \left\{ d \in Z \mid \exists\, x^k \to x,\ t_k \downarrow 0 \text{ such that } x^k \in S \text{ and } (x^k - x)/t_k \to d \right\}$$

the *tangent cone* to S at x. If $S \subseteq H$ is convex, we also define the *normal cone*

$$\mathcal{N}_S(x) := \{\psi \in H : \langle \psi, y - x\rangle \le 0\ \forall y \in S\} = (S - x)^\circ, \tag{1}$$

where $^\circ$ denotes the polar cone [2, 8]. Note that we also have the characterization $\mathcal{N}_S(x) = \mathcal{T}_S(x)^\circ$. Finally, we say that a cone C is pointed if $C \cap (-C) = \{0\}$.

Throughout this paper, we will use the following important concept [8, 12].

Definition 2.1. We say that K is *C^2-cone reducible at $y_0 \in K$* if there exist a pointed closed convex cone $C \subseteq Z$ in some finite-dimensional space Z, a neighborhood N of y_0, and a twice continuously differentiable mapping $\Xi : N \to Z$ such that $\Xi(y_0) = 0$, $\Xi'(y_0)$ is onto, and $K \cap N = \Xi^{-1}(C) \cap N$. We say that K is *C^2-cone reducible* if the above holds at every $y_0 \in K$.

The C^2-cone reducibility of the set K plays a crucial role for both first and second order optimality conditions as we shall see below. As noted in the introduction, there are quite a few classes of constraints for which C^2-reducibility is known. In particular, the set K is C^2-cone reducible if it arises from an arbitrary (finite) combination of equality, inequality, semidefiniteness, or second-order cone constraints [26].

2.1 Constraint Qualifications and KKT Conditions

Consider the Lagrange function

$$\mathcal{L} : X \times H \to \mathbb{R}, \quad \mathcal{L}(x, \lambda) := f(x) + \langle \lambda, G(x)\rangle.$$

Then the KKT conditions can be stated as follows.

Definition 2.2. A tuple $(\bar{x}, \bar{\lambda}) \in X \times H$ is a *KKT point* of (P) if

$$\mathcal{L}'(\bar{x}, \bar{\lambda}) = 0 \quad \text{and} \quad \bar{\lambda} \in \mathcal{N}_K(G(\bar{x})).$$

We call $\bar{x}$ a *stationary point* if $(\bar{x}, \bar{\lambda})$ is a KKT point for some multiplier $\bar{\lambda}$, and denote by $\mathcal{M}(\bar{x})$ the set of such multipliers.

The relation between the optimization problem (P) and its KKT conditions is well-known [8]: if $\bar{x}$ is a local solution of the problem and a suitable constraint qualification (see below) holds, then $\bar{x}$ is a KKT point. Conversely, the KKT conditions always imply the "abstract" first order stationarity $f'(\bar{x})d \ge 0$ for all $d \in \mathcal{T}_M(\bar{x})$, where $M := G^{-1}(K)$ is the feasible set of (P). In particular, if the problem is convex, then the KKT conditions are always sufficient for local and global optimality.

Definition 2.3. Let $\bar{x} \in X$ be a feasible point. We say that the *Robinson constraint qualification (RCQ)* holds in $\bar{x}$ if

$$G'(\bar{x})X + \mathcal{T}_K(G(\bar{x})) = H. \tag{2}$$

If $(\bar{x}, \bar{\lambda}) \in X \times H$ is a KKT point of the problem, we say that the *strict Robinson condition (SRC)* holds in $(\bar{x}, \bar{\lambda})$ if

$$G'(\bar{x})X + \mathcal{T}_K(G(\bar{x})) \cap \bar{\lambda}^{\perp} = H. \tag{3}$$

The Robinson constraint qualification was introduced in [24] and has since become a ubiquitous tool in optimization theory [8, 21]. For standard nonlinear programming, RCQ boils down to the Mangasarian-Fromovitz constraint qualification, see [8, Eq. 2.191].

The strict Robinson condition is a tightened version of the Robinson constraint qualification aimed at ensuring the uniqueness of the Lagrange multiplier as well as certain stability properties. This condition is also called the "strict Robinson constraint qualification" in the literature [12, 29]. It should be noted, however, that SRC is not a constraint qualification in the conventional sense since it presupposes the existence of $\bar{\lambda}$ and therefore depends not only on the constraint function but on the problem as a whole. A similar observation was made in [28] for a strict version of the Mangasarian-Fromovitz constraint qualification which can be seen as a special case of SRC.

Remark 2.4. A related condition used in the book [8] is the so-called *strict constraint qualification*, which is defined as $0 \in \operatorname{int}[G(\bar{x}) + G'(\bar{x})X - K_0]$, where $K_0 := \{y \in K : \langle \bar{\lambda}, y - G(\bar{x}) \rangle = 0\}$, see [8, Def. 4.46]. By [8, Cor. 2.98], this condition is equivalent to

$$G'(\bar{x})X + \mathcal{T}_{K_0}(G(\bar{x})) = H.$$

Hence, there are some similarities between the strict constraint qualification and our notion of SRC. In fact, the former implies the latter [8, p. 299], but the two conditions are not equivalent in general, even for C^2-cone reducible sets K. Two concrete examples demonstrating this fact are given in Example 2.5 below.

Example 2.5. (a) Let $X := \mathbb{R}$, $H := \mathbb{R}^2$, and consider the optimization problem (P) with $f(x) := x$, $G(x) := (x, 0)$, and K the closed unit ball in H. Clearly, $\bar{x} := -1$ is the global minimizer of this problem, and it is easy to see that $\bar{\lambda} := (-1, 0)$ is the corresponding (unique) Lagrange multiplier. Moreover, the set K is C^2-cone reducible in $\bar{y} := G(\bar{x}) = (-1, 0)$ to the cone $C := [0, +\infty)$ by means of the mapping $\Xi(x) := 1 - x_1^2 - x_2^2$. A straightforward calculation shows that $\mathcal{T}_K(\bar{y}) \cap \bar{\lambda}^{\perp} = \bar{\lambda}^{\perp}$; on the other hand, the set K_0 is given by $K_0 = \{\bar{y}\}$, and it follows that $\mathcal{T}_{K_0}(\bar{y}) = \{0\}$, cf. Figure 1. We conclude that $G'(\bar{x})X + \mathcal{T}_K(\bar{y}) \cap \bar{\lambda}^{\perp} = H$ and $G'(\bar{x})X + \mathcal{T}_{K_0}(\bar{y}) \neq H$.

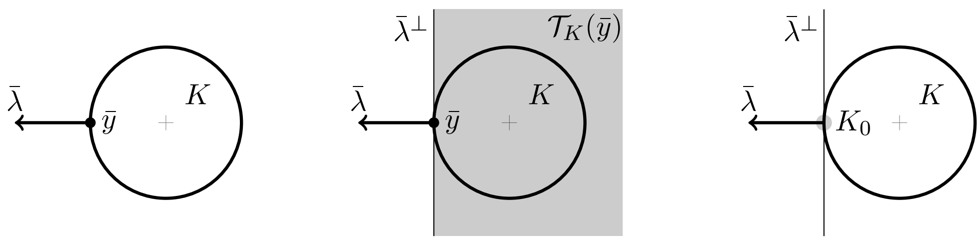

Figure 1: The setting of Example 2.5 (a), the tangent cone to K, and the set K_0.

(b) This example is a second-order cone program. Let $X := \mathbb{R}$, $H := \mathbb{R}^3$, $f(x) := -2x$, $G(x) := (x, 0, 2-x)$, and $K := \left\{ y \in \mathbb{R}^3 : y_3 \geq \sqrt{y_1^2 + y_2^2} \right\}$. An easy calculation shows that $\bar{x} := 1$ is the global minimizer of the problem, and $\bar{\lambda} := (1, 0, -1)$ is the corresponding (unique) Lagrange multiplier. Moreover, with $\bar{y} := G(\bar{x}) = (1, 0, 1)$, we have

$$\mathcal{N}_K(\bar{y}) = \{\alpha\bar{\lambda} : \alpha \geq 0\} \quad \text{and} \quad \mathcal{T}_K(\bar{y}) = \mathcal{N}_K(\bar{y})^\circ = \{\bar{\lambda}\}^\circ.$$

Hence, $\mathcal{T}_K(\bar{y}) \cap \bar{\lambda}^\perp = \bar{\lambda}^\perp$. On the other hand, the set K_0 is the intersection of K with the plane $\bar{\lambda}^\perp$, which is given by $K_0 = \{\alpha\bar{y} : \alpha \geq 0\}$. Therefore, $\mathcal{T}_{K_0}(\bar{y}) = \operatorname{span}(\bar{y})$, and it follows that $G'(\bar{x})X + \mathcal{T}_K(\bar{y}) \cap \bar{\lambda}^\perp = H$ but $G'(\bar{x})X + \mathcal{T}_{K_0}(\bar{y}) \neq H$.

2.2 Second-Order Conditions and Local Error Bounds

We have already mentioned that the KKT conditions from Definition 2.2 are sufficient for optimality if the underlying problem is convex. In the absence of convexity, local optimality can still be deduced if a suitable second-order sufficient condition is satisfied. For the theory in this paper, we will need a second-order condition which crucially depends on the C^2-cone reducibility of the set K. For the corresponding definition, let

$$\sigma(y, C) := \sup\{\langle y, z\rangle : z \in C\}$$

be the *support function* of a closed convex set (or cone) C. Moreover, for a point $y \in K$ and a direction $h \in H$, let

$$\mathcal{T}_K^2(y, h) := \left\{ w \in H : \operatorname{dist}(y + th + \tfrac{1}{2}t^2 w, K) = o(t^2),\ t \geq 0 \right\}$$

be the *second-order tangent set* to K in the direction h, where dist is the distance function. Note that there are actually different notions of second-order tangent sets in the literature, in particular inner and outer tangent sets [8], but these coincide for C^2-cone reducible sets K [8, Prop. 3.136]. Hence, this distinction is not necessary for our purposes.

Definition 2.6. We say that the *second-order sufficient condition (SOSC)* holds in a KKT point $\bar{x}$ if the set K is C^2-cone reducible at $G(\bar{x})$ and

$$\sup_{\lambda \in \mathcal{M}(\bar{x})} \left\{ \mathcal{L}''(\bar{x}, \lambda)(d, d) - \sigma\big(\lambda, \mathcal{T}_K^2(G(\bar{x}), G'(\bar{x})d)\big) \right\} > 0 \tag{4}$$

for all $d \in C(\bar{x}) \setminus \{0\}$, where $C(\bar{x}) := \{d \in X : G'(\bar{x})d \in \mathcal{T}_K(G(\bar{x})),\ f'(\bar{x})d \leq 0\}$.

It is well-known that this form of SOSC implies the quadratic growth of the objective function, i.e. there is a $c > 0$ such that $f(x) \geq f(\bar{x}) + c\|x - \bar{x}\|_X^2$ for all feasible points x near $\bar{x}$, see [8, Thm. 3.86]. In particular, $\bar{x}$ is a strict local minimizer of the problem.

We now turn to another consequence of SOSC when used in conjunction with the strict Robinson condition (SRC). These two conditions imply a primal-dual error bound in a neighborhood of a KKT point $(\bar{x}, \bar{\lambda})$ in terms of the residual mapping

$$R(x, \lambda) := \|\mathcal{L}'(x, \lambda)\|_X + \|G(x) - P_K(G(x) + \lambda)\|_H,$$

see [12]. A precise statement of the error bound is found in the following theorem which actually generalizes [12, Remark 5] by not requiring any a-priori proximity of λ to $\bar{\lambda}$.

Theorem 2.7. *Assume that the problem* (P) *admits a KKT point* $(\bar{x}, \bar{\lambda})$ *which satisfies SOSC and SRC. Then* $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$ *and there is a* $c > 0$ *such that, for all* $(x, \lambda) \in X \times H$ *with* x *sufficiently close to* $\bar{x}$ *and* $R(x, \lambda)$ *sufficiently small,*

$$\|x - \bar{x}\|_X + \|\lambda - \bar{\lambda}\|_H \leq cR(x, \lambda). \tag{5}$$

Proof. The uniqueness follows from [8, Prop. 4.50]. By [15, Thm. 3.1], the above error bound is equivalent to the following upper Lipschitz stability of the KKT system: there is a $c > 0$ such that, for every $(\alpha, \beta) \in X \times H$ sufficiently small, any solution (x, λ) with x near $\bar{x}$ of the perturbed KKT conditions

$$\mathcal{L}'(x, \lambda) = \alpha, \quad \lambda \in \mathcal{N}_K(G(x) - \beta) \tag{6}$$

satisfies the estimate $\|x - \bar{x}\|_X + \|\lambda - \bar{\lambda}\|_H \leq c\|(\alpha, \beta)\|_{X \times H}$. To prove this property, let $\{(\alpha^k, \beta^k)\} \subseteq X \times H$ be a zero sequence and (x^k, λ^k) a sequence of corresponding solutions of (6) such that $x^k \to \bar{x}$. By [8, Prop. 4.43], the sequence $\{\lambda^k\}$ is bounded. Moreover, it follows from (6) and simple continuity arguments that every accumulation point of $\{\lambda^k\}$ is a multiplier corresponding to $\bar{x}$. Since $\mathcal{M}(\bar{x}) = \{\bar{\lambda}\}$, this implies $\lambda^k \to \bar{\lambda}$. The result now follows from [12, Thm. 24]. □

The (semi-local) error bound property is central to this paper and allows us to deduce local convergence properties for certain algorithms, in particular augmented Lagrangian methods. As in [15], we note that the function R is locally Lipschitz-continuous with respect to (x, λ) and globally so with respect to λ. Hence, we can extend the one-sided error bound (5) to

$$c_1 R(x, \lambda) \leq \|x - \bar{x}\|_X + \|\lambda - \bar{\lambda}\|_H \leq c_2 R(x, \lambda) \tag{7}$$

for suitable constants $c_1, c_2 > 0$ and all $(x, \lambda) \in X \times H$ with x near $\bar{x}$.

3 The Augmented Lagrangian Method

We now present the augmented Lagrangian method for the optimization problem (P), which is fundamentally similar to [15, Alg. 4.1]. Consider the augmented Lagrange function

$$\mathcal{L}_\rho : X \times H \to \mathbb{R}, \quad \mathcal{L}_\rho(x, \lambda) := f(x) + \frac{\rho}{2} d_K^2\left(G(x) + \frac{\lambda}{\rho}\right). \tag{8}$$

Note that there are multiple variants of $\mathcal{L}_\rho$ in the literature. In particular, when deriving the augmented Lagrangian through the standard slack variable approach, one formally obtains a function which includes an additional term depending on the multiplier λ. Since this term plays no role for the minimization of $\mathcal{L}_\rho$ with respect to x, we will deal with the "reduced" augmented Lagrangian (8) for the sake of simplicity.

For the construction of our algorithm, we will need a means of controlling the penalty parameters. To this end, we define the auxiliary function

$$V(x, \lambda, \rho) := \left\| G(x) - P_K\left(G(x) + \frac{\lambda}{\rho}\right) \right\|_H. \tag{9}$$

The function V is a composite measure of feasibility and complementarity which arises naturally in the aforementioned slack variable approach.

Algorithm 3.1 (Augmented Lagrangian method)**.**

(S.0) *Let* $(x^0, \lambda^0) \in X \times H$*,* $B \subseteq H$ *bounded,* $\rho_0 > 0$*,* $\gamma > 1$*,* $\tau \in (0, 1)$*, and set* $k := 0$*.*

(S.1) *If* (x^k, λ^k) *satisfies a suitable termination criterion: STOP.*

(S.2) *Choose* $w^k \in B$ *and compute a minimizer* x^{k+1} *of* $\mathcal{L}_{\rho_k}(\cdot, w^k)$*.*

(S.3) *Update the vector of multipliers to*

$$\lambda^{k+1} := \rho_k \left[G(x^{k+1}) + \frac{w^k}{\rho_k} - P_K\left(G(x^{k+1}) + \frac{w^k}{\rho_k} \right) \right]. \tag{10}$$

(S.4) *If* $k = 0$ *or*

$$V(x^{k+1}, w^k, \rho_k) \leq \tau V(x^k, w^{k-1}, \rho_{k-1}) \tag{11}$$

holds, set $\rho_{k+1} := \rho_k$*; otherwise, set* $\rho_{k+1} := \gamma \rho_k$*.*

(S.5) *Set* $k \leftarrow k + 1$ *and go to* (S.1)*.*

Note that Algorithm 3.1 uses a safeguarded multiplier sequence $\{w^k\}$ in certain places where classical augmented Lagrangian methods use the sequence $\{\lambda^k\}$. This bounding scheme is in the spirit of recent developments [6, 15, 23] and is crucial to establishing strong global convergence results for the above and similar methods [1, 5, 6, 16]. In practice, one usually tries to keep w^k as "close" as possible to λ^k, e.g. by defining $w^k := P_B(\lambda^k)$, where B (the bounded set from the algorithm) is chosen suitably to allow cheap projections.

4 Local Convergence Analysis

This section is dedicated to the local convergence analysis of Algorithm 3.1. The first step in our local analysis is to analyze the behavior of local minimizers of the augmented Lagrangian in a neighborhood of $\bar{x}$.

Lemma 4.1. *Let* $(\bar{x}, \bar{\lambda})$ *be a KKT point satisfying SOSC and* $B \subseteq H$ *a bounded set. Then there are* $\bar{\rho}, r > 0$ *such that, for every* $\rho \geq \bar{\rho}$ *and* $w \in B$*, the function* $\mathcal{L}_\rho(x, w)$ *has a local minimizer* $x = x_\rho(w)$ *in* $B_r(\bar{x})$*. Moreover,* $x_\rho \to \bar{x}$ *uniformly on* B *as* $\rho \to \infty$*.*

Proof. Since SOSC holds, there is an $r > 0$ such that $\bar{x}$ is a strict local solution of the problem. For each $\rho > 0$ and $w \in B$, let $x = x_\rho(w)$ be a solution of

$$\min_x \; \mathcal{L}_\rho(x, w) \quad \text{s.t.} \quad x \in B_r(\bar{x}).$$

(Note that x_ρ is not necessarily unique.) We show that $x_\rho \to \bar{x}$ uniformly on B as $\rho \to \infty$. In particular, this implies the existence of a $\bar{\rho} > 0$ such that $x_\rho(w)$ lies in the interior of $B_r(\bar{x})$ for all $\rho \geq \bar{\rho}$, $w \in B$. Assume, by contradiction, that the uniform convergence does not hold. Then there are $\varepsilon > 0$, $\rho_k \to \infty$ and $\{w^k\} \subseteq B$ such that $\|x_{\rho_k}(w^k) - \bar{x}\|_X \geq \varepsilon$ for all k. Due to compactness, the sequence $y^k := x_{\rho_k}(w^k)$ has an accumulation point $\hat{x} \in B_r(\bar{x})$. We claim that $\hat{x} = \bar{x}$, which yields the desired contradiction.

The nonexpansiveness of the distance function d_K implies $d_K(G(\bar{x}) + w^k/\rho_k) \le \|w^k\|_H/\rho_k$ for all k. Thus, by the minimizing property of y^k, we have

$$f(y^k) + \frac{\rho_k}{2} d_K^2\left(G(y^k) + \frac{w^k}{\rho_k}\right) \le \mathcal{L}_{\rho_k}(\bar{x}, w^k) \le f(\bar{x}) + \frac{\|w^k\|_H^2}{2\rho_k}. \tag{12}$$

This implies $d_K(G(y^k)+w^k/\rho_k) \to 0$, hence $d_K(G(y^k)) \to 0$ and $G(\hat{x}) \in K$. Moreover, (12) also yields $\limsup_{k\to\infty} f(y^k) \le f(\bar{x})$. Therefore, $f(\hat{x}) \le f(\bar{x})$, which implies $\hat{x} = \bar{x}$. □

Of course, since SOSC is a local condition, we cannot rule out the possibility that the augmented Lagrangian has local or global minimizers arbitrarily far from $\bar{x}$ or is even unbounded from below. Since we are mainly interested in a local analysis of the method, we will assume that the algorithm finds, at least for sufficiently large k, the minimizers from Lemma 4.1. This enables us to prove the following result which implies the strong convergence of $\{(x^k, \lambda^k)\}$ as well as an estimate for the rate of convergence. Note that this result is essentially a consequence of the theory established in [15]. For the convenience of the reader, we include a complete proof here.

Theorem 4.2. *Let $(\bar{x}, \bar{\lambda})$ be a KKT point satisfying SOSC and SRC and assume that, for k sufficiently large, x^{k+1} is one of the minimizers from Lemma 4.1. Then there is $\bar{\rho} > 0$ such that, if $\rho_k \ge \bar{\rho}$ for sufficiently large k, then $(x^k, \lambda^k) \to (\bar{x}, \bar{\lambda})$. If, in addition, $w^k = \lambda^k$ for sufficiently large k, then there exists $c > 0$ such that*

$$\|x^{k+1} - \bar{x}\|_X + \|\lambda^{k+1} - \bar{\lambda}\|_H \le \frac{c}{\rho_k}\big(\|x^k - \bar{x}\|_X + \|\lambda^k - \bar{\lambda}\|_H\big) \tag{13}$$

for all k sufficiently large. Moreover, $\{\rho_k\}$ remains bounded.

Proof. By Lemma 4.1, we can choose $\bar{\rho} > 0$ large enough such that, whenever $\rho_k \ge \bar{\rho}$, then x^k lies in a neighborhood of $\bar{x}$ where the error bound property (5) holds. Consider now the sequence $R_k := R(x^k, \lambda^k) = \|G(x^k) - P_K(G(x^k) + \lambda^k)\|_H$. We first show that $R_k \to 0$. Let $s^{k+1} := P_K(G(x^{k+1}) + w^k/\rho_k)$. Then $s^{k+1} \in K$ and $\lambda^{k+1} \in \mathcal{N}_K(s^{k+1})$ by [2, Prop. 6.46]. We now use the fact that $y \mapsto y - P_K(y + \lambda^{k+1})$ is nonexpansive, which is an easy consequence of [2, Cor. 4.10]. Therefore, the inverse triangle inequality yields

$$R_{k+1} \le \|G(x^{k+1}) - s^{k+1}\|_H + \|s^{k+1} - P_K(s^{k+1} + \lambda^{k+1})\|_H. \tag{14}$$

The last term is equal to zero since $\lambda^{k+1} \in \mathcal{N}_K(s^{k+1})$, cf. [2, Cor. 6.46]. Now, if $\{\rho_k\}$ is bounded, then the penalty updating scheme (11) implies $\|G(x^{k+1}) - s^{k+1}\|_H \to 0$, and $R_{k+1} \to 0$ follows. On the other hand, if $\rho_k \to \infty$ (which actually cannot occur, but for the sake of the proof we need to cover this case), then $x^{k+1} \to \bar{x}$ by Lemma 4.1, and

$$\|G(x^{k+1}) - s^{k+1}\|_H \le \|s^{k+1} - P_K(G(x^{k+1}))\|_H + d_K(G(x^{k+1})) \to 0$$

by the nonexpansiveness of P_K and the boundedness of $\{w^k\}$. Hence, in this case, we also obtain $R_{k+1} \to 0$.

The convergence $R_k \to 0$ implies, by virtue of the error bound property from Theorem 2.7, that $(x^k, \lambda^k) \to (\bar{x}, \bar{\lambda})$. Assume now that $w^k = \lambda^k$ for sufficiently large k. We prove the convergence rates by first showing that there is a $c > 0$ such that $R_{k+1} \le (c/\rho_k) R_k$ for all $k \in \mathbb{N}$ sufficiently large. Using (14) and the definition of λ^{k+1}, it follows that

$$R_{k+1} \le \|G(x^{k+1}) - s^{k+1}\|_H = \frac{\|\lambda^{k+1} - \lambda^k\|_H}{\rho_k} \le \frac{1}{\rho_k}\big(\|\lambda^{k+1} - \bar{\lambda}\|_H + \|\lambda^k - \bar{\lambda}\|_H\big). \tag{15}$$

By the error bound property (5), there is a $c_1 > 0$ such that $\|\lambda^k - \bar{\lambda}\|_H \le c_1 R_k$ for all $k \in \mathbb{N}$ sufficiently large (recall that $x^k \to \bar{x}$). Hence, we obtain $R_{k+1} \le (c_1/\rho_k)(R_{k+1} + R_k)$, or equivalently

$$\left(1 - \frac{c_1}{\rho_k}\right) R_{k+1} \le \frac{c_1}{\rho_k} R_k$$

for $k \in \mathbb{N}$ sufficiently large. Increasing the threshold value $\bar{\rho}$ if necessary, we may assume that $1 - c_1/\rho_k \ge 1/2$ and thus $R_{k+1} \le (2c_1/\rho_k) R_k$. Using the two-sided error bound (7), it is easy to deduce the convergence estimate (13).

Finally, let us show that $\{\rho_k\}$ remains bounded. To this end, we need to show that $V_{k+1} \le \tau V_k$ holds eventually, where $V_{k+1} := V(x^{k+1}, \lambda^k, \rho_k) = \|G(x^{k+1}) - s^{k+1}\|_H$ is the quantity used to determine the penalty update (11). From (15), we have $V_{k+1} \ge R_{k+1}$ and

$$V_{k+1} = \frac{\|\lambda^{k+1} - \lambda^k\|_H}{\rho_k} \le \frac{1}{\rho_k}\big(\|\lambda^{k+1} - \bar{\lambda}\|_H + \|\lambda^k - \bar{\lambda}\|_H\big) \le \frac{c_1}{\rho_k}(R_{k+1} + R_k).$$

Putting these inequalities together yields

$$\frac{V_{k+1}}{V_k} \le \frac{c_1}{\rho_k R_k}(R_{k+1} + R_k) \le \frac{c_1}{\rho_k}\left(1 + \frac{R_{k+1}}{R_k}\right).$$

If we now assume that $\rho_k \to \infty$, then it is easy to conclude that $V_{k+1}/V_k \to 0$. Hence, $V_{k+1}/V_k \le \tau$ for sufficiently large k, which contradicts $\rho_k \to \infty$. $\square$

5 Applications

In this section, we describe some applications of the above theory, in particular to semidefinite and second-order cone programming. Note that, for both these problem classes, we obtain convergence results which are stronger than those in the literature.

5.1 Semidefinite Programming

Semidefinite programming (SDP), linear or nonlinear, revolves around constraints which impose semidefiniteness of certain matrices. Throughout this section, we write $\mathcal{S}^n$ for the space of symmetric $n \times n$-matrices, equipped with the scalar product $\langle A, B \rangle := \operatorname{tr}(A^T B)$, $\mathcal{S}^n_+$ ($\mathcal{S}^n_-$) for the subsets of positive (negative) semidefinite matrices, and $A \succeq 0$ ($A \preceq 0$) for positive (negative) semidefiniteness. With these definitions, a typical SDP is given by

$$\min \; f(x) \quad \text{s.t.} \quad g(x) \le 0, \; h(x) = 0, \; \mathcal{G}(x) \preceq 0, \tag{16}$$

where $x \in X$ for some finite-dimensional space X and $g : X \to \mathbb{R}^m$, $h : X \to \mathbb{R}^p$, $\mathcal{G} : X \to \mathcal{S}^n$ are given mappings. This problem corresponds to our setting (P) with

$H := \mathbb{R}^m \times \mathbb{R}^p \times \mathcal{S}^n$, $G(x) := (g(x), h(x), \mathcal{G}(x))$, and $K := \mathbb{R}^m_- \times \{0\}^p \times \mathcal{S}^n_-$. Note that K is C^2-cone reducible because it is a Cartesian product of C^2-cone reducible sets [26].

For semidefinite programming, the Lagrange multiplier occurring in the KKT conditions can be split as $\bar{\lambda} = (\bar{\mu}, \bar{\nu}, \bar{\Gamma})$ with $\bar{\mu} \in \mathbb{R}^m$, $\bar{\nu} \in \mathbb{R}^p$, and $\bar{\Gamma} \in \mathcal{S}^n$. With an obvious change of notation, the Lagrange function now becomes

$$\mathcal{L}(x, \mu, \nu, \Gamma) := f(x) + \mu^T g(x) + \nu^T h(x) + \langle \Gamma, \mathcal{G}(x) \rangle,$$

and the KKT conditions from Definition 2.2 take on the form

$$\mathcal{L}'(\bar{x}, \bar{\mu}, \bar{\nu}, \bar{\Gamma}) = 0, \quad 0 \le \bar{\mu} \perp g(\bar{x}) \le 0, \quad h(\bar{x}) = 0, \quad 0 \preceq \bar{\Gamma} \perp \mathcal{G}(\bar{x}) \preceq 0.$$

Most of the conditions in this paper such as the strict Robinson condition or the second-order sufficient condition can be reformulated explicitly in the case of SDP. For SRC, a characterization was essentially obtained in [25] (see also [29]). As for SOSC, the σ-term occurring in (4) can be calculated explicitly [8], and the condition can therefore be rewritten as

$$\sup_{(\mu,\nu,\Gamma) \in \mathcal{M}(\bar{x})} \left\{ \mathcal{L}''(\bar{x}, \mu, \nu, \Gamma)(d, d) - 2\left\langle \Gamma, (\mathcal{G}'(\bar{x})d)\mathcal{G}(\bar{x})^\dagger(\mathcal{G}'(\bar{x})d) \right\rangle \right\} > 0 \tag{17}$$

for all $d \in C(\bar{x}) \setminus \{0\}$, see [8, 26, 29]. Note that the functions g and h provide no contribution to the σ-term since they represent constraints for which the corresponding factor in the set K is polyhedral.

Theorem 5.1. *Assume that $(\bar{x}, \bar{\mu}, \bar{\nu}, \bar{\Gamma})$ is a KKT point of the nonlinear SDP (16) satisfying the SOSC (17) and SRC. Then, under the assumptions of Theorem 4.2, the sequence $\{(x^k, \mu^k, \nu^k, \Gamma^k)\}$ generated by Algorithm 3.1 converges to $(\bar{x}, \bar{\mu}, \bar{\nu}, \bar{\Gamma})$, the rate of convergence is proportional to $1/\rho_k$, and $\{\rho_k\}$ remains bounded.*

Proof. This is a consequence of Theorem 4.2. □

Note that similar results were obtained in [27]; however, the theory established therein uses the so-called *strong SOSC* and *constraint nondegeneracy* assumptions, which are stronger than SOSC and SRC respectively, and assumes that the multiplier λ is close to the optimal multiplier $\bar{\lambda}$, which follows automatically from our analysis.

For extensive numerical results on augmented Lagrangian methods for SDP, we refer the reader to the existing literature [17, 18, 31]. We finish this section with some remarks on the linear case. Given a linear SDP of the form

$$\min \langle c, x \rangle \quad \text{s.t.} \quad Ax = b, \ x \succeq 0,$$

where $c \in \mathcal{S}^n$, $b \in \mathbb{R}^m$, and $A : \mathcal{S}^n \to \mathbb{R}^m$ is a linear operator, it is customary to apply the augmented Lagrangian method to the *dual* problem [31]

$$\max b^T y \quad \text{s.t.} \quad A^* y - c \succeq 0,$$

since this yields subproblems which are smooth, unconstrained minimization problems on $\mathbb{R}^m$. It turns out that SOSC and SRC for the dual problem are closely related to the corresponding primal properties. In fact, assuming that the problem admits a unique primal-dual solution pair, it can be shown that SOSC for the primal problem is equivalent to SRC for the dual problem [31, Cor. 2.2]. By duality, this also holds with SOSC and SRC interchanged. Hence, if both conditions hold for the primal problem, then they also hold for the dual problem (primal-dual uniqueness follows automatically in this case).

5.2 Second-Order Cone Programming

For second-order cone programs (SOCP), the theoretical analysis is very similar to semidefinite programming. Throughout this section, we write $w := (w_0, \bar{w})$ for a generic element in $\mathbb{R}^{1+m}$. Assume, for the sake of simplicity, that the optimization problem is given by (P) with K the second-order cone, i.e.

$$K := \{(w_0, \bar{w}) \in \mathbb{R}^{1+m} : w_0 \geq \|\bar{w}\|_2\},$$

where $\|\cdot\|_2$ is the Euclidean norm. The analysis below can easily be extended to the case where additional inequality, equality, or multiple second-order cone constraints are present. In any case, the resulting set K is C^2-cone reducible [26].

As in the case of semidefinite programming, the second-order sufficient condition from Definition 2.6 can be reformulated to take into account the particular structure of the problem. The resulting condition is given by

$$\sup_{\lambda \in \mathcal{M}(\bar{x})} \left\{ \mathcal{L}''(\bar{x}, \lambda)(d, d) + d^T \mathcal{H}(\bar{x}, \lambda) d \right\} > 0 \tag{18}$$

for all $d \in C(\bar{x}) \setminus \{0\}$, where

$$\mathcal{H}(\bar{x}, \lambda) := -\frac{\lambda_0}{G_0(\bar{x})} G'(\bar{x})^T \begin{pmatrix} 1 & 0 \\ 0 & -I_m \end{pmatrix} G'(\bar{x})$$

if $G(\bar{x}) \in \mathrm{bd}(K) \setminus \{0\}$ and $\mathcal{H}(\bar{x}, \lambda) := 0$ otherwise, see [7, 20, 30].

Theorem 5.2. *Assume that $(\bar{x}, \bar{\lambda})$ is a KKT point of the nonlinear SOCP (16) satisfying the SOSC (18) and SRC. Then, under the assumptions of Theorem 4.2, the sequence $\{(x^k, \lambda^k)\}$ generated by Algorithm 3.1 converges to $(\bar{x}, \bar{\lambda})$, the rate of convergence is proportional to $1/\rho_k$, and $\{\rho_k\}$ remains bounded.*

Proof. This is a consequence of Theorem 4.2. □

As with semidefinite programming, similar results to Theorem 5.2 have been obtained in the literature [19, 20]. However, both these papers use assumptions which are stronger than ours, in particular the *constraint nondegeneracy* assumption together with either SOSC and strict complementarity or the so-called *strong SOSC*. Moreover, the analysis in both papers assumes that λ is close to $\bar{\lambda}$, which follows automatically in our case.

6 Final Remarks

We have shown that the augmented Lagrangian method converges locally for C^2-cone reducible constrained optimization problems under the second-order sufficient condition and a strict version of the Robinson constraint qualification. Notably, our analysis does not require any assumptions on the initial multiplier estimate.

Moreover, we have shown that the aforementioned assumptions guarantee a primal-dual error bound on the KKT system which does not require any (a-priori) proximity of the multiplier to the optimal one.

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(4):1286–1309, 2007.

[2] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2011.

[3] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York-London, 1982.

[4] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1995.

[5] 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(1, Ser. A):139–162, 2010.

[6] E. G. Birgin and J. M. Martínez. *Practical Augmented Lagrangian Methods for Constrained Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014.

[7] J. F. Bonnans and H. Ramírez C. Perturbation analysis of second-order cone programming problems. *Math. Program.*, 104(2-3, Ser. B):205–227, 2005.

[8] J. F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000.

[9] A. R. Conn, N. I. M. Gould, and P. L. Toint. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM J. Numer. Anal.*, 28(2):545–572, 1991.

[10] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust-region methods*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA, 2000.

[11] L. Contesse-Becker. Extended convergence results for the method of multipliers for non-strictly binding inequality constraints. *J. Optim. Theory Appl.*, 79(2):273–310, 1993.

[12] C. Ding, D. Sun, and L. Zhang. Characterization of the robust isolated calmness for a class of conic programming problems. *SIAM J. Optim.*, 27(1):67–90, 2017.

[13] D. Fernández and M. V. Solodov. Local convergence of exact and inexact augmented Lagrangian methods under the second-order sufficient optimality condition. *SIAM J. Optim.*, 22(2):384–407, 2012.

[14] K. Ito and K. Kunisch. The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces. *Math. Programming*, 46(3, (Ser. A)):341–360, 1990.

[15] C. Kanzow and D. Steck. On error bounds and multiplier methods for variational problems in Banach spaces. *Technical Report*, Institute of Mathematics, University of Würzburg, September 2017.

[16] C. Kanzow, D. Steck, and D. Wachsmuth. An augmented Lagrangian method for optimization problems in Banach spaces. *Technical Report*, Institute of Mathematics, University of Würzburg, December 2016.

[17] M. Kočvara and M. Stingl. PENNON: a code for convex nonlinear and semidefinite programming. *Optim. Methods Softw.*, 18(3):317–333, 2003.

[18] M. Kočvara and M. Stingl. PENNON: a generalized augmented Lagrangian method for semidefinite programming. In *High performance algorithms and software for nonlinear optimization (Erice, 2001)*, volume 82 of *Appl. Optim.*, pages 303–321. Kluwer Acad. Publ., Norwell, MA, 2003.

[19] Y.-J. Liu and L.-W. Zhang. Convergence analysis of the augmented Lagrangian method for nonlinear second-order cone optimization problems. *Nonlinear Anal.*, 67(5):1359–1373, 2007.

[20] Y. J. Liu and L. W. Zhang. Convergence of the augmented Lagrangian method for nonlinear optimization problems over second-order cones. *J. Optim. Theory Appl.*, 139(3):557–575, 2008.

[21] B. S. Mordukhovich. *Variational analysis and generalized differentiation. I*, volume 330 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2006. Basic theory.

[22] J. Nocedal and S. J. Wright. *Numerical Optimization.* Springer, New York, second edition, 2006.

[23] J.-S. Pang and M. Fukushima. Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games. *Comput. Manag. Sci.*, 2(1):21–56, 2005.

[24] S. M. Robinson. Stability theory for systems of inequalities. II. Differentiable nonlinear systems. *SIAM J. Numer. Anal.*, 13(4):497–513, 1976.

[25] A. Shapiro. On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints. *SIAM J. Optim.*, 7(2):508–518, 1997.

[26] A. Shapiro. Sensitivity analysis of generalized equations. *J. Math. Sci.*, 115(4):2554–2565, 2003.

[27] D. Sun, J. Sun, and L. Zhang. The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Math. Program.*, 114(2, Ser. A):349–391, 2008.

[28] G. Wachsmuth. On LICQ and the uniqueness of Lagrange multipliers. *Oper. Res. Lett.*, 41(1):78–80, 2013.

[29] Y. Zhang and L. Zhang. On the upper Lipschitz property of the KKT mapping for nonlinear semidefinite optimization. *Oper. Res. Lett.*, 44(4):474–478, 2016.

[30] Y. Zhang, L. Zhang, J. Wu, and K. Wang. Characterizations of local upper Lipschitz property of perturbed solutions to nonlinear second-order cone programs. *Optimization*, 66(7):1079–1103, 2017.

[31] X.-Y. Zhao, D. Sun, and K.-C. Toh. A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM J. Optim.*, 20(4):1737–1765, 2010.