Local and Global Analysis of Multiplier Methods for Constrained Optimization in Banach Spaces*

Eike Börgens† Christian Kanzow† Daniel Steck†

January 22, 2019

Abstract. We propose an augmented Lagrangian method for the solution of constrained optimization problems in Banach spaces. The framework we consider is very general and encompasses a host of standard problems such as nonlinear programming, semidefinite programming, and optimal control. We analyze several convergence-related aspects of the method, including global convergence, the attainment of feasibility, and local convergence. In particular, a result is presented which guarantees local convergence of the method under the sole assumption that a solution of the problem satisfies the second-order sufficient condition. Finally, we give detailed numerical results for optimal control problems and mathematical programs with complementarity constraints.

Keywords. Augmented Lagrangian method, multiplier-penalty method, Banach space, global convergence, second-order sufficient condition, strong local convergence.

AMS subject classifications. 49K, 49M, 65K, 90C.

1 Introduction

The present manuscript is the latest in a series of articles on augmented Lagrangian methods in Banach spaces. Throughout this paper, we consider the *Banach space optimization problem*

$$\underset{x \in C}{\text{minimize}}\ f(x) \quad \text{subject to} \quad G(x) \in K, \tag{P_Y}$$

where X, Y are (real) Banach spaces, $f : X \to \mathbb{R}$ and $G : X \to Y$ suitable mappings, and $C \subseteq X$, $K \subseteq Y$ nonempty closed convex sets. Problems akin to (P_Y) have long been identified as a suitable framework for generic optimization; see [7] for more details. In particular, this framework encompasses standard nonlinear programming, semidefinite programming, second-order cone programs, optimal control, and many more.

Here, we apply a safeguarded augmented Lagrangian method (ALM) in order to solve (P_Y). The classical augmented Lagrangian approach (or multiplier method) is one of the standard techniques for the solution of finite-dimensional optimization problems and therefore described in almost all textbooks; see, for instance, [4, 5, 11, 34]. The need for bounded dual variables was recognized by many authors [10, 36, 37], eventually giving rise to the concept of *multiplier safeguarding* and the ALGENCAN software package [1, 6]. The underlying

*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; {eike.boergens,kanzow,daniel.steck}@mathematik.uni-wuerzburg.de

modification of the augmented Lagrangian method uses a slightly different update of the Lagrange multiplier and turns out to have very strong global convergence properties; see [6] and, in particular, [26] for more details.

On the other hand, methods of augmented Lagrangian type have also received considerable attention in the context of variational inequalities and certain Banach space optimization problems [3, 16, 18, 20–22, 31, 44], but many of these papers suffer from drawbacks such as a focus on particular problem classes, convex problems, finite-rank constraints, etc. A first attempt at augmented Lagrangian methods for *general* Banach space problems was given by the authors in [24]. That paper is aimed at optimization problems with L^2-type inequality constraints and uses a generalized notion of multiplier safeguarding combined with suitable embeddings to obtain a globally convergent algorithm. The underlying theory has since evolved quite significantly and, in addition, been extended to other problem classes such as generalized Nash equilibrium problems and quasi-variational inequalities [25, 28].

Keeping in mind the above discussion, the purpose of the present paper is twofold. Our first aim is to give a global convergence analysis of the safeguarded ALM under fairly general assumptions. Part of this theory can be seen as a special case of the highly evolved theory for quasi-variational inequalities from [28], but for the more concrete case of constrained optimization. As a result, the theory in the present setting is much more accessible and may therefore be useful to a larger audience. In any case, we provide in this paper a far more general and widely applicable theory than the previous optimization-specific paper [29], thus significantly enhancing the literature for the optimization case.

The second and major contribution of this paper is a local analysis of safeguarded multiplier methods in Banach spaces based on an established notion of second-order sufficient conditions (SOSC). The main result in this direction is that the augmented Lagrangian algorithm is locally well-defined and *strongly* convergent under the aforementioned SOSC. An attractive feature of the theory is that it does not require any constraint qualifications (which can be problematic in the infinite-dimensional case), thus opening up a large spectrum of applications on degenerate problems such as mathematical programs with complementarity constraints (MPCCs; see [23] for ALMs on finite-dimensional MPCCs).

Note that there exist some papers dealing with local convergence of optimization methods in finite dimensions, especially stabilized SQP-type methods, under SOSC alone [13, 15, 17, 45, 46]. The paper [14] shows that SOSC is also sufficient to prove local convergence of the (classical) augmented Lagrangian technique in finite dimensions, provided that the primal iterates and Lagrange multipliers are sufficiently close to a KKT point.

We are not aware of any paper in the infinite-dimensional setting where it is shown that one can get local convergence of augmented Lagrangian methods under the second-order condition without any other assumption. Hence, we believe that our result is the first of this kind. We also stress that this is not just a simple generalization of the analysis from [14], but that our assumptions are also weaker in the sense that we do not need any assumption regarding the multipliers to be close to an optimal Lagrange multiplier. In return, we get convergence of the primal iterates only, and we do not deal with convergence rates. (A rate-of-convergence analysis under suitable assumptions can be found in [27].)

Discussion of the Problem Setting

As mentioned before, many authors deal with the augmented Lagrangian method for specific problem classes only. In this paper, we will deal with the general framework (P_Y) under the assumption that the constraint $G(x) \in K$ can, roughly speaking, be interpreted as a Hilbert space constraint. More precisely, we assume that there is a dense embedding $e : Y \hookrightarrow H$ for some real Hilbert space H, and that $\mathcal{K} \subseteq H$ is a closed convex set with $e^{-1}(\mathcal{K}) = K$. Hence, problem (P_Y) is equivalent to

$$\underset{x \in C}{\text{minimize}}\ f(x) \quad \text{subject to} \quad e(G(x)) \in \mathcal{K}. \tag{P_H}$$

As commonly done when dealing with embeddings, we will omit the mapping e in most formulas. Clearly, our setting encompasses the case where Y itself is a Hilbert space. In that case, we can simply set $H := Y$, $\mathcal{K} := K$, and the problems (P_Y), (P_H) are identical.

In practice, however, the distinction between Y and H is usually necessary (cf. some of the examples in Section 8 and in [24]) for the treatment of the underlying optimization problem. The main reason for this is that our multiplier-penalty method makes extensive use of Hilbert-space techniques such as projections. On the other hand, we cannot entirely drop the space Y since, in many cases, the optimization problem admits a Lagrange multiplier in Y^* but not in H. Note that Y and H satisfy the chain

$$Y \overset{e}{\hookrightarrow} H \cong H^* \overset{e^*}{\hookrightarrow} Y^*, \tag{1.1}$$

which is occasionally referred to as a Gelfand triple. Hence, the notion of multipliers in H is stronger than that in Y^*. In fact, we will see that the question of whether or not (P_Y) admits a multiplier in H has direct consequences for the behavior of our algorithm.

The rest of this paper is organized as follows. Section 2 collects some preliminary results from optimization and Banach space theory. We continue with a precise statement of the main algorithm in Section 3 and convergence analyses pertaining to global minimization and KKT points in Sections 4 and 5, respectively. Section 6 contains some preliminary work on second-order conditions, and Section 7 includes a detailed discussion of the behavior of the ALM under second-order assumptions. Finally, Section 8 contains some applications and corresponding numerical results, and we give some concluding remarks in Section 9.

Notation

Throughout the paper, we denote strong and weak convergence by $\to$ and $\rightharpoonup$, respectively. Moreover, duality pairings are written as $\langle \cdot, \cdot \rangle$, scalar products in Hilbert spaces as $(\cdot, \cdot)$, and norms are denoted by $\|\cdot\|$ with an appropriate subscript to emphasize the corresponding space (e.g., $\|\cdot\|_X$). If S is a nonempty subset of a normed space, we denote by $d_S = \operatorname{dist}(\cdot, S)$ the distance to S with respect to the underlying norm. If S is a closed convex subset of a Hilbert space, we write P_S for the projection onto S. Throughout the paper, we denote by $\Phi \subseteq X$ the feasible set of (P_Y), and use a prime $'$ to denote the derivative of a function with respect to the primal variable $x \in X$.

2 Preliminaries

Given Banach spaces X and Z, an operator $T : X \to Z$ is called *bounded* if it maps bounded sets to bounded sets, *weakly sequentially continuous* if it maps weakly convergent to weakly convergent sequences, and *completely continuous* if it maps weakly convergent to strongly convergent sequences. It is well known that, given a differentiable completely continuous operator T, the Fréchet derivative $T'(x) \in L(X, Z)$ is completely continuous for all $x \in X$; see [12, Thm. 1.5.1]. It is also possible (but slightly more involved) to give sufficient conditions for the complete continuity of the derivative mapping $T' : X \to L(X, Z)$; see [35]. A function $f\colon X \to \mathbb{R}$ is called *weakly sequentially lower semicontinuous (weakly lsc)* if for all $x \in X$ and all weakly convergent sequences $x^k \rightharpoonup x$, it is valid that $\liminf_{k\to\infty} f(x^k) \geq f(x)$.

For a convex subset S of X (or Y, H), we write

$$S^\circ := \{\psi \in X^* : \langle \psi, s \rangle \leq 0 \ \forall s \in S\}$$

for the *polar* cone of S. Moreover, if $x \in S$ is a given point, we denote by

$$\mathcal{R}_S(x) := \{\alpha(s - x) \mid \alpha \geq 0,\ s \in S\}, \quad \mathcal{T}_S(x) := \operatorname{cl}\big(\mathcal{R}_S(x)\big), \quad \mathcal{N}_S(x) := \mathcal{T}_S(x)^\circ,$$

the *radial cone* (also called the *cone of feasible directions*), the *tangent cone*, and the *normal cone* of S at x, respectively. If $x \notin S$, we set $\mathcal{T}_S(x) := \mathcal{R}_S(x) := \mathcal{N}_S(x) := \emptyset$. All of these sets are convex cones; moreover, with the possible exception of $\mathcal{R}_S(x)$, they are closed.

The following result is a well known and easily proved closedness property of the multifunction $\mathcal{N}_C(\cdot)$.

Lemma 2.1. *Let X be a real Banach space and $C \subseteq X$ a closed convex set. Let $x \in C$, $\phi \in X^*$, $\{x^k\} \subseteq C$, $\{\phi^k\} \subseteq X^*$ such that $\phi^k \in \mathcal{N}_C(x^k)$ for all k, and assume that either* (i) *$x^k \rightharpoonup x$ and $\phi^k \to \phi$, or* (ii) *$x^k \to x$ and $\phi^k \rightharpoonup^* \phi$. Then $\phi \in \mathcal{N}_C(x)$.*

We now turn to the optimality conditions of the optimization problem (P_Y). To this end, we define the *Lagrange function* or *Lagrangian* of the problem as

$$\mathcal{L} : X \times Y^* \to \mathbb{R}, \quad \mathcal{L}(x, \lambda) := f(x) + \langle \lambda, G(x) \rangle. \tag{2.1}$$

This function occurs quite prominently in the Karush–Kuhn–Tucker conditions of (P_Y). Recall that $\mathcal{L}'$ denotes the derivative of $\mathcal{L}$ with respect to x.

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

$$-\mathcal{L}'(\bar{x}, \bar{\lambda}) \in \mathcal{N}_C(\bar{x}) \quad \text{and} \quad \bar{\lambda} \in \mathcal{N}_K(G(\bar{x})). \tag{2.2}$$

Note that (2.2) necessarily implies that $\bar{x} \in C$ and $G(\bar{x}) \in K$, since otherwise at least one of the corresponding normal cones would be empty.

Furthermore, we remark that the KKT conditions from Definition 2.2 are defined with respect to the space Y, but it is rather straightforward to make a similar definition for the problem (P_H) pertaining to the space H. Whenever we wish to distinguish these two, we will refer to them as the KKT conditions in Y and in H, respectively. Using the embeddings from (1.1), it is easy to see that the latter imply the former.

For the analysis of KKT points and Lagrange multipliers, it is inevitable to use some form of constraint qualification. The following is one of the most widely used conditions in infinite dimensions. Note that int_Y denotes the interior with respect to the norm in Y.

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

$$0 \in \operatorname{int}_Y\big[G(x) + G'(x)(C - x) - K\big], \tag{2.3}$$

If $x \in X$ is an arbitrary, not necessary feasible point and (2.3) holds, then we say that the *extended Robinson constraint qualification (extended RCQ, ERCQ)* holds in x.

Again, we note that the (extended) Robinson constraint qualification could be defined in H instead of Y. Similarly to before, the continuous embedding $Y \hookrightarrow H$ implies that RCQ in H is stronger than in Y.

For our purposes, it will be convenient to have an analogue of RCQ defined that is not restricted to feasible points; this is the extended RCQ (ERCQ). Note that the condition defining ERCQ is the same as for the standard Robinson constraint qualification. The only difference is that, for the latter, the point x has to be feasible, whereas ERCQ is defined for arbitrary points.

If $x \in X$ is a (local) minimizer of (P_Y) and RCQ holds in x, then there exists a Lagrange multiplier $\lambda \in Y^*$ such that (x, λ) is a KKT point of (P_Y); see, for instance, [7].

Assume now that we have a point $\hat{x}$ which is "almost" a solution of (P_Y). A popular definition in this context is that of *ε-minimizers*: given $\varepsilon > 0$, we say that $\hat{x} \in \Phi$ is an ε-minimizer of (P_Y) if $f(\hat{x}) \leq f(x) + \varepsilon$ for all $x \in \Phi$. For such approximate minimizers, it is indeed possible to obtain an inexact analogue of the KKT conditions. This result is usually called Ekeland's variational principle.

Proposition 2.4 (Ekeland's variational principle, [7, Thm. 3.23])**.** *Let $\bar{x} \in \Phi$ be an ε-minimizer of (P_Y), let $\delta := \varepsilon^{1/2}$, and assume that RCQ holds at every $x \in B_\delta(\bar{x}) \cap \Phi$. Then there exist another ε-minimizer $\hat{x}$ of (P_Y) and $\lambda \in Y^*$ such that $\|\hat{x} - \bar{x}\|_X \leq \delta$,*

$$\operatorname{dist}\big(-\mathcal{L}'(\hat{x}, \lambda), \mathcal{N}_C(\hat{x})\big) \leq \delta, \quad \text{and} \quad \lambda \in \mathcal{N}_K(G(\hat{x})).$$

The final ingredient we will need in this paper is a generalized form of continuity of the function f (or, more precisely, its derivative). In the infinite-dimensional literature, variational problems are often treated in the scenario where the variational operator is monotone (in this case, this would mean the convexity of f). To achieve a higher level of generality, we will use the notion of *pseudomonotonicity* as introduced by H. Brezis [8]. This can be seen as a generalization of monotonicity, but it is in fact satisfied for large classes of nonmonotone operators simply by virtue of certain continuity properties (see below).

Definition 2.5 (Brezis pseudomonotonicity)**.** We say that an operator $F : X \to X^*$ is *pseudomonotone* if, whenever

$$\{x^k\} \subseteq X, \quad x^k \rightharpoonup x, \quad \text{and} \quad \limsup_{k\to\infty} \big\langle F(x^k), x^k - x \big\rangle \leq 0,$$

then

$$\langle F(x), x - y\rangle \leq \liminf_{k\to\infty} \big\langle F(x^k), x^k - y \big\rangle \quad \text{for all } y \in X.$$

The notion of pseudomonotonicity will play a fundamental role in Section 5. Some sufficient conditions for pseudomonotone operators are summarized in the following lemma.

Lemma 2.6 (Sufficient conditions for pseudomonotonicity)**.** *Let X be a real Banach space and $T, U : X \to X^*$ given operators. Then:*

(a) *If T is monotone and continuous, then T is pseudomonotone.*

(b) *If, for every $y \in X$, the mapping $x \mapsto \langle T(x), x - y\rangle$ is weakly sequentially lsc, then T is pseudomonotone.*

(c) *If T is completely continuous, then T is pseudomonotone.*

(d) *If T is continuous and $\dim(X) < +\infty$, then T is pseudomonotone.*

(e) *If T and U are pseudomonotone, then $T+U$ is pseudomonotone.*

Proof. (b) is obvious. The remaining assertions can be found in [47, Prop. 27.6]. Note that [47] defines pseudomonotonicity only on reflexive Banach spaces, but this property is not used in the proof of the result. □

Note that, if the underlying space X is finite dimensional, then pseudomonotonicity is implied by ordinary continuity, and the converse holds for bounded operators. Thus, in this case, Definition 2.5 is satisfied by any continuous operator.

3 The Multiplier-Penalty Method

We now turn to the multiplier-penalty method for the optimization problem (P_Y), whose analysis is the main subject of this work. To this end, we define the augmented Lagrangian of (P_H) as follows.

Definition 3.1 (Augmented Lagrange function)**.** For $\rho > 0$, the *augmented Lagrange function* or *augmented Lagrangian* of (P_H) is the function

$$\mathcal{L}_\rho : X \times H \to \mathbb{R}, \quad \mathcal{L}_\rho(x, \lambda) := f(x) + \frac{\rho}{2} d_{\mathcal{K}}^2\left(G(x) + \frac{\lambda}{\rho}\right) - \frac{\|\lambda\|_H^2}{2\rho}. \tag{3.1}$$

Note that there are other variants of $\mathcal{L}_\rho$ in the literature. However, these differ from (3.1) only by an additive constant (with respect to x).

A basic property of the augmented Lagrangian is stated in the following lemma.

Lemma 3.2. *Let $\mathcal{L}_\rho : X \times H \to \mathbb{R}$ be the augmented Lagrangian* (3.1)*. If $x \in X$ is a feasible point, then $\mathcal{L}_\rho(x, \lambda) \leq f(x)$ for all $\lambda \in H$.*

Proof. If $G(x) \in \mathcal{K}$, then $d_{\mathcal{K}}(G(x) + \lambda/\rho) \leq \|\lambda\|_H/\rho$ by the nonexpansiveness of the distance function. Hence, $\mathcal{L}_\rho(x, \lambda) \leq f(x) + (\rho/2)\|\lambda\|_H^2/\rho^2 - \|\lambda\|_H^2/(2\rho) = f(x)$. □

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

$$V(x, \lambda, \rho) = \left\| G(x) - P_{\mathcal{K}}\left(G(x) + \frac{\lambda}{\rho}\right)\right\|_H, \tag{3.2}$$

The definitions above enable us to formulate our algorithm as follows.

Algorithm 3.3 (ALM for constrained optimization)**.** Let $(x^0, \lambda^0) \in X \times H$, $\rho_0 > 0$, let $B \subseteq H$ be a nonempty bounded set, $\gamma > 1$, $\tau \in (0, 1)$, and set $k := 0$.

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

Step 2. Choose $w^k \in B$ and compute an approximate solution x^{k+1} of the problem

$$\underset{x \in C}{\text{minimize}}\ \mathcal{L}_{\rho_k}(x, w^k). \tag{3.3a}$$

Step 3. Update the vector of multipliers to

$$\lambda^{k+1} := \rho_k\left[G(x^{k+1}) + \frac{w^k}{\rho_k} - P_{\mathcal{K}}\left(G(x^{k+1}) + \frac{w^k}{\rho_k}\right)\right]. \tag{3.3b}$$

Step 4. Let $V_{k+1} := V(x^{k+1}, w^k, \rho_k)$ and set

$$\rho_{k+1} := \begin{cases} \rho_k, & \text{if } k = 0 \text{ or } V_{k+1} \le \tau V_k, \\ \gamma\rho_k, & \text{otherwise.} \end{cases} \tag{3.3c}$$

Step 5. Set $k \leftarrow k+1$ and go to Step 1.

Note that we have not specified what constitutes an "approximate solution" in Step 2. Clearly, there are multiple possibilities when solving the subproblem (3.3a); for instance, we could look for global minima, local minima, or KKT points. Moreover, the convergence properties of the overall method depend on the manner in which the subproblems are solved.

4 Convergence to Global Minimizers

In this section, we analyze the convergence properties of Algorithm 3.3 under the assumption that we can solve the subproblems in an (essentially) global sense. This is of course a rather restrictive requirement and can, in general, only be expected under certain convexity assumptions. However, the resulting theory is still appealing due to its simplicity. Indeed, the results below merely require some rather mild form of continuity (no differentiability), and can easily be extended to the case where the function f is extended-valued, i.e., it is allowed to take on the value $+\infty$.

Assumption 4.1 (Global minimization)**.** We make the following assumptions:

(i) f and $d_{\mathcal{K}} \circ G$ are weakly sequentially lsc on C.

(ii) The sequence $\{x^k\}$ satisfies $x^k \in C$ for all k.

(iii) For every $x \in C$, there is a null sequence $\{\varepsilon_k\} \subseteq \mathbb{R}$ such that $\mathcal{L}_{\rho_k}(x^{k+1}, w^k) \le \mathcal{L}_{\rho_k}(x, w^k) + \varepsilon_{k+1}$ for all k.

Some remarks are due. Recall that, for convex functions, weak sequential lower semicontinuity is implied by ordinary continuity (see [9, Cor. 3.9]). Thus, if f is a continuous convex function, then f is weakly sequentially lsc.

A similar comment applies to the weak sequential lower semicontinuity of the function $d_{\mathcal{K}} \circ G$. Indeed, there are two rather general situations in which this condition is satisfied: if G is weakly sequentially continuous, then $d_{\mathcal{K}} \circ G$ is weakly sequentially lsc since $d_{\mathcal{K}}$ is so by its continuity and convexity. On the other hand, if G is continuous and concave with respect to the recession cone of $\mathcal{K}$ in the sense of [27] (for instance, G could be affine), then it can be seen that $d_{\mathcal{K}} \circ G$ is a continuous convex function and thus again weakly sequentially lsc. Let us also remark that, if G is continuous and affine, then both the above cases apply.

Finally, another salient feature of Assumption 4.1 is the dependence of the sequence $\{\varepsilon_k\}$ on the comparison point $x \in C$. The motivation behind this is that, if (P_Y) is a smooth convex problem and the point x^{k+1} is "nearly stationary" in the sense that $\operatorname{dist}(-\mathcal{L}'_{\rho_k}(x^{k+1}, w^k), \mathcal{N}_C(x^{k+1})) \le \delta$ for some (small) $\delta > 0$, then, by convexity, we obtain an estimate of the form

$$\begin{aligned} \mathcal{L}_{\rho_k}(x, w^k) &\ge \mathcal{L}_{\rho_k}(x^{k+1}, w^k) + \mathcal{L}'_{\rho_k}(x^{k+1}, w^k)(x - x^{k+1}) \\ &\ge \mathcal{L}_{\rho_k}(x^{k+1}, w^k) - \delta \|x^{k+1} - x\|_X. \end{aligned}$$

This suggests that we should allow the sequence $\{\varepsilon_k\}$ in Assumption 4.1 to depend on the point x. In any case, the stated assumption is satisfied automatically if x^{k+1} is a global ε_{k+1}-minimizer of $\mathcal{L}_{\rho_k}(\cdot, w^k)$ for some null sequence $\{\varepsilon_k\}$.

We now turn to the convergence analysis of Algorithm 3.3 under Assumption 4.1. The theory is divided into separate analyses of feasibility and optimality. Since the augmented Lagrangian method is, at its heart, a penalty-type algorithm, the attainment of feasibility is particularly important for the success of the algorithm. A closer look at the definition of the augmented Lagrangian suggests that, if ρ is large, then the minimization of $\mathcal{L}_\rho$ essentially reduces to that of the infeasibility measure $d_{\mathcal{K}}^2(G(x))$. Hence, we can expect (weak) limit points of the sequence $\{x^k\}$ to be minimizers of this auxiliary function, which means that, roughly speaking, these points are "as feasible as possible." A precise statement of this assertion can be found in the following lemma.

Lemma 4.2. *Let $\{x^k\}$ be generated by Algorithm 3.3, let Assumption 4.1 hold, and let $\bar{x}$ be a weak limit point of $\{x^k\}$. Then $\bar{x}$ is a global minimizer of the function $d_{\mathcal{K}} \circ G$ on C. In particular, if the feasible set of (P_Y) is nonempty, then $\bar{x}$ is feasible.*

Proof. Note that C is weakly sequentially closed since it is closed and convex, hence $\bar{x} \in C$. To show the desired minimization property, we first consider the case where $\{\rho_k\}$ remains bounded. Then (3.3c) and the definition of V yield

$$0 \le d_{\mathcal{K}}(G(x^{k+1})) \le \left\| G(x^{k+1}) - P_{\mathcal{K}}\left(G(x^{k+1}) + \frac{w^k}{\rho_k} \right) \right\|_H = V_{k+1} \to 0.$$

The weak sequential lower semicontinuity of $d_{\mathcal{K}} \circ G$ (by Assumption 4.1) therefore implies that $d_{\mathcal{K}}(G(\bar{x})) = 0$. Hence, $\bar{x}$ is feasible and there is nothing to prove.

Assume now that $\rho_k \to \infty$, and let $x^{k+1} \rightharpoonup_I \bar{x}$ for some subset $I \subseteq \mathbb{N}$. Let $x \in C$ be an arbitrary point and let $\{\varepsilon_k\}$ be the corresponding null sequence from Assumption 4.1. Then $\mathcal{L}_{\rho_k}(x^{k+1}, w^k) \le \mathcal{L}_{\rho_k}(x, w^k) + \varepsilon_{k+1}$ for all k, which implies

$$f(x^{k+1}) + \frac{\rho_k}{2} d_{\mathcal{K}}^2\left(G(x^{k+1}) + \frac{w^k}{\rho_k} \right) \le f(x) + \frac{\rho_k}{2} d_{\mathcal{K}}^2\left(G(x) + \frac{w^k}{\rho_k} \right) + \varepsilon_{k+1}. \tag{4.1}$$

Observe now that $\{f(x^{k+1})\}_{k \in I}$ is bounded from below since f is weakly sequentially lsc. Hence, dividing both sides by ρ_k and taking the lim inf for $k \in I$, we obtain

$$\liminf_{k \in I} d_{\mathcal{K}}^2\left(G(x^{k+1}) + \frac{w^k}{\rho_k} \right) \le \liminf_{k \in I} d_{\mathcal{K}}^2\left(G(x) + \frac{w^k}{\rho_k} \right) = d_{\mathcal{K}}^2(G(x)).$$

Using the fact that $w^k/\rho_k \to 0$ and that $d_{\mathcal{K}} \circ G$ is weakly sequentially lsc, it follows that the left-hand side is greater than or equal to $d_{\mathcal{K}}^2(G(\bar{x}))$. This completes the proof. □

The idea to link the feasibility properties of the iterates $\{x^k\}$ to the minimization of the infeasibility measure $d_{\mathcal{K}}^2 \circ G$ is a recurring theme in the convergence theory of augmented Lagrangian methods. In fact, we will encounter similar statements in the context of stationary points.

Let us now turn to the optimality part.

Theorem 4.3 (Convergence to global minimizers)**.** *Let $\{x^k\}$ be generated by Algorithm under Assumption 4.1* (ii)–(iii)*. Then*

$$\limsup_{k \to \infty} f(x^k) \le f(x) \quad \textit{for all } x \in \Phi.$$

If furthermore Assumption 4.1 (i) *holds and the feasible set of* (P_Y) *is nonempty, then every weak limit point of* $\{x^k\}$ *is a global solution of* (P_Y).

Proof. Let $x \in X$ be an arbitrary feasible point, and let $\{\varepsilon_k\}$ be the sequence from Assumption 4.1. By Lemma 3.2 and the assumption, we have

$$f(x^{k+1}) + \frac{\rho_k}{2} d_{\mathcal{K}}^2\left(G(x^{k+1}) + \frac{w^k}{\rho_k}\right) - \frac{\|w^k\|_H^2}{2\rho_k} \leq \mathcal{L}_{\rho_k}(x, w^k) + \varepsilon_{k+1} \leq f(x) + \varepsilon_{k+1}. \tag{4.2}$$

Clearly, if $\rho_k \to \infty$, then $\|w^k\|_H^2/(2\rho_k) \to 0$. In this case, the nonnegativity of $d_{\mathcal{K}}$ and the fact that $\varepsilon_k \to 0$ imply $\limsup_{k\to\infty} f(x^{k+1}) \leq f(x)$.

Consider now the case where $\{\rho_k\}$ remains bounded. The triangle inequality yields

$$d_{\mathcal{K}}\left(G(x^{k+1}) + \frac{w^k}{\rho_k}\right) \leq \left\|\frac{w^k}{\rho_k}\right\|_H + \left\|G(x^{k+1}) - P_{\mathcal{K}}\left(G(x^{k+1}) + \frac{w^k}{\rho_k}\right)\right\|_H.$$

The last term converges to zero by the penalty updating scheme (3.3c). Using the boundedness of $\{w^k\}$ and squaring on both sides, it is easy to deduce that

$$\limsup_{k\to\infty}\left[d_{\mathcal{K}}^2\left(G(x^{k+1}) + \frac{w^k}{\rho_k}\right) - \left\|\frac{w^k}{\rho_k}\right\|_H^2\right] \leq 0.$$

Since $\{\rho_k\}$ is bounded, it follows again from (4.2) that $\limsup_{k\to\infty} f(x^{k+1}) \leq f(x)$.

Now let Assumption 4.1 be satisfied and the feasible set of (P_Y) be nonempty. Suppose that $x^{k+1} \rightharpoonup_I \bar{x}$ for some (infinite) subset $I \subseteq \mathbb{N}$. Then $\bar{x}$ is feasible by Lemma 4.2, and the weak sequential lower semicontinuity of f implies that $f(\bar{x}) \leq \liminf_{k\in I} f(x^{k+1}) \leq f(x)$ for every feasible x. Hence, $\bar{x}$ is a global solution of (P_Y). □

If the problem is convex with strongly convex objective, then it is possible to considerably strengthen the results of the previous theorem. Recall that, in this case, the weak sequential lower semicontinuity of f from Assumption 4.1 is implied by (ordinary) continuity. Recall also that a sufficient condition for the convexity of the feasible set Φ is the concavity of G with respect to the recession cone of $\mathcal{K}$, see [27]. In this case the distance function $d_{\mathcal{K}} \circ G$ is convex, and thus the weak sequential lower semicontinuity from Assumption 4.1 is implied by (ordinary) continuity of G.

Corollary 4.4. *Let* $\{x^k\}$ *be generated by Algorithm 3.3 and let Assumption 4.1 hold. Assume that* X *is reflexive,* f *is strongly convex on* C*, and the feasible set of* (P_Y) *is nonempty and convex. Then* $\{x^k\}$ *converges strongly to the unique solution of* (P_Y).

Proof. Note that Assumption 4.1 implies that the feasible set Φ is closed. Since f is strongly convex, the existence and uniqueness of the solution $\bar{x}$ follows from standard arguments. Now, denoting by $c > 0$ the modulus of convexity of f, it follows that

$$\frac{c}{8}\|x^{k+1} - \bar{x}\|_X^2 \leq \frac{f(x^{k+1}) + f(\bar{x})}{2} - f\left(\frac{x^{k+1} + \bar{x}}{2}\right) \tag{4.3}$$

for all k. Moreover, by Theorem 4.3, we have $\limsup_{k\to\infty} f(x^{k+1}) \leq f(\bar{x})$. Taking into account that f is bounded from below, it follows from (4.3) that $\{x^k\}$ is bounded. Since X is reflexive and every weak limit point of $\{x^k\}$ is a solution of (P_Y) by Theorem 4.3, it follows that $x^k \rightharpoonup \bar{x}$. Since f is weakly sequentially lsc and $\limsup_{k\to\infty} f(x^{k+1}) \leq f(\bar{x})$, we conclude that $f(x^{k+1}) \to f(\bar{x})$. Moreover, since $(x^k + \bar{x})/2 \rightharpoonup \bar{x}$, we also have $f(\bar{x}) \leq \liminf_{k\to\infty} f((x^{k+1} + \bar{x})/2)$. Hence, (4.3) implies that $\|x^{k+1} - \bar{x}\|_X \to 0$. □

5 Stationarity of Limit Points

The theory on global minimization which we have developed in the preceding section is certainly appealing from a theoretical point of view. However, the practical relevance of the corresponding results is essentially limited to problems where some form of convexity is present. It therefore seems natural to conduct a dedicated analysis for the augmented Lagrangian method which, instead of global minimization, takes into account the optimality and stationary concepts from (P_Y).

The present section is dedicated to precisely this approach. To that end, we assume that the functions defining the optimization problem are continuously differentiable and that we are able to compute local minimizers or stationary points of the subproblems (3.3a) which occur in the algorithm. Note that the first-order optimality conditions of these problems are given by

$$-\mathcal{L}'_{\rho_k}(x, w^k) \in \mathcal{N}_C(x).$$

Similarly to the previous section, we will allow for certain inexactness terms. A natural way of doing this is by considering the inexact first-order optimality condition

$$\varepsilon^{k+1} - \mathcal{L}'_{\rho_k}(x, w^k) \in \mathcal{N}_C(x),$$

where $\varepsilon^{k+1} \in X^*$ is an error term. For $k \to \infty$, the degree of inexactness should vanish in the sense that $\varepsilon^k \to 0$. Hence, we arrive at the following assumption.

Assumption 5.1 (Convergence to KKT points)**.** We assume that

(i) f and G are continuously differentiable on X,

(ii) the derivative f' is bounded and pseudomonotone,

(iii) G and G' are completely continuous on C, and

(iv) $x^{k+1} \in C$ and $\varepsilon^{k+1} - \mathcal{L}'_{\rho_k}(x^{k+1}, w^k) \in \mathcal{N}_C(x^{k+1})$ for all k, where $\varepsilon^k \to 0$.

Notice that $\mathcal{L}_{\rho_k}$ is continuously differentiable by the differentiability of $d_{\mathcal{K}}^2$. The derivative $\mathcal{L}'_{\rho_k}$ (with respect to x) can be calculated by applying the chain rule together with a standard projection theorem. One obtains

$$\mathcal{L}'_{\rho_k}(x, w^k) = f'(x) + \rho_k G'(x)^* \left[G(x) + \frac{w^k}{\rho_k} - P_{\mathcal{K}}\left(G(x) + \frac{w^k}{\rho_k} \right) \right] \tag{5.1}$$

and, in particular, $\mathcal{L}'_{\rho_k}(x^{k+1}, w^k) = \mathcal{L}'(x^{k+1}, \lambda^{k+1})$.

As in the previous section, we treat the questions of feasibility and optimality in a separate manner. For the feasibility part, we relate the augmented Lagrangian to the infeasibility measure $d_{\mathcal{K}}^2 \circ G$.

Lemma 5.2. *Let $\{x^k\}$ be generated by Algorithm 3.3 under Assumption 5.1, and let $\bar{x}$ be a weak limit point of $\{x^k\}$. Then $\bar{x}$ is a stationary point of the problem $\min_{x \in C} d_{\mathcal{K}}^2(G(x))$. If ERCQ holds in $\bar{x}$ with respect to the constraint system of (P_Y), then $G(\bar{x}) \in K$.*

Proof. Let $x^{k+1} \rightharpoonup_I \bar{x}$ for some index set $I \subseteq \mathbb{N}$. Observe that $\bar{x} \in C$ by the weak sequential closedness of C. If $\{\rho_k\}$ is bounded, then we can argue as in Lemma 4.2 to see that $\bar{x}$ is feasible, and there is nothing to prove. If $\rho_k \to \infty$, then Assumption 5.1 implies that

$$\varepsilon^{k+1} - \mathcal{L}'_{\rho_k}(x^{k+1}, w^k) \in \mathcal{N}_C(x^{k+1})$$

for all $k \in \mathbb{N}$. We now divide this inclusion by ρ_k, use (5.1) and the fact that $\mathcal{N}_C(x^{k+1})$ is a cone. It follows that

$$\frac{\varepsilon^{k+1} - f'(x^{k+1})}{\rho_k} - G'(x^{k+1})^* \left[G(x^{k+1}) + \frac{w^k}{\rho_k} - P_{\mathcal{K}}\left(G(x^{k+1}) + \frac{w^k}{\rho_k} \right) \right] \in \mathcal{N}_C(x^{k+1}).$$

We now take the limit $k \to_I \infty$, use the boundedness of $\{f'(x^{k+1})\}$, the complete continuity of G and G' (by Assumption 5.1), and Lemma 2.1. This yields $G'(\bar{x})^*[P_{\mathcal{K}}(G(\bar{x})) - G(\bar{x})] \in \mathcal{N}_C(\bar{x})$, which is precisely the first-order optimality condition of $\min_{x \in C} d_{\mathcal{K}}^2(G(x))$.

Now, assume that ERCQ holds in $\bar{x}$, and let $r > 0$ be such that $B_r^Y \subseteq G(\bar{x}) + G'(\bar{x})(C - \bar{x}) - K$, where B_r^Y is the closed r-ball around zero in Y. Then, for any $y \in B_r^Y$, there are $z \in C$ and $w \in K$ such that $y = G(\bar{x}) + G'(\bar{x})(z - \bar{x}) - w$. Thus, we have

$$\begin{aligned} \langle G(\bar{x}) - P_{\mathcal{K}}(G(\bar{x})), y \rangle = {} & \big\langle G'(\bar{x})^* \big[G(\bar{x}) - P_{\mathcal{K}}(G(\bar{x}))\big], z - \bar{x} \big\rangle \\ & + \langle G(\bar{x}) - P_{\mathcal{K}}(G(\bar{x})), G(\bar{x}) - w \rangle. \end{aligned}$$

Observe that $G'(\bar{x})^*[G(\bar{x}) - P_{\mathcal{K}}(G(\bar{x}))]$ is just the derivative of $\frac{1}{2} d_{\mathcal{K}}^2 \circ G$ in $\bar{x}$. Hence, the first term above is nonnegative by the stationarity of $\bar{x}$, and so is the second term by standard projection inequalities. Thus, $\langle G(\bar{x}) - P_{\mathcal{K}}(G(\bar{x})), y \rangle \geq 0$ for all $y \in B_r^Y$, which implies $\langle G(\bar{x}) - P_{\mathcal{K}}(G(\bar{x})), y \rangle = 0$ for all $y \in B_r^Y$ and, since Y is dense in H, it follows that $G(\bar{x}) - P_{\mathcal{K}}(G(\bar{x})) = 0$. This completes the proof. □

The above lemma indicates that weak limit points of the sequence $\{x^k\}$ have a strong tendency to be feasible points. Apart from the heuristic appeal of the result, there are several nontrivial cases where Lemma 5.2 automatically implies the feasibility of the limit point $\bar{x}$. Here, two cases in particular deserve a special mention: first, if ERCQ is fulfilled, as already mentioned in the Lemma. Second, if the function $d_{\mathcal{K}}^2 \circ G$ is convex as already discussed above, and it follows that $\bar{x}$ is a global minimizer of this function. Hence, if the feasible set Φ is nonempty, then $\bar{x} \in \Phi$.

The assertion of the next lemma can be described roughly as a kind of "asymptotic normality" between λ^k and $G(x^k)$, the proof can be found in [28, Lem. 5.2].

Lemma 5.3. *There is a null sequence $\{r_k\} \subseteq [0, \infty)$ such that $(\lambda^k, y - G(x^k)) \leq r_k$ for all $y \in \mathcal{K}$ and $k \in \mathbb{N}$.*

Note that, by virtue of the Gel'fand triple $Y \hookrightarrow H \hookrightarrow Y^*$, the inequality from Lemma 5.3 also holds if we replace $\mathcal{K}$ by K and the scalar product by the duality pairing on $Y^* \times Y$. Recall furthermore that the KKT conditions of (P_Y) postulate the existence of a Lagrange multiplier $\bar{\lambda} \in \mathcal{N}_K(G(\bar{x})) \subseteq Y^*$. The assertion of Lemma 5.3 is essentially an asymptotic analogue of this condition, and it will prove useful in the convergence analysis.

Example 5.4 (Cone constraints)**.** If the set $\mathcal{K}$ is a closed convex *cone*, then some parts of Algorithm 3.3 and Lemma 5.3 can be simplified. In this case, we can use the Moreau decomposition ([2, Thm. 14.3]) to restate the multiplier update (3.3b) as $\lambda^{k+1} = P_{\mathcal{K}^\circ}(w^k + \rho_k G(x^{k+1}))$, therefore $\lambda^k \in \mathcal{K}^\circ$ for all k. This implies that $(\lambda^k, y) \leq 0$ for all $y \in \mathcal{K}$, and it is easy to see that the assertion of Lemma 5.3 is then equivalent to

$$\liminf_{k \to \infty} \big(\lambda^k, G(x^k)\big) \geq 0.$$

We now analyze the optimality properties of limit points. Recall that $\mathcal{L}'_{\rho_k}(x^{k+1}, w^k) = \mathcal{L}'(x^{k+1}, \lambda^{k+1})$ for all k. Hence, combining Assumption 5.1 and Lemma 5.3, we obtain the *asymptotic KKT conditions* (for $k \geq 1$)

$$\varepsilon^k - \mathcal{L}'(x^k, \lambda^k) \in \mathcal{N}_C(x^k) \quad \text{and} \quad \langle \lambda^k, y - G(x^k) \rangle \leq r_k \ \forall y \in K. \tag{5.2}$$

Note that the second inequality also holds with K replaced by $\mathcal{K}$, by the embedding. Hence, our main goal is to use these asymptotic KKT conditions to obtain the optimality of weak limit points of $\{x^k\}$. The main result in this direction is the following.

Theorem 5.5. *Let $\{(x^k, \lambda^k)\}$ be generated by Algorithm 3.3 under Assumption 5.1, let $x^k \rightharpoonup_I \bar{x}$ for some index set $I \subseteq \mathbb{N}$, and let $\bar{x}$ satisfy ERCQ with respect to the constraint system of (P_Y). Then $\bar{x}$ is a stationary point of (P_Y), the sequence $\{\lambda^k\}_{k \in I}$ is bounded in Y^*, and each of its weak-* limit points belongs to $\Lambda(\bar{x})$.*

Proof. Note that $\bar{x}$ is feasible by Lemma 5.2. We first prove the boundedness of $\{\lambda^k\}$ in Y^*. Applying the generalized open mapping theorem [7, Thm. 2.70] to the multifunction $\mathcal{W}(u) := G(\bar{x}) + G'(\bar{x})u - K$ on the domain $C - \bar{x}$, we obtain the existence of $r > 0$ such that

$$B_r^Y \subseteq G(\bar{x}) + G'(\bar{x})\big[(C - \bar{x}) \cap B_1^X\big] - K,$$

where B_r^Y again denotes the closed r-ball around zero in Y. Now, let $\{y^k\} \subseteq Y$ be a sequence of unit vectors such that $\langle \lambda^k, y^k \rangle \geq \frac{1}{2}\|\lambda^k\|_{Y^*}$. Then $-ry^k \in B_r^Y$ and therefore

$$-ry^k = G(\bar{x}) + G'(\bar{x})(v^k - \bar{x}) - z^k$$

with $\{v^k\} \subseteq C$ a bounded sequence and $\{z^k\} \subseteq K$. It follows that $ry^k = z^k - G(x^k) - G'(x^k)(v^k - x^k) + \delta^k$ with $\delta^k \to 0$ as $k \to \infty$. Let k be large enough so that $\|\delta^k\|_Y \leq r/4$. Then, by the asymptotic KKT conditions (5.2), we obtain

$$\begin{aligned} \frac{r}{2}\|\lambda^k\|_{Y^*} \leq \langle \lambda^k, ry^k \rangle &\leq \langle \lambda^k, z^k - G(x^k) \rangle - \langle \lambda^k, G'(x^k)(v^k - x^k) \rangle + \frac{r}{4}\|\lambda^k\|_{Y^*} \\ &\leq \langle \lambda^k, z^k - G(x^k) \rangle + \langle f'(x^k) - \varepsilon^k, v^k - x^k \rangle + \frac{r}{4}\|\lambda^k\|_{Y^*}. \end{aligned}$$

Now, using again (5.2) and the boundedness of f', it follows that the first two terms are bounded from above by some constant $c > 0$. Hence, $\frac{r}{4}\|\lambda^k\|_{Y^*} \leq c$.

We now show the second assertion. Let $I \subseteq \mathbb{N}$ be an (infinite) subset such that $\lambda^k \rightharpoonup_I^* \bar{\lambda}$ in Y^*. By (5.2) and Lemma 2.1, we have $\bar{\lambda} \in \mathcal{N}_K(G(\bar{x}))$. Now, let $y \in C$ be arbitrary. Then, by (5.2),

$$\langle \varepsilon^k, y - x^k \rangle \leq \langle f'(x^k), y - x^k \rangle + \langle \lambda^k, G'(x^k)(y - x^k) \rangle. \tag{5.3}$$

By complete continuity, we have $G'(x^k) \to G'(\bar{x})$. The complete continuity of G implies that $G'(\bar{x})$ is a compact linear operator, cf. [12, Thm. 1.5.1], thus $G'(\bar{x})(y - x^k) \to G'(\bar{x})(y - \bar{x})$. Inserting $y := \bar{x}$ into (5.3) yields $\liminf_{k \to \infty} \langle f'(x^k), \bar{x} - x^k \rangle \geq 0$. Hence, by pseudomonotonicity, we obtain that, for all $y \in C$,

$$\langle f'(\bar{x}), y - \bar{x} \rangle + \langle \bar{\lambda}, G'(\bar{x})(y - \bar{x}) \rangle \geq \limsup_{k \to \infty} \big[\langle f'(x^k), y - x^k \rangle + \langle \lambda^k, G'(x^k)(y - x^k) \rangle\big] \geq 0.$$

But this means that $-\mathcal{L}(\bar{x}, \bar{\lambda}) \in \mathcal{N}_C(\bar{x})$. Hence, $(\bar{x}, \bar{\lambda})$ is a KKT point of (P_Y). $\square$

Observe that the sequence $\{\lambda^k\}$ is only bounded in Y^* and not necessarily in H. If the extended RCQ holds with respect to the transformed constraint $G(x) \in \mathcal{K}$ (instead of the original condition $G(x) \in K$), then the result remains true with Y^* replaced by H. However, this assumption is too restrictive for many applications, in particular those where (P_Y) is regular (in the constraint qualification sense) with respect to the original space Y, but not with respect to the larger space H.

In the context of optimality properties, it is worthwhile to briefly discuss the case of bounded penalty parameters. This is particularly interesting because any assertion made under this assumption is a *necessary* condition for the boundedness of $\{\rho_k\}$. It turns out that no constraint qualifications are needed in the bounded case, and the algorithm produces a Lagrange multiplier in H.

Corollary 5.6. *Let $\{(x^k, \lambda^k)\}$ be generated by Algorithm 3.3, let Assumption 5.1 hold, and let $\bar{x}$ be a weak limit point of $\{x^k\}$. If $\{\rho_k\}$ remains bounded, then $\{\lambda^k\}$ has a bounded subsequence in H, and $\bar{x}$ satisfies the KKT conditions of (P_H) with a multiplier in H.*

Proof. By (5.2), the sequence $\{(x^k, \lambda^k)\}$ is an asymptotic KKT sequence for (P_Y). Now, let $x^{k+1} \rightharpoonup_I \bar{x}$ on some subset $I \subseteq \mathbb{N}$, and assume that $\{\rho_k\}$ remains bounded. By arguing as in the proof of Lemma 5.2, it follows that $\bar{x} \in \Phi$. Moreover, by the definition

$$\lambda^{k+1} = \rho_k \left[G(x^{k+1}) + \frac{w^k}{\rho_k} - P_{\mathcal{K}}\left(G(x^{k+1}) + \frac{w^k}{\rho_k} \right) \right]$$

of λ^{k+1}, and the boundedness of all the involved quantities, the sequence $\{\lambda^{k+1}\}_{k \in I}$ is bounded in H. Thus, this sequence admits a weak limit point in H, and this point is a Lagrange multiplier in $\bar{x}$ by (5.2). □

The above result implies that $\{\rho_k\}$ can only remain bounded if (P_Y) admits a multiplier in H.

6 Second-Order Sufficient Conditions and Quadratic Growth

In the finite dimensional context the second order sufficient condition (SOSC) is well known to yield nice local properties such as local quadratic growth. This section uses of an infinite-dimensional analogue of SOSC to derive an extended quadratic growth condition which forms the basis of suitable strong convergence results. The particular definition of SOSC that we use is basically taken from [7], with a mild modification to account for the additional constraint set C.

It should be noted that, as with constraint qualifications and KKT conditions, second-order conditions for (P_Y) can be formulated either with respect to Y or H. In this section, to avoid unnecessary notational overhead, we will simply formulate the second-order condition and its consequences with respect to Y. The results below all remain true when Y is replaced by H (note that the choice $Y := H$ is even admissible in our framework).

Let $(\bar{x}, \bar{\lambda}) \in X \times Y^*$ be a KKT point of (P_Y). Throughout this section, we assume that f and G are continuously differentiable in a neighborhood of $\bar{x}$, and twice differentiable in $\bar{x}$. Consider, for $\eta > 0$, the *extended critical cone*

$$\mathcal{C}_\eta(\bar{x}) := \left\{ d \in \mathcal{T}_C(\bar{x}) : \begin{array}{l} f'(\bar{x})d \leq \eta \|d\|_X, \\ \operatorname{dist}(G'(\bar{x})d, \mathcal{T}_K(G(\bar{x}))) \leq \eta \|d\|_X \end{array} \right\}. \tag{6.1}$$

Note that $\mathcal{C}_\eta$ depends on $\bar{x}$ only. The following is the general form of second-order sufficient conditions which we will use throughout this section.

Definition 6.1 (Second-order sufficient condition)**.** We say that the *second-order sufficient condition (SOSC)* holds in a KKT point $(\bar{x}, \bar{\lambda}) \in X \times Y^*$ of (P_Y) if there are $\eta, c > 0$ such that

$$\mathcal{L}''(\bar{x}, \bar{\lambda})(d, d) \geq c\|d\|_X^2 \quad \text{for all } d \in \mathcal{C}_\eta(\bar{x}).$$

As mentioned before, the extended critical cone and SOSC can also be formulated with respect to $\mathcal{K}$ and H for KKT pairs $(\bar{x}, \bar{y}) \in X \times H$.

The above should be considered the "basic" second order condition which can be stated without any assumptions on the specific structure of (P_Y). For many problem classes, it is possible to state more refined second-order conditions which are either equivalent to Definition 6.1 or turn out to have similar implications. Some information in this direction can be found, for instance, in [7, Section 3.3].

One of the most important consequences of second-order conditions is the local quadratic growth of the objective function on the feasible set, i.e., the existence of $c > 0$ such that $f(x) \geq f(\bar{x}) + c\|x - \bar{x}\|_X^2$ for all $x \in \Phi$ near $\bar{x}$, see, for instance, [7, Thm. 3.63]. Here, we will prove a slightly stronger version of this statement with the aim of applying it to the augmented Lagrangian method. In this context, it will be essential to discuss the impact of SOSC on sequences of points $\{x^k\}$ which are not necessarily feasible but satisfy some kind of asymptotic feasibility, e.g., of the form $d_K(G(x^k)) \to 0$. It turns out that the quadratic growth condition can be extended to such points.

For the statement of this result, we use the Landau symbol $a_k = o(b_k)$ for nonnegative real sequences $\{a_k\}$ and $\{b_k\}$, which means that $a_k \leq z_k b_k$ for some null sequence $\{z_k\}$. The sequences $\{a_k\}$ and $\{b_k\}$ themselves are not required to converge to zero.

Theorem 6.2 (Extended quadratic growth)**.** *Let SOSC hold in a KKT point $(\bar{x}, \bar{\lambda}) \in X \times Y^*$ of (P_Y). Then there are $r, c > 0$ such that, for every sequence $\{x^k\} \subseteq B_r(\bar{x}) \cap C$ with $d_K(G(x^k)) = o(\|x^k - \bar{x}\|_X)$, we have*

$$\liminf_{k\to\infty} \left[f(x^k) - f(\bar{x}) - c\|x^k - \bar{x}\|_X^2 \right] \geq 0. \tag{6.2}$$

Proof. Let $\eta, \bar{c} > 0$ be the constants from SOSC and choose r small enough so that

$$\left| f(\bar{x} + d) - f(\bar{x}) - f'(\bar{x})d \right| \leq \frac{\eta}{2}\|d\|_X, \tag{6.3}$$

$$\left\| G(\bar{x} + d) - G(\bar{x}) - G'(\bar{x})d \right\|_Y \leq \frac{\eta}{2}\|d\|_X, \tag{6.4}$$

$$\text{and} \quad \left| \mathcal{L}(\bar{x} + d, \bar{\lambda}) - \mathcal{L}(\bar{x}, \bar{\lambda}) - \mathcal{L}'(\bar{x}, \bar{\lambda})d - \frac{1}{2}\mathcal{L}''(\bar{x}, \bar{\lambda})(d, d) \right| \leq \frac{\bar{c}}{4}\|d\|_X^2 \tag{6.5}$$

for all $d \in X$ with $\|d\|_X \leq r$. Furthermore, set

$$c := \min\left\{ \frac{\eta}{2r}, \frac{\bar{c}}{4} \right\}. \tag{6.6}$$

Now, let $\{x^k\} \subseteq B_r(\bar{x}) \cap C$ be a sequence with $d_K(G(x^k)) = o(\|x^k - \bar{x}\|_X)$, and set $d^k := x^k - \bar{x}$. Passing onto a subsequence if necessary, we assume that the lim inf in (6.2) is a proper limit. If $f'(\bar{x})d^k > \eta\|d^k\|_X$ for infinitely many k, then by (6.3) and (6.6) we obtain

$$f(x^k) - f(\bar{x}) \geq f'(\bar{x})d^k - \frac{\eta}{2}\|d^k\|_X \geq \frac{\eta}{2}\|d^k\|_X \geq c\|d^k\|_X^2$$

for all these k, which implies (6.2). We now consider the case where $f'(\bar{x})d^k \le \eta\|d^k\|_X$ for all but finitely many k. From (6.4), the fact that $K - G(\bar{x}) \subseteq \mathcal{T}_K(G(\bar{x}))$, and $d_K(G(x^k)) = o(\|d^k\|_X)$, it is easy to deduce that

$$\operatorname{dist}\big(G'(\bar{x})d^k, \mathcal{T}_K(G(\bar{x}))\big) \le \operatorname{dist}\big(G(\bar{x}) + G'(\bar{x})d^k, K\big) \le \frac{\eta}{2}\|d^k\|_X + o(\|d^k\|_X).$$

Hence, $d^k \in \mathcal{C}_\eta(\bar{x})$ for sufficiently large k. Applying (6.5), (6.6) and SOSC yields

$$\mathcal{L}(x^k, \bar{\lambda}) - \mathcal{L}(\bar{x}, \bar{\lambda}) - \mathcal{L}'(\bar{x}, \bar{\lambda})d^k \ge \frac{\bar{c}}{2}\|d^k\|_X^2 - \frac{\bar{c}}{4}\|d^k\|_X^2 \ge c\|d^k\|_X^2. \tag{6.7}$$

Observe now that $-\mathcal{L}'(\bar{x}, \bar{\lambda})d^k \le 0$ since $(\bar{x}, \bar{\lambda})$ is a KKT point and $d^k \in C - \bar{x}$. Moreover,

$$\mathcal{L}(x^k, \bar{\lambda}) - \mathcal{L}(\bar{x}, \bar{\lambda}) = f(x^k) - f(\bar{x}) + \big\langle \bar{\lambda}, G(x^k) - G(\bar{x}) \big\rangle,$$

and the last term is asymptotically nonpositive since $\bar{\lambda} \in \mathcal{N}_K(G(\bar{x}))$. Inserting this into (6.7), we obtain $f(x^k) - f(\bar{x}) \ge c\|d^k\|_X^2 + o(1)$, and the result follows. □

The ordinary quadratic growth condition can be seen as a direct corollary of the above theorem.

We now give a second consequence of Theorem 6.2 which will be particularly useful for later results. The main idea is that we can use the theorem to give a sufficient condition for a sequence of asymptotically feasible points to converge to $\bar{x}$.

Corollary 6.3. *Let $(\bar{x}, \bar{\lambda}) \in X \times Y^*$ be a KKT point of (P_Y) satisfying SOSC. Then there exists $r > 0$ such that, whenever $\{x^k\} \subseteq B_r(\bar{x}) \cap C$ is a sequence with $d_K(G(x^k)) \to 0$ and $\limsup_{k\to\infty} f(x^k) \le f(\bar{x})$, then $x^k \to \bar{x}$ (strongly) in X.*

Proof. Let $r, c > 0$ be as in Theorem 6.2 and $\{x^k\} \subseteq B_r(\bar{x}) \cap C$ a sequence with the stated properties. Assume that $\{x^k\}$ does not converge to $\bar{x}$. Passing onto a subsequence if necessary, we may assume that $\|x^k - \bar{x}\|_X \ge \varepsilon$ for all k and some $\varepsilon > 0$. Then $d_K(G(x^k)) = o(\|x^k - \bar{x}\|_X)$ holds trivially; hence, by Theorem 6.2, we obtain

$$0 \le \liminf_{k\to\infty} \big[f(x^k) - f(\bar{x}) - c\|x^k - \bar{x}\|_X^2\big] \le -c \limsup_{k\to\infty} \|x^k - \bar{x}\|_X^2,$$

where we used the fact that $\limsup_{k\to\infty} f(x^k) \le f(\bar{x})$ by assumption. It follows that $\|x^k - \bar{x}\|_X \to 0$, which is the desired contradiction. □

Note that the above result uses the distance function d_K with respect to $K \subseteq Y$ (and therefore depends on the norm $\|\cdot\|_Y$). In our setting (P_H), if the second-order condition is taken with respect to H, then this becomes the distance to $\mathcal{K} \subseteq H$ which depends on $\|\cdot\|_H$.

7 Existence of Local Minimizers and Strong Convergence of the Algorithm

We now consider the subproblems from Algorithm 3.3, the existence of solutions thereof, and the closely linked strong convergence of the method. A recent analysis in [14] for finite-dimensional nonlinear programming showed that the subproblems generated by augmented Lagrangian methods admit minimizers if the underlying optimization problem satisfies the second-order sufficient condition (SOSC) and the primal-dual iterates (x^k, λ^k) are close to a KKT point $(\bar{x}, \bar{\lambda})$.

Here, we extend these results to our general setting from (P_Y), (P_H) and show the existence of minimizers using only the proximity of x^k to $\bar{x}$, whereas no assumption regarding the proximity of the multipliers λ^k is required. Closely linked to the existence of local minimizers is also the local (and *strong*) convergence of Algorithm 3.3.

The basic approach to the existence of local minimizers is the following. Let $r > 0$ be a sufficiently small radius, $B \subseteq H$ a bounded set, and consider, for $\rho > 0$ and $w \in B$, the "localized" problem

$$\underset{x \in X}{\text{minimize}}\ \mathcal{L}_\rho(x, w) \quad \text{subject to} \quad x \in B_r(\bar{x}) \cap C. \tag{7.1}$$

Under suitable assumptions, this problem admits minimizers (or approximate minimizers) in $B_r(\bar{x}) \cap C$. If we can now show that, for sufficiently large ρ, these minimizers actually lie in the interior of $B_r(\bar{x})$, then the spherical constraint in (7.1) is superfluous and we obtain local minimizers of $\mathcal{L}_\rho(\cdot, w)$ subject to $x \in C$.

Note that, for the existence of exact minimizers of (7.1), one in general needs suitable weak compactness and lower semicontinuity assumptions; however, the existence of *approximate* minimizers follows trivially if $r > 0$ is sufficiently small, since in that case the objective in (7.1) is bounded from below on $B_r(\bar{x}) \cap C$ by continuity.

The above localization property can equivalently, and more conveniently, be stated in terms of sequences. We need to show that, whenever $\{w^k\} \subseteq B$ is an arbitrary (bounded) sequence, $\rho_k \to \infty$, and, for all k, y^{k+1} is an (approximate) solution of

$$\underset{x \in X}{\text{minimize}}\ \mathcal{L}_{\rho_k}(x, w^k) \quad \text{subject to} \quad x \in B_r(\bar{x}) \cap C, \tag{7.2}$$

then $\|y^{k+1} - \bar{x}\|_X < r$ for all k sufficiently large. Indeed, we will show that any such sequence converges strongly to $\bar{x}$, and the existence of local minimizers of the augmented Lagrangian subproblems follows directly by the above arguments.

To prove the convergence of minimizers of (7.2) to $\bar{x}$, we will make use of Corollary 6.3, which is a consequence of the second-order sufficient condition. This result guarantees the convergence $y^{k+1} \to \bar{x}$ if we are able to show that $d_\mathcal{K}(G(y^{k+1})) \to 0$ and $\limsup_{k\to\infty} f(y^{k+1}) \le f(\bar{x})$ as $k \to \infty$.

Lemma 7.1. *Assume that there is a KKT point $(\bar{x}, \bar{\lambda}) \in X \times H$ of (P_H) which satisfies the SOSC from Definition 6.1 with respect to H. Then there is a radius $r > 0$ such that the following holds: whenever $\{w^k\} \subseteq H$ is a bounded sequence, $\rho_k \to \infty$, $\varepsilon_k \downarrow 0$, and, for all k, y^{k+1} is an ε_{k+1}-minimizer of (7.2), then $y^{k+1} \to \bar{x}$.*

Proof. Let $r > 0$ be as in Corollary 6.3. Shrinking r if necessary, we may assume that f is bounded on $B_r(\bar{x})$. In particular, it follows that $\mathcal{L}_\rho(\cdot, w)$ is bounded from below on $B_r(\bar{x})$ for all $\rho > 0$ and $w \in B$.

Now, let $\{y^{k+1}\}$ be as specified. Then the ε_{k+1}-minimality of y^{k+1} and Lemma 3.2 yield

$$f(y^{k+1}) + \frac{\rho_k}{2} d_\mathcal{K}^2\left(G(y^{k+1}) + \frac{w^k}{\rho_k}\right) - \frac{\|w^k\|_H^2}{2\rho_k} \le \mathcal{L}_{\rho_k}(\bar{x}, w^k) + \varepsilon_{k+1} \le f(\bar{x}) + \varepsilon_{k+1} \tag{7.3}$$

for all k. Dividing by ρ_k and using the boundedness of $\{w^k\}$ and $\{f(y^{k+1})\}$, it follows that $d_\mathcal{K}(G(y^{k+1}) + w^k/\rho_k) \to 0$ and thus $d_\mathcal{K}(G(y^{k+1})) \to 0$. Moreover, since $\rho_k \to \infty$, we also obtain from (7.3) that $\limsup_{k\to\infty} f(y^{k+1}) \le f(\bar{x})$. Hence, the desired convergence follows from Corollary 6.3. □

Lemma 7.1 implies that the augmented Lagrangian subproblem admits approximate local minimizers in a neighborhood of $\bar{x}$, provided that ρ is large enough and ε sufficiently small.

Since practical algorithms typically compute stationary points of the subproblems, it is important to clarify whether such points can also be found, at least in an approximate sense. This is indeed true and follows from a careful application of Ekeland's variational principle (Proposition 2.4).

Theorem 7.2. *Let SOSC hold with respect to the space H, and let $B \subseteq H$ be a bounded set. Then there are $\bar{\rho}, \bar{\varepsilon}, r > 0$ such that, for all $w \in B$, $\rho \geq \bar{\rho}$, and $\varepsilon \in (0, \bar{\varepsilon})$, there is a point $x = x_{\rho,\varepsilon}(w) \in C$ with $\|x - \bar{x}\|_X < r$ and the following properties:*

(i) *x is an ε-minimizer of $\mathcal{L}_\rho(\cdot, w)$ on $B_r(\bar{x}) \cap C$,*

(ii) *x satisfies $\operatorname{dist}(-\mathcal{L}'_\rho(x, w), \mathcal{N}_C(x)) \leq \varepsilon^{1/2}$, and*

(iii) *$x = x_{\rho,\varepsilon}(w) \to \bar{x}$ uniformly on B as $\rho \to \infty$ and $\varepsilon \to 0$.*

Proof. Let $r > 0$ be as in Lemma 7.1. For $\rho > 0$ and $w \in B$, consider the problem

$$\underset{x \in X}{\text{minimize}} \ \mathcal{L}_\rho(x, w) \quad \text{subject to} \quad x \in C_r := B_r(\bar{x}) \cap C.$$

Observe that the constraint $x \in B_r(\bar{x}) \cap C$ trivially satisfies the Robinson constraint qualification. Hence, by Ekeland's variational principle (Proposition 2.4), there are points $x = x_{\rho,\varepsilon}(w)$ such that x satisfies (i) and, in addition, $\operatorname{dist}(-\mathcal{L}'_\rho(x, w), \mathcal{N}_{C_r}(x)) \leq \varepsilon^{1/2}$. By Lemma 7.1, it follows that $x_{\rho,\varepsilon} \to \bar{x}$ uniformly on B as $\rho \to \infty$ and $\varepsilon \to 0$. Hence, there are $\bar{\rho}, \bar{\varepsilon} > 0$ such that $\|x_{\rho,\varepsilon}(w) - \bar{x}\|_X < r$ for all $\rho \geq \bar{\rho}$, $\varepsilon \in (0, \bar{\varepsilon})$, and $w \in B$. But $\mathcal{N}_{C_r}(x) = \mathcal{N}_C(x)$ whenever $x \in C$ and $\|x - \bar{x}\|_X < r$. Hence, the result follows. □

If X is reflexive and the augmented Lagrangian $\mathcal{L}_\rho(\cdot, w)$ is weakly sequentially lsc, then the assertions of the above theorem remain valid if we replace the ε-minimizers by exact minimizers. In this case, we obtain points $x = x_\rho(w)$ which satisfy (i) and (ii) with $\varepsilon := 0$ and which converge to $\bar{x}$ uniformly on B as $\rho \to \infty$.

It should be noted that the above result is not specifically tied to the augmented Lagrangian method (Algorithm 3.3). When applied to the algorithm, the result has two important consequences: first, it shows that the augmented subproblems admit approximate local solutions if ρ is sufficiently large, and secondly, it follows that these solutions converge to $\bar{x}$ if $\rho_k \to \infty$. Indeed, in the setting of Algorithm 3.3, this convergence also holds if $\{\rho_k\}$ is bounded, so that Theorem 7.2 effectively yields a strong convergence result for the algorithm. The details are contained in the following corollary.

Corollary 7.3. *Let the assumptions of Theorem 7.2 be satisfied. Assume that, for sufficiently large k, Algorithm 3.3 chooses x^k as one of the approximate minimizers from Theorem 7.2, or any other ε_k-minimizer of (7.1) with $r > 0$ and $\varepsilon_k \downarrow 0$. Then $x^k \to \bar{x}$ strongly.*

Proof. If $\rho_k \to \infty$, then the desired convergence follows from Lemma 7.1. On the other hand, if $\{\rho_k\}$ remains bounded, then we can argue as in Lemma 4.2 to see that $d_\mathcal{K}(G(x^k)) \to 0$. Moreover, Theorem 4.3 yields $\limsup_{k\to\infty} f(x^k) \leq f(x)$, so that the convergence $x^k \to \bar{x}$ finally follows from Corollary 6.3. □

In the remaining part of this section, we briefly analyze the case where the second-order condition holds not with respect to H but Y. This is the more natural condition in some application problems (since (P_Y) is originally formulated in Y), but it is a little more intricate to discuss in the context of augmented Lagrangian methods since H, not Y, plays the role of the augmentation space. However, by using complete continuity techniques, it is nevertheless possible to deduce an analogue of Theorem 7.2 for the case of SOSC with respect to Y.

Proposition 7.4. *Let $(\bar{x}, \bar{\lambda}) \in X \times Y^*$ be a KKT point of (P_Y) which satisfies SOSC with respect to the space Y, and $B \subseteq H$ a bounded set. Assume that*

(i) *the space X is reflexive,*

(ii) *f is weakly sequentially lsc on X, and*

(iii) *G is completely continuous from X into Y.*

Then there are $\bar{\rho}, r > 0$ such that, for every $w \in B$ and $\rho \geq \bar{\rho}$, the problem $\min_{x\in C} \mathcal{L}_\rho(x, w)$ admits a local minimizer $x = x_\rho(w)$ in $B_r(\bar{x}) \cap C$, and $x_\rho \to \bar{x}$ uniformly on B as $\rho \to \infty$.

Proof. We argue similarly to the proof of Lemma 7.1 and Theorem 7.2. Let $r > 0$ be small enough so that $\bar{x}$ is a strict minimizer of f on $B_r(\bar{x}) \cap \Phi$, and such that the assertions of Corollary 6.3 hold. Let $\{w^k\} \subseteq B$ and $\rho_k \to \infty$ be arbitrary sequences and, for each k, let $y^{k+1} \in B_r(\bar{x}) \cap C$ be a (global) minimizer of (7.2). Note that y^{k+1} exists since $B_r(\bar{x})$ is weakly compact and $\mathcal{L}_\rho(\cdot, w)$ is weakly sequentially lsc under the given assumptions.

We need to show that $y^{k+1} \to \bar{x}$ as $k \to \infty$. By SOSC, $\bar{x}$ is a strict local minimizer of (P_Y); hence, a subsequence-subsequence argument shows that $y^{k+1} \rightharpoonup \bar{x}$. Now, Theorem 4.3 and the weak sequential lower semi-continuity of f yield $f(y^{k+1}) \to f(\bar{x})$. Since G is completely continuous, we obtain $G(y^{k+1}) \to G(\bar{x})$, which implies that $d_K(G(y^{k+1})) \to 0$ (note the K instead of $\mathcal{K}$). Thus, Corollary 6.3 yields $y^{k+1} \to \bar{x}$, and the proof is complete. □

The above result is relevant, for instance, for optimal control problems involving partial differential equations and state constraints, and can also be seen as generalization of a related result from [30].

Similarly to before, it is possible to formulate the previous results in terms of Algorithm 3.3 and obtain a strong convergence result.

Corollary 7.5. *Let the assumptions of Proposition 7.4 be satisfied. Assume that, for sufficiently large k, Algorithm 3.3 chooses x^k as one of the minimizers from Proposition 7.4, or any other minimizer of (7.1) with $r > 0$. Then $x^k \to \bar{x}$ strongly.*

Proof. Let $r > 0$ be as in Proposition 7.4 and recall that, in particular, $\bar{x}$ is a strict minimizer of f on $B_r(\bar{x}) \cap \Phi$. If $\rho_k \to \infty$, then the desired convergence follows as in Proposition 7.4. On the other hand, if $\{\rho_k\}$ remains bounded, then we can argue as in Theorem 4.3 to see that every weak limit point of $\{x^k\}$ is a minimizer of f on $B_r(\bar{x}) \cap \Phi$. Since $\bar{x}$ is the strict minimizer of f on this set, a subsequence-subsequence argument yields $x^k \rightharpoonup \bar{x}$. As in Proposition 7.4, this implies $f(x^k) \to f(\bar{x})$ and $G(x^k) \to G(\bar{x})$. Hence, $d_K(G(x^k)) \to 0$, and Corollary 6.3 yields $x^k \to \bar{x}$. □

8 Applications

We now apply the augmented Lagrangian method to some example problems. Note that our algorithm is a generalization of the one from [29]; hence, the examples given there apply to our framework as well. This section is therefore dedicated to applications which are covered by our theory but not by [29]. In particular, we will consider problems where the additional constraint $x \in C$ cannot be eliminated, and we will give an example which demonstrates the consequences of second-order conditions established in Sections 6 and 7.

All our examples follow the general pattern that X, Y, H are (infinite-dimensional) function spaces on some bounded domain $\Omega \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$. Given such a function space, we write Y_+ (Y_-) for the positive (negative) cone in Y. The sets K and $\mathcal{K}$ often model inequality constraints which involve these cones. In each of the subsections, we first give a general overview about the problem in question and then present some numerical results on a simple domain Ω.

In practice, Algorithm 3.3 is then applied to a (finite-dimensional) discretization of the corresponding problem. Hence, we implemented the algorithm for finite-dimensional problems. The implementation was done in MATLAB and uses the parameters

$$\lambda^0 := 0, \quad B := [0, 10^6]^m, \quad \rho_0 := 1, \quad \gamma := 5, \quad \tau := 0.9$$

(where m is the appropriate dimension), together with a problem-dependent starting point x^0. The overall stopping criterion which we use for our algorithm is given by

$$\|\nabla f(x) + \nabla G(x)\lambda\|_\infty \le 10^{-4} \quad \text{and} \quad \|\min\{-G(x), \lambda\}\|_\infty \le 10^{-4}. \tag{8.1}$$

If further equality constraints $h(x) = 0$ are present, we also check $\|h(x)\|_\infty \le 10^{-4}$. Thus we verify an inexact KKT condition for the discretized problem. To make the termination criterion independent from the discretization fidelity, we use the infinity norm. Furthermore, in each outer iteration, we solve the corresponding subproblem in Step 2 by computing a point x^{k+1} which satisfies

$$\left\|\mathcal{L}'_{\rho_k}(x^{k+1}, w^k)\right\|_\infty \le 10^{-6}.$$

Hence if the solver for the subproblem does not fail, the first condition of (8.1) is always true, cf. (5.1). We now turn to our test problems.

8.1 State-Constrained Optimal Control Problems

We now turn to a rather prominent class of optimization problems with partial differential equation (PDE) constraints. Let $\Omega \subseteq \mathbb{R}^n$, $n \in \{2, 3\}$, be a bounded Lipschitz domain. We consider the optimal control problem given by the functional

$$J(y, u) := \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}, \tag{8.2}$$

where $y \in H^1_0(\Omega) \cap C(\overline{\Omega})$ and $u \in L^2(\Omega)$, together with the PDE, state and control constraints

$$-\Delta y + d(y) = u \quad \text{in } H^{-1}(\Omega), \qquad y \ge y_c \quad \text{in } \Omega, \quad \text{and} \quad u_a \le u \le u_b \quad \text{in } \Omega. \tag{8.3}$$

Here, $\Delta : H^1_0(\Omega) \to H^{-1}(\Omega)$ is the Laplace operator, $\alpha > 0$ is a regularization parameter, and $u_a, u_b \in L^\infty(\Omega)$, $y_d \in L^2(\Omega)$, $y_c \in C(\overline{\Omega})$, $y_c \le 0$ on $\partial\Omega$, are given functions. The nonlinearity d in the elliptic equation is induced by a function $d : \mathbb{R} \to \mathbb{R}$, which is assumed to be sufficiently regular and monotonically increasing.

We now follow a standard technique in PDE-constrained optimization and eliminate the variable y. By elliptic regularity results, the PDE in (8.3) admits, for any given $u \in L^2(\Omega)$, a unique solution $y = S(u) \in H^1_0(\Omega) \cap C(\overline{\Omega})$. The resulting mapping $S : L^2(\Omega) \to H^1_0(\Omega) \cap C(\overline{\Omega})$

is completely continuous and continuously differentiable, see [40], and thus we can restate (8.2) as the smooth minimization problem

$$\underset{u\in L^2(\Omega)}{\text{minimize}}\ \bar{J}(u) := J(S(u),u) \quad \text{subject to} \quad S(u) \geq y_c. \tag{8.4}$$

This is usually called the *reduced form* of the problem. In the context of our general framework (P_Y), we have the data

$$X := L^2(\Omega), \quad C := \{u \in L^2(\Omega) \mid u_a \leq u \leq u_b\}, \quad Y := C(\overline{\Omega}), \quad G(u) := S(u) - y_c,$$
$$K := Y_+, \quad H := L^2(\Omega), \quad \mathcal{K} := H_+.$$

We can now apply the augmented Lagrangian method to eliminate the constraint $S(u) \geq y_c$, thus obtaining a sequence of penalized problems. For efficiency reasons, we tackle these problems by reintroducing the state variable y and writing the problems as

$$\begin{aligned} &\underset{y\in H_0^1(\Omega),\ u\in L^2(\Omega)}{\text{minimize}} && J(y,u) + \tfrac{\rho_k}{2}\left\|\max\Big(0, y_c - y + \tfrac{w^k}{\rho_k}\Big)\right\|^2 \\ &\text{subject to} && -\Delta y = u \quad \text{and} \quad u_a \leq u \leq u_b. \end{aligned} \tag{8.5}$$

We now discuss the applicability of the convergence results. Due to the nonconvexity of the problem, we can only expect to compute stationary points of the augmented subproblems. This makes the theory from Section 5 a natural candidate for the present situation. To apply the main results from that section, we need to verify the following properties:

- The mapping $G' : X \to L(X,Y)$ is completely continuous. In the present setting, since $X = L^2(\Omega)$ is reflexive and $G'(u) \in L(X,Y)$ is compact for all u (by [12, Thm. 1.5.1]), this is equivalent to the following property: whenever $u^k \rightharpoonup u$ and $h^k \rightharpoonup h$ in X, then $G'(u^k)h^k \to G'(u)h$ strongly in Y. A proof of this statement (for the Neumann case) can be found in [30, Lem. 4.7].
- The mapping $\bar{J}' : X \to X^*$ is bounded and pseudomonotone. Note that $\bar{J}'(u) = S'(u)^*(S(u) - y_d) + \alpha u$ for all $u \in X$. As seen above, S and S' are completely continuous, hence bounded (since X is reflexive). This implies the boundedness of $\bar{J}'$. The pseudomonotonicity follows from the fact that the first term in $\bar{J}'$ is completely continuous and the second term is monotone (see Lemma 2.6).

Moreover, we need the Robinson constraint qualification (RCQ) to hold at feasible points of (8.4). For this, the following observation is helpful. Since the set K has a nonempty interior, RCQ is equivalent to the linearized Slater condition

$$\exists \hat{u} \in X : \quad G(u) + G'(u)(\hat{u} - u) \in \operatorname{int}(K),$$

which is a standard assumption in the optimal control context. If the linearized Slater condition holds, then we obtain RCQ (and its extended version from Definition 2.3), see [40, p. 332]. This implies that the main results from Section 5 are applicable. In particular, we obtain from Theorem 5.5 that every weak limit point $\bar{u}$ of the sequence $\{u^k\}$ is a stationary point of (8.4), the corresponding subsequence of $\{\lambda^k\}$ is bounded in $C(\overline{\Omega})^*$, and its weak-* limit points are Lagrange multipliers in $\bar{u}$.

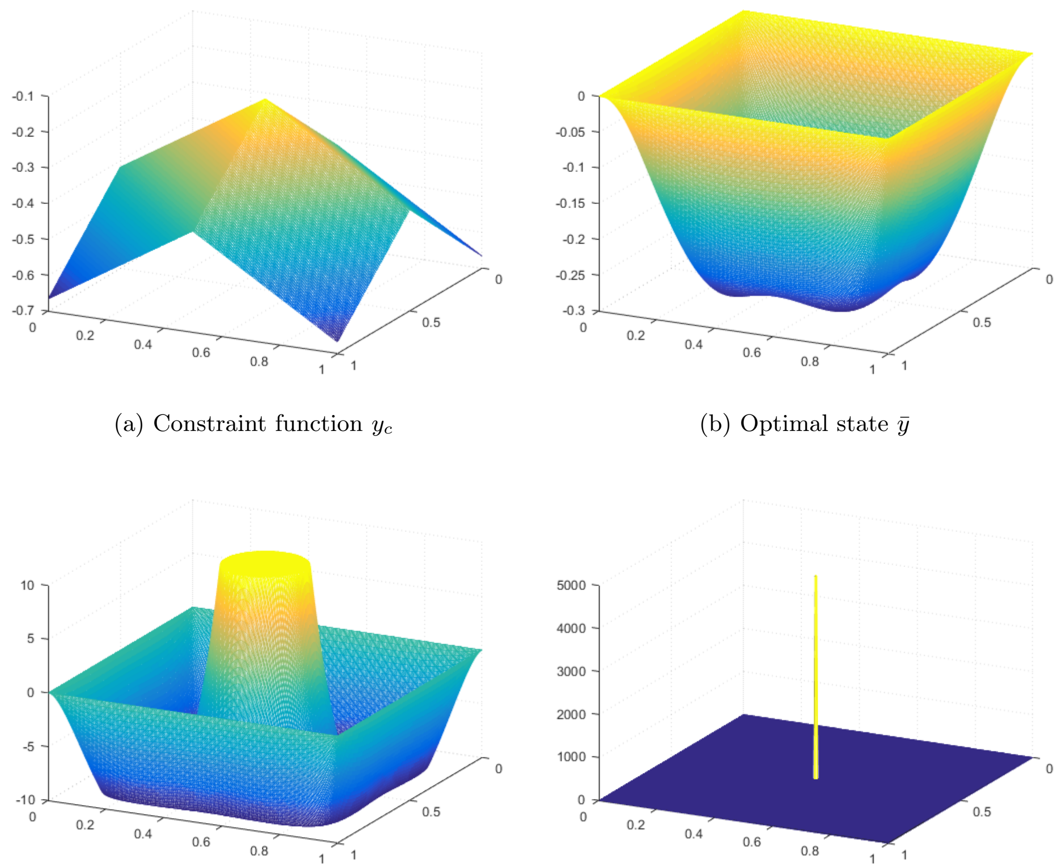
We now turn to numerical results. The following test problem is a small modification of the example presented in [33]. Let $\Omega := (0,1)^2$, $d(y) := y^3$, $\alpha := 10^{-3}$, $u_a = -10$ $u_b = 10$, and

$$y_c(x) := -\frac{2}{3} + \frac{1}{2}\min\{x_1 + x_2, 1 + x_1 - x_2, 1 - x_1 + x_2, 2 - x_1 - x_2\}.$$

n	16	32	64	128	256
outer it.	27	30	38	46	52
inner it.	211	238	289	332	370
final ρ_k	5^4	5^5	5^7	5^8	5^9

Table 1: Numerical results for the optimal control problem from Section 8.1.

Clearly, in this setting, (8.2) and its reformulation (8.4) are nonconvex problems. The augmented subproblems are solved by applying the MATLAB function `fmincon`, where the Hessian of the objective is approximated by a generalized second-order derivative, known as *Newton derivative*, in the sense of [21]. Table 1 contains the resulting iteration numbers and final penalty parameters for the augmented Lagrangian method. The method scales well with increasing dimension.



(a) Constraint function y_c

(b) Optimal state $\bar{y}$

(c) Optimal control $\bar{u}$

(d) Lagrange multiplier $\bar{\lambda}$

Figure 1: Numerical results for the optimal control problem from Section 8.1 ($n = 256$).

The state constraint y_c and the results of our method are given in Figure 1. It is interesting to note that the multiplier $\bar{\lambda}$ appears to be much less regular than the optimal control $\bar{u}$ and state $\bar{y}$. This is not surprising because, due to our construction, we have

$$\bar{u} \in L^2(\Omega), \quad \bar{y} \in C(\overline{\Omega}), \quad \text{and} \quad \bar{\lambda} \in C(\overline{\Omega})^*.$$

The latter is well-known to be the space of Radon measures on $\overline{\Omega}$, which is a superset of $L^2(\Omega)$. In fact, the convergence data shows that the (discrete) L^2-norm of $\bar{\lambda}$ grows approximately linearly as n increases, possibly even diverging to $+\infty$, which suggests that the underlying (infinite-dimensional) problem (8.4) does not admit a multiplier in $L^2(\Omega)$ but only in $C(\overline{\Omega})^*$.

8.2 Complementarity Constraints

We now give an example whose main purpose is to illustrate the results related to second-order sufficient conditions; see Sections 6 and 7. Since the theory established there does not assume any constraint qualifications, a prime candidate for a suitable example is an optimization problem with complementarity constraints (MPCC). Such problems are known to be highly irregular even in the finite-dimensional case [19, 32, 39], and the behavior of augmented Lagrangian methods for finite-dimensional MPCCs was analyzed in some detail in [23].

Here, we will consider MPCCs in infinite dimensions; some recent literature on this topic includes [42, 43]. In particular, it is known that the Robinson regularity condition is never satisfied for such problems, and the set of standard Lagrange multipliers in a local minimum may be empty or unbounded.

Now, let $\Omega \subseteq \mathbb{R}^n$, $n \in \mathbb{N}$, be a bounded domain and consider the problem

$$\min\ f(x,y) \quad \text{s.t.} \quad x, y \geq 0,\ (x,y)_{L^2(\Omega)} = 0, \tag{8.6}$$

where $x, y \in L^2(\Omega)$. This problem can be put into our general framework (P_Y), (P_H) by setting

$$X := L^2(\Omega) \times L^2(\Omega), \quad Y := H := L^2(\Omega) \times L^2(\Omega) \times \mathbb{R},$$

and

$$G(x,y) := \big(x, y, (x,y)_{L^2(\Omega)}\big), \quad K := \mathcal{K} := L^2_+(\Omega) \times L^2_+(\Omega) \times \{0\},$$

where $L^2_+(\Omega)$ denotes the nonnegative cone in $L^2(\Omega)$. The Lagrangian of this problem is given by

$$\mathcal{L}(x,y,\lambda) = f(x,y) - (\lambda_1, x)_{L^2(\Omega)} - (\lambda_2, y)_{L^2(\Omega)} + \lambda_3 (x,y)_{L^2(\Omega)},$$

where $\lambda_1, \lambda_2 \in L^2(\Omega)$ and $\lambda_3 \in \mathbb{R}$. The KKT conditions can be written as

$$\begin{aligned} D_x f(\bar{x},\bar{y}) - \lambda_1 + \lambda_3 \bar{y} &= 0, \\ D_y f(\bar{x},\bar{y}) - \lambda_2 + \lambda_3 \bar{x} &= 0, \end{aligned}$$

where $\bar{x}, \bar{y}, \lambda_1, \lambda_2 \geq 0$ in $L^2(\Omega)$ satisfy the complementarity conditions

$$(\bar{x},\bar{y})_{L^2(\Omega)} = (\lambda_1, \bar{x})_{L^2(\Omega)} = (\lambda_2, \bar{y})_{L^2(\Omega)} = 0.$$

We now consider a fairly simple example by setting

$$f(x,y) = \frac{1}{2}\|x-1\|^2 + \frac{1}{2}\|y-2\|^2,$$

where 1 and 2 are understood to be the corresponding constant functions. A straightforward calculation shows that $(\bar{x},\bar{y}) := (0,2)$ is the unique solution of the problem. Moreover, the KKT conditions hold for any multiplier triple satisfying

$$\lambda_1 - 2\lambda_3 = -1 \quad \text{and} \quad \lambda_2 = 0.$$

The simplest such triple is given by $\bar{\lambda} = (0, 0, 1/2)$, and the corresponding Lagrange function is a quadratic functional with

$$\mathcal{L}''(x, y, \bar{\lambda}) = \begin{pmatrix} I & \frac{1}{2}I \\ \frac{1}{2}I & I \end{pmatrix},$$

where I is the identity operator on $L^2(\Omega)$ and the derivative is understood with respect to (x, y). This implies that $(x, y) \mapsto \mathcal{L}(x, y, \bar{\lambda})$ is strongly convex. In particular, SOSC is satisfied for this example, and we expect the augmented Lagrangian method to behave well due to the assertions of Corollary 7.3.

We now give some numerical results which use $\Omega = (0, 1)$ and a discretization of $n \in \mathbb{N}$ points. The resulting iteration numbers are given as follows.

n	16	32	64	128	256
outer it.	3	3	3	3	3
inner it.	13	11	13	12	14
final ρ_k	1	1	1	1	1

As suggested by our theoretical analysis, the augmented Lagrangian method performs quite well on this problem. Note that we do not give an image displaying the results because the solution is simply given by $\bar{x} = 0$, $\bar{y} = 2$. We did, however, check that the algorithm in fact finds the correct solution.

8.3 Optimal Control of the Obstacle Problem

Our final example is based on [41] and incorporates elements from both of the previous examples. Note that this example is highly irregular and its theoretical background in view of the previous convergence theorems (e.g. 4.3, 5.5, 7.3 and 7.5) is rather doubtful. Nevertheless, we have decided to include the example in our presentation because the multiplier-penalty method performs fairly well for this problem.

Recall that the standard obstacle problem [20, 38] has the form

$$\min \ \frac{1}{2}\|\nabla y\|^2_{L^2(\Omega)} - \langle f, y\rangle_{H^{-1}(\Omega), H^1_0(\Omega)} \quad \text{s.t.} \quad y \geq \psi,$$

where $\Omega \subseteq \mathbb{R}^n$ is a bounded domain, $f \in H^{-1}(\Omega)$, $y \in H^1_0(\Omega)$, and $\psi \in H^1(\Omega)$ satisfies $\psi \leq 0$ on $\partial\Omega$. The KKT conditions of this problem can be written tersely as

$$-\Delta y = f + \xi \quad \text{and} \quad 0 \geq \psi - y \perp \xi \geq 0$$

with a suitable multiplier $\xi \in H^{-1}(\Omega)$. We now consider this problem in an optimal control framework where the fixed function f is replaced by $f + u$ with $u \in L^2(\Omega)$. Hence, we obtain lower-level complementarity constraints of the form

$$-\Delta y = u + f + \xi \quad \text{and} \quad 0 \geq \psi - y \perp \xi \geq 0.$$

Finally, choosing a similar objective function as in Section 8.1, we arrive at the optimal control problem

$$\begin{aligned} \min \quad & \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \\ \text{s.t.} \quad & -\Delta y = u + \xi + f, \\ & 0 \geq \psi - y \perp \xi \geq 0 \end{aligned} \tag{8.7}$$

where $(y, u, \xi) \in H_0^1(\Omega) \times L^2(\Omega) \times H^{-1}(\Omega)$ and $y_d \in L^2(\Omega)$.

The above problem was analyzed in depth in [41]. Among other things, it was shown that local minima of (8.7) are strongly stationary under moderate assumptions. However, the problem is still highly irregular in many aspects. For instance, since the problem is an infinite-dimensional MPCC, strong stationarity is weaker than the standard KKT conditions [42]. In particular, it is not clear whether a KKT point exists at all. This issue is somewhat mitigated by the fact that we will eventually consider finite-dimensional discretizations of (8.7) anyway; for these, the aforementioned equivalence of strong stationarity and the KKT conditions is satisfied, and the latter are known to be necessary optimality conditions under certain assumptions [39].

Another issue regarding the minimization problem (8.7) is the violation of the Robinson condition at any point. This follows from the presence of the complementarity condition

$$0 \geq \psi - y \perp \xi \geq 0, \tag{8.8}$$

cf. [42, 43]. As a consequence of this irregularity, we cannot expect the multiplier sequence generated by the multiplier-penalty method to remain bounded.

On the positive side, the objective function in (8.7) is strongly convex with respect to (y, u). Note that ξ is effectively an implicit variable due to the structure of the constraints; hence, we may consider (8.7) as a problem in (y, u) only. The strong convexity of the objective suggests that the problem may satisfy a quadratic growth condition akin to the one from Theorem 6.2, or even the second-order sufficient condition, provided a KKT point exists. However, since the solution of (8.7) is nontrivial to calculate, we cannot analyze these properties in more detail. In our implementation the complementarity constraint (8.8) is augmented in the Lagrange function, while the PDE-constraint is treated in the solver for the subproblem as in (8.5).

We now give some numerical results which use $\Omega = (0, 1)^2$ and a discretization of n^2 ($n \in \mathbb{N}$) points. The problem parameters are given by $f = 0$, $\alpha = 10^{-3}$, the starting point x^0 is the solution of the standard obstacle problem (which we also computed with Algorithm 3.3), and the desired state is given by $y_d := \psi$, i.e., it is the obstacle itself.

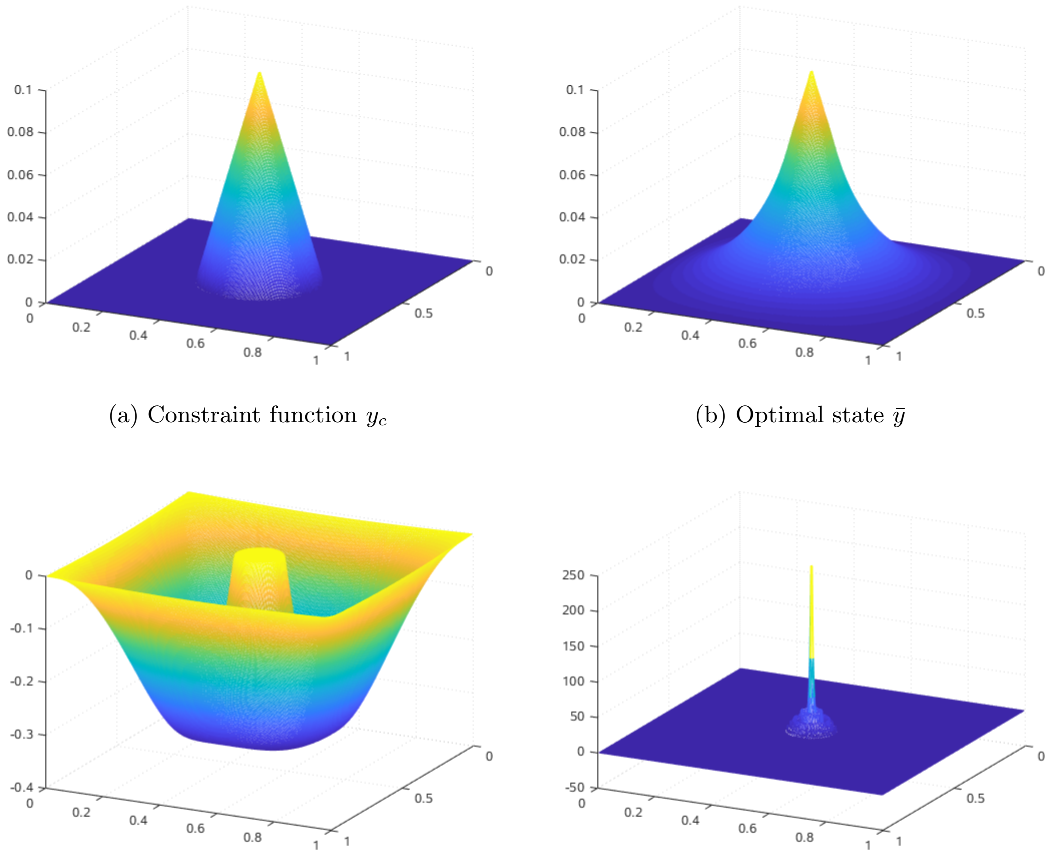
The treatment of the augmented Lagrangian subproblems requires special care because the (discretized) equality constraint $h(x, y) := x^T y$ induces an addend of the form

$$\begin{pmatrix} y \\ x \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix}^T = \begin{pmatrix} yy^T & yx^T \\ xy^T & xx^T \end{pmatrix}$$

in the Hessian matrix of the augmented Lagrangian. This term makes the overall Hessian nearly dense and thus hard to deal with. It is possible to decompose the Hessian into a sparse part and the above rank-one term, and use the well-known Sherman–Morrison formula for the solution of linear equations involving the Hessian, but the tailored nature of this approach rules out most standard Newton-type implementations. For the present problem, we solved this issue by using a semismooth SQP method with a simple trust-region type globalization. Let us stress that, while this method works sufficiently well for our purposes, the resolution of the inner problems is not the main subject of the present paper, and a more detailed discussion of possible methods is outside the scope of our work.

The resulting iteration numbers can be found below.

n	16	32	64	128	256
outer it.	13	17	22	33	35
inner it.	97	733	2732	8484	19007
final ρ_k	5^2	5^4	5^4	5^5	5^5



(a) Constraint function y_c (b) Optimal state $\bar{y}$

(c) Optimal control $\bar{u}$ (d) Optimal multiplier $\bar{\xi}$

Figure 2: Numerical results for the optimal control of the obstacle problem with $n = 256$.

Note that the augmented Lagrangian method is able to solve the obstacle-type control problem with a fairly modest growth of iteration numbers and penalty parameters. The Newtonian method used for the resolution of the subproblems is sufficiently robust and efficient to yield an overall solution of the problem, but incurs more substantial iteration numbers with increasing discretization fidelity. As discussed above, this matter is down to the specific choice of inner algorithm and no indicator of the performance of the outer algorithmic framework.

9 Final Remarks

We have presented an augmented Lagrangian method for the solution of a very general class of minimization problems. This method is in the spirit of recent developments for similar finite-dimensional methods [6] and generalizes a related approach for L^2-type inequality constraints [29]. The numerical results we have presented build upon the earlier work of the authors in [29] and appear quite promising.

A key feature of our method is the ability to solve optimization problems with arbitrary inclusion constraints. This generality allows us to treat various problem classes, ranging from classical nonlinear programming to more specific problems such as semidefinite programming and optimal control. It follows that all the results we have established hold for any of these problem classes.

The convergence analysis we have presented includes, on the one hand, a generalization of the global analysis from [29]. This encompasses results pertaining to global minima of the subproblems as well as stationary points. On the other hand, we have established a novel result implying local convergence of the augmented Lagrangian method under the second-order sufficient optimality condition alone. To the best of our knowledge, this is the first such result for general cone-constrained optimization problems. It also extends a recent result in [14].

We believe that the sum of these contributions makes the augmented Lagrangian method a viable technique for generic minimization problems in Banach spaces. Therefore, we hope that this report will facilitate research in similar directions and contribute to the rich theoretical background of multiplier-penalty methods in both finite and infinite dimensions.

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*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, Cham, second edition, 2017. With a foreword by Hédy Attouch.

[3] M. Bergounioux and K. Kunisch. Augmented Lagrangian techniques for elliptic state constrained optimal control problems. *SIAM J. Control Optim.*, 35(5):1524–1543, 1997.

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

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

[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 A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000.

[8] H. Brezis. Équations et inéquations non linéaires dans les espaces vectoriels en dualité. *Ann. Inst. Fourier (Grenoble)*, 18(fasc. 1):115–175, 1968.

[9] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Universitext. Springer, New York, 2011.

[10] 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.

[11] 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.

[12] S. V. Emelyanov, S. K. Korovin, N. A. Bobylev, and A. V. Bulatov. *Homotopy of extremal problems*, volume 11 of *De Gruyter Series in Nonlinear Analysis and Applications*. Walter de Gruyter & Co., Berlin, 2007. Theory and applications.

[13] D. Fernández and M. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. *Math. Program.*, 125(1, Ser. A):47–73, 2010.

[14] 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.

[15] A. Fischer. Local behavior of an iterative framework for generalized equations with nonisolated solutions. *Math. Program.*, 94(1, Ser. A):91–124, 2002.

[16] M. Fortin and R. Glowinski. *Augmented Lagrangian Methods. Applications to the Numerical Solution of Boundary-Value Problems*, volume 15 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1983. Translated from the French by B. Hunt and D. C. Spicer.

[17] W. W. Hager. Stabilized sequential quadratic programming. *Comput. Optim. Appl.*, 12(1-3):253–273, 1999.

[18] M. Hintermüller and K. Kunisch. Feasible and noninterior path-following in constrained minimization with low multiplier regularity. *SIAM J. Control Optim.*, 45(4):1198–1221, 2006.

[19] T. Hoheisel, C. Kanzow, and A. Schwartz. Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints. *Math. Program.*, 137(1-2, Ser. A):257–288, 2013.

[20] K. Ito and K. Kunisch. An augmented Lagrangian technique for variational inequalities. *Appl. Math. Optim.*, 21(3):223–241, 1990.

[21] K. Ito and K. Kunisch. *Lagrange Multiplier Approach to Variational Problems and Applications*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

[22] A. Iusem and R. Gárciga Otero. Augmented Lagrangian methods for cone-constrained convex optimization in Banach spaces. *J. Nonlinear Convex Anal.*, 3(2):155–176, 2002.

[23] A. F. Izmailov, M. V. Solodov, and E. I. Uskov. Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. *SIAM J. Optim.*, 22(4):1579–1606, 2012.

[24] C. Kanzow. On the multiplier-penalty-approach for quasi-variational inequalities. *Math. Program.*, 160(1-2, Ser. A):33–63, 2016.

[25] C. Kanzow, V. Karl, D. Steck, and D. Wachsmuth. The multiplier-penalty method for generalized Nash equilibrium problems in Banach spaces. *SIAM J. Optim.*, to appear.

[26] C. Kanzow and D. Steck. An example comparing the standard and safeguarded augmented Lagrangian methods. *Oper. Res. Lett.*, 45(6):598–603, 2017.

[27] C. Kanzow and D. Steck. On error bounds and multiplier methods for variational problems in Banach spaces. *SIAM J. Control Optim.*, 56(3):1716–1738, 2018.

[28] C. Kanzow and D. Steck. Quasi-variational inequalities in Banach spaces: theory and augmented Lagrangian methods. *ArXiv e-prints*, December 2018. ArXiv:1810.00406.

[29] C. Kanzow, D. Steck, and D. Wachsmuth. An augmented Lagrangian method for optimization problems in Banach spaces. *SIAM J. Control Optim.*, 56(1):272–291, 2018.

[30] V. Karl, I. Neitzel, and D. Wachsmuth. A Lagrange multiplier method for semilinear elliptic state constrained optimal control problems. *ArXiv e-prints*, June 2018. ArXiv:1806.08124.

[31] V. Karl and D. Wachsmuth. An augmented Lagrange method for elliptic state constrained optimal control problems. *Comput. Optim. Appl.*, 69(3):857–880, 2018.

[32] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, 1996.

[33] I. Neitzel, J. Pfefferer, and A. Rösch. Finite element discretization of state-constrained elliptic optimal control problems with semilinear state equation. *SIAM J. Control Optim.*, 53(2):874–904, 2015.

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

[35] K. J. Palmer. On the complete continuity of differentiable mappings. *J. Austral. Math. Soc.*, 9:441–444, 1969.

[36] 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.

[37] R. T. Rockafellar. Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM J. Control*, 12:268–285, 1974.

[38] J.-F. Rodrigues. *Obstacle Problems in Mathematical Physics.* North-Holland Publishing Co., Amsterdam, 1987.

[39] H. Scheel and S. Scholtes. Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. *Math. Oper. Res.*, 25(1):1–22, 2000.

[40] F. Tröltzsch. *Optimal Control of Partial Differential Equations.* American Mathematical Society, Providence, RI, 2010.

[41] G. Wachsmuth. Strong stationarity for optimal control of the obstacle problem with control constraints. *SIAM J. Optim.*, 24(4):1914–1932, 2014.

[42] G. Wachsmuth. Mathematical programs with complementarity constraints in Banach spaces. *J. Optim. Theory Appl.*, 166(2):480–507, 2015.

[43] G. Wachsmuth. Strong Stationarity for Optimization Problems with Complementarity Constraints in Absence of Polyhedricity. *Set-Valued Var. Anal.*, 25(1):133–175, 2017.

[44] A. P. Wierzbicki and S. Kurcyusz. Projection on a cone, penalty functionals and duality theory for problems with inequaltity constraints in Hilbert space. *SIAM J. Control Optimization*, 15(1):25–56, 1977.

[45] S. J. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Comput. Optim. Appl.*, 11(3):253–275, 1998.

[46] S. J. Wright. Modifying SQP for degenerate problems. *SIAM J. Optim.*, 13(2):470–497, 2002.

[47] E. Zeidler. *Nonlinear Functional Analysis and its Applications. II/B: Nonlinear Monotone Operators.* Springer-Verlag, New York, 1990. Translated from the German by the author and Leo F. Boron.