Deutsche Forschungsgemeinschaft
Priority Programme 1962

An Augmented Lagrangian Method for Optimization Problems in Banach Spaces

Christian Kanzow, Daniel Steck, Daniel Wachsmuth

SPP 1962
Non-smooth and Complementarity-based Distributed Parameter Systems: Simulation and Hierarchical Optimization

Preprint Number SPP1962-003

received on December 12, 2016

Edited by
SPP1962 at Weierstrass Institute for Applied Analysis and Stochastics (WIAS)
Leibniz Institute in the Forschungsverbund Berlin e.V.
Mohrenstraße 39, 10117 Berlin, Germany
E-Mail: spp1962@wias-berlin.de

World Wide Web: http://spp1962.wias-berlin.de/

An Augmented Lagrangian Method for Optimization Problems in Banach Spaces*

Christian Kanzow[†] Daniel Steck[†] Daniel Wachsmuth[†]

December 2, 2016

Abstract

We propose a variant of the classical augmented Lagrangian method for constrained optimization problems in Banach spaces. Our theoretical framework does not require any convexity or second-order assumptions and allows the treatment of inequality constraints with infinite-dimensional image space. Moreover, we discuss the convergence properties of our algorithm with regard to feasibility, global optimality and KKT conditions. Some numerical results are given to illustrate the practical viability of the method.

1 Introduction

Let X, Y be (real) Banach spaces and let $f : X \to \mathbb{R}$, $g : X \to Y$ be given mappings. The aim of this paper is to describe an augmented Lagrangian method for the solution of the constrained optimization problem

$$\min \; f(x) \quad \text{subject to (s.t.)} \quad g(x) \leq 0.$$

We assume that $Y \hookrightarrow L^2(\Omega)$ for some measure space Ω, where the natural order on $L^2(\Omega)$ induces the order on Y. A detailed description together with some remarks about this setting is given in Section 2.

Augmented Lagrangian methods for the solution of optimization problems belong to the most famous and successful algorithms for the solution of finite-dimensional problems and are described in almost all text books on continuous optimization, see, e.g. [5, 22]. Their generalization to infinite-dimensional problems has received considerable attention throughout the last decades [12, 13, 15, 16, 17, 18]. However, most existing approaches

*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 numbers KA 1296/24-1 and Wa 3626/3-1.

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

either assume a very specific problem structure [16], require strong second-order conditions [13] or consider only the case where Y is finite-dimensional [15, 18].

The contribution of the present paper is to overcome these limitations and to provide a general convergence theory for infinite-dimensional problems. To this end, we extend some of the recent contributions on the convergence of certain modified augmented Lagrangian methods from the finite- to the infinite-dimensional case, cf. [7] and references therein for more details regarding some of the newer convergence results in the finite-dimensional setting. The main difference between the classical augmented Lagrangian approach and its modified version consists of a more controlled way of the multiplier update which is responsible for a stronger global convergence theory.

This paper is organized as follows. In Section 2, we give a detailed overview of our problem setting and assumptions. Section 3 contains a precise statement of the algorithm, and we conduct a convergence analysis dedicated to global optimization in Section 4. Starting with Section 5, we assume that the mappings which constitute our problem are continuously differentiable, and establish some theoretical foundations regarding KKT conditions and constraint qualifications. In Section 6, we apply these insights to our algorithm and deduce corresponding convergence results. Finally, Section 7 contains practical applications and we conclude with some final remarks in Section 8.

Notation: We use standard notation such as $\langle \cdot, \cdot \rangle$ for the duality pairing on Y, $(\cdot, \cdot)_Z$ for the scalar product on Z, and $\perp$ to denote orthogonality in Z. The norms on X, Y, etc. are denoted by $\| \cdot \|$, where an index (as in $\| \cdot \|_X$) is appended if necessary. Furthermore, we write $\to$, $\rightharpoonup$, and $\rightharpoonup^*$ for strong, weak, and weak-* convergence. Finally, we use the abbreviation lsc for a lower semicontinuous function.

2 Preliminaries and Assumptions

We denote by $e : Y \to Z$ the (linear and continuous) embedding of Y into $Z := L^2(\Omega)$, and by K_Y, K_Z the respective nonnegative cones in Y and Z, i.e.

$$K_Z = \{z \in Z \mid z(t) \geq 0 \text{ a.e.}\} \quad \text{and} \quad K_Y = \{y \in Y \mid e(y) \in K_Z\}.$$

Note that the adjoint mapping e^* embeds Z^* into Y^*. Hence, we have the chain

$$Y \hookrightarrow Z \cong Z^* \hookrightarrow Y^*. \tag{1}$$

The main reason for the specific configuration of our spaces Y and Z is that $Z = L^2(\Omega)$ is a Hilbert lattice [3, 29] and, hence, the order on Z has some strong structural properties which may not hold on Y. For instance, the order on a Hilbert lattice always satisfies the relation

$$0 \leq z_1 \leq z_2 \quad \Longrightarrow \quad \|z_1\| \leq \|z_2\|,$$

which is trivial for L^2-spaces but does not hold for, say, H^1 or H^1_0. (Hence, these spaces are not Hilbert lattices.) We will put the properties of Z to fruitful use by performing the augmentation which constitutes our algorithm in Z. To simplify this, we denote by z_+ and z_- the positive and negative parts of $z \in Z$, i.e.

$$z_+ = \max\{z, 0\} \quad \text{and} \quad z_- = \max\{-z, 0\}. \tag{2}$$

These operations coincide with the similarly denoted lattice operations on Z and have a variety of properties which are easily verified by resorting to the pointwise definition (2). For instance, we have $z = z_+ - z_-$ and $z_+ \perp z_-$ for every $z \in Z$.

Recall that, as in the introduction, we are concerned with the optimization problem

$$\min \ f(x) \quad \text{s.t.} \quad g(x) \le 0, \tag{3}$$

where $Y \hookrightarrow Z = L^2(\Omega)$. Here, the inequality $g(x) \le 0$ has to be understood with respect to the order induced by the cone K_Y, which is implicitly given by the order on Z through the embedding e.

The following is a list of assumptions which we will use throughout this paper.

Assumption 2.1 (General assumptions on the problem setting)**.**

(A1) f and $\|g_+\|_Z$ are weakly lower semicontinuous.

(A2) f and g are continuously Fréchet-differentiable.

(A3) $y \mapsto |y|$ is well-defined and continuous on Y.

(A4) The unit ball in Y^* is weak-* sequentially compact.

Most of the theorems we will encounter later use only a subset of these assumptions. Hence, we will usually list the assumptions for each theorem explicitly by referencing to the names (A1)-(A4).

One assumption which might require some elaboration is the weak lower semicontinuity of $\|g_+\|_Z$. To this end, note that there are various theorems which characterize the weak lower semicontinuity of convex functions, e.g. [4, Thm. 9.1]. Hence, if $\|g_+\|$ is convex (which is certainly true if g is convex with respect to the order in Y), then the (strong) lower semicontinuity of g already implies the weak lower semicontinuity. We conclude that (A1) holds, in particular, for every lsc. convex function f and any mapping $g \in \mathcal{L}(X, Y)$.

On a further note, the above remarks offer another criterion for the weak lower semicontinuity of $\|g_+\|_Z$. Since $y \mapsto \|y_+\|$ obviously has this property, we conclude that it is sufficient for g to be weakly (sequentially) continuous.

Regarding the space Y which is embedded into Z, recall that (A3) assumed the operation $y \mapsto |y|$ to be well-defined and continuous on Y. (Note that this assumption holds automatically if $Y = Z$, but in many applications, Y is only a subset of Z, cf. the first remark below.) Hence, the same holds for the mappings y_+, y_-, min, max, etc., which may be defined in terms of their counterparts on Z. This allows us to use the lattice structure of Z on Y – at least to a certain extent.

We now give some general remarks about the setting (3).

- Clearly, one motivation for this setting is the case where Ω is a bounded domain in $\mathbb{R}^d$ and Y is one of the spaces $H^1(\Omega)$, $H_0^1(\Omega)$, or $C(\bar{\Omega})$. Problems of this type will be our main application in Section 7. Note that (A3) is satisfied for these spaces, cf. [8, 19] for a proof in H^1.

- In theory, we could easily generalize our work by allowing Z to be a Hilbert lattice [3, 20, 26, 29]. However, it turns out [20, Cor. 2.7.5] that every Hilbert lattice is (isometrically and lattice) isomorphic to $L^2(\Omega)$ for some measure space Ω. Hence, this seemingly more general setting is already covered by ours.
- Related to the previous point, we note that our setting also covers the case $Y = \mathbb{R}^m$, which is a Hilbert lattice and can be identified with $L^2(\Omega)$ on the discrete measure space $\Omega = \{1, \dots, m\}$.

We conclude this section by proving a lemma for later reference. Recall that $(\cdot,\cdot)_Z$ denotes the scalar product in $Z = L^2(\Omega)$.

Lemma 2.2. *Let (a^k) and (b^k) be bounded sequences in Z. Then $\min\{a^k, b^k\} \to 0$ implies $\big(a^k, b^k\big)_Z \to 0$.*

Proof. This follows from $\big(a^k, b^k\big)_Z = \big(\min\{a^k, b^k\}, \max\{a^k, b^k\}\big)_Z$. □

Note that the above lemma becomes false if we drop the boundedness of one of the sequences. For instance, consider the case where $\Omega = \{1\}$ and $Z = L^2(\Omega)$, which can be identified with $\mathbb{R}$. Then the sequences $a^k = k$ and $b^k = 1/k$ provide a simple counterexample.

3 An Augmented Lagrangian Method

This section gives a detailed statement of our augmented Lagrangian method for the solution of the optimization problem (3). It is motivated by the finite-dimensional discussion in, e.g., [7] and differs from the traditional augmented Lagrangian method as applied, e.g., in [12, 16] to a class of infinite-dimensional problems, in a more controlled updating of the Lagrange multiplier estimates.

We begin by defining the augmented Lagrangian

$$L_\rho : X \times Z \to \mathbb{R}, \quad L_\rho(x, \lambda) = f(x) + \frac{\rho}{2} \left\| \left(g(x) + \frac{\lambda}{\rho} \right)_+ \right\|_Z^2. \tag{4}$$

This enables us to formulate the following algorithm for the solution of (3), which is a variant of the (finite-dimensional) method from [7] in the context of our optimization problem (3). In fact, formally, the method looks almost identical to the one from [7], but some of the notations related to the order in Y or Z have a different and more general meaning than those in the finite-dimensional literature.

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

(S.0) Let $(x^0, \lambda^0) \in X \times Z$, $\rho_0 > 0$, $u^{\max} \in K_Z$, $\gamma > 1$, $\tau \in (0, 1)$, and set $k = 0$.

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

(S.2) Choose $0 \le u^k \le u^{\max}$ *and compute an approximate solution* x^{k+1} *of*

$$\min_x \; L_{\rho_k}(x, u^k). \tag{5}$$

(S.3) Set $\lambda^{k+1} = \big(u^k + \rho_k g(x^{k+1})\big)_+$. *If* $k = 0$ *or*

$$\left\| \min\left\{ -g(x^{k+1}), \frac{u^k}{\rho_k} \right\} \right\|_Z \le \tau \left\| \min\left\{ -g(x^k), \frac{u^{k-1}}{\rho_{k-1}} \right\} \right\|_Z \tag{6}$$

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

(S.4) Set $k \leftarrow k + 1$ *and go to (S.1).*

Note that the case $k = 0$ is considered separately in Step 2 for formal reasons only since u^{k-1} and ρ_{k-1} are not defined for this value of the iteration counter. In any case, the treatment of this initial step has no influence on our convergence theory.

One of the most important aspects of the above algorithm is the sequence (u^k). Note that $u^k \le u^{\max}$ implies that (u^k) is bounded in Z. However, the precise choice of u^k does not affect the (theoretical) convergence properties of the method; for instance, we could always choose $u^k = 0$ and thus obtain a simplified algorithm which is essentially a quadratic penalty method. However, the much more natural choice is $u^k = \min\{\lambda^k, u^{\max}\}$. That is, u^k is a bounded analogue of the possibly unbounded multiplier λ^k.

Another part of Algorithm 3.1 which needs some explanation is our notion of an "approximate solution" in Step 2. The reason we have not specified this part is because we will carry out two distinct convergence analyses which each require different assumptions.

4 Global Minimization

We begin by considering Algorithm 3.1 from a global optimization perspective. Note that most of the analysis in this section can be carried out in the more general case where f is an extended real-valued function, i.e. f maps to $\mathbb{R} \cup \{\infty\}$.

The global optimization perspective is particularly valid for convex problems, where we can expect to solve the subproblems in Step 2 in a global sense. This is reflected in the following assumption, which we require throughout this section.

Assumption 4.1. In Step 2 of Algorithm 3.1, we obtain x^{k+1} such that there is a sequence $\varepsilon_k \downarrow 0$ with

$$L_{\rho_k}(x^{k+1}, u^k) \le L_{\rho_k}(x, u^k) + \varepsilon_k \quad \forall x \in X.$$

Assumption 4.1 is quite natural and basically asserts that we finish each inner iteration with a point that is (globally) optimal within some tolerance ε_k, and that this tolerance vanishes asymptotically. Note that $\|g_+\|_Z$ being weakly lsc implies a slightly stronger statement. If $x^k \rightharpoonup x$ and $\alpha^k \to 0$, then the nonexpansiveness of $z \mapsto z_+$ together with (A1) implies that

$$\liminf_{k\to\infty} \left\|(g(x^k) + \alpha^k)_+\right\|_Z = \liminf_{k\to\infty} \left\|g_+(x^k)\right\|_Z \geq \|g_+(x)\|_Z. \tag{7}$$

This fact will be used in the proof of the following theorem.

Theorem 4.2. *Suppose that (A1) and Assumption 4.1 hold. Let (x^k) be a sequence generated by Algorithm 3.1, and let $\bar{x}$ be a weak limit point of (x^k). Then:*

(a) *$\bar{x}$ is a global minimum of the function $\|g_+(x)\|_Z^2$.*

(b) *If $\bar{x}$ is feasible, then $\bar{x}$ is a solution of the optimization problem* (3).

Proof. (a): We first consider the case where (ρ_k) is bounded. Recalling (6), we obtain

$$\|g_+(x^{k+1})\|_Z \leq \left\|\min\left\{-g(x^{k+1}), \frac{u^k}{\rho_k}\right\}\right\|_Z \to 0.$$

Hence (A1) implies that $\bar{x}$ is feasible and the assertion follows trivially.

Next, we consider the case where $\rho_k \to \infty$. Let $\mathcal{K} \subset \mathbb{N}$ be such that $x^{k+1} \rightharpoonup_{\mathcal{K}} \bar{x}$ and assume that there is an $x \in X$ with $\|g_+(x)\|_Z^2 < \|g_+(\bar{x})\|_Z^2$. By (7), the boundedness of (u^k), and the fact that $\rho_k \to \infty$, there is a constant $c > 0$ such that

$$\left\|\left(g(x^{k+1}) + \frac{u^k}{\rho_k}\right)_+\right\|_Z^2 > \left\|\left(g(x) + \frac{u^k}{\rho_k}\right)_+\right\|_Z^2 + c$$

holds for all $k \in \mathcal{K}$ sufficiently large. Hence,

$$L_{\rho_k}(x^{k+1}, u^k) > L_{\rho_k}(x, u^k) + \frac{\rho_k c}{2} + f(x^{k+1}) - f(x).$$

Using Assumption 4.1, we arrive at the inequality

$$\varepsilon_k > \frac{\rho_k c}{2} + f(x^{k+1}) - f(x),$$

where $\varepsilon_k \to 0$. Since $(f(x^{k+1}))_{\mathcal{K}}$ is bounded from below by the weak lower semicontinuity of f, this is a contradiction.

(b): Let $\mathcal{K} \subset \mathbb{N}$ be such that $x^{k+1} \rightharpoonup_{\mathcal{K}} \bar{x}$, and let x be any other feasible point. From Assumption 4.1, we get

$$L_{\rho_k}(x^{k+1}, u^k) \leq L_{\rho_k}(x, u^k) + \varepsilon_k.$$

Again, we distinguish two cases. First assume that $\rho_k \to \infty$. By the definition of the augmented Lagrangian, we have (recall that x is feasible)

$$f(x^{k+1}) \leq f(x) + \frac{\rho_k}{2}\left\|\left(g(x) + \frac{u^k}{\rho_k}\right)_+\right\|_Z^2 + \varepsilon_k \leq f(x) + \frac{\|u^k\|_Z^2}{2\rho_k} + \varepsilon_k.$$

Taking limits in the above inequality, using the boundedness of (u^k) and the weak lower semicontinuity of f, we get $f(\bar{x}) \le f(x)$.

Next, consider the case where (ρ_k) is bounded. Using the feasibility of x and a similar inequality to above, it follows that

$$f(x^{k+1}) + \frac{\rho_k}{2} \left\| \left(g(x^{k+1}) + \frac{u^k}{\rho_k} \right)_+ \right\|^2 \le f(x) + \frac{\rho_k}{2} \left\| \frac{u^k}{\rho_k} \right\|_Z^2 + \varepsilon_k.$$

But

$$\left(g(x^{k+1}) + \frac{u^k}{\rho_k} \right)_+ = \frac{u^k}{\rho_k} - \min \left\{ -g(x^{k+1}), \frac{u^k}{\rho_k} \right\}$$

and the latter part tends to 0 because of (6). This implies $f(\bar{x}) \le f(x)$. $\square$

Note that, for part (a) of the theorem, we did not fully use $\varepsilon_k \downarrow 0$; we only used the fact that (ε_k) is bounded. Hence, this result remains true under weaker conditions than those given in Assumption 4.1. Furthermore, note that Theorem 4.2 does not require any differentiability assumption, though, in practice, the approximate solution of the subproblems in (S.2) of Algorithm 4.1 might be easier under differentiability assumptions. Finally, note that, in view of statement (a), the weak limit point $\bar{x}$ is always feasible if the feasible set of the optimization problem (3) is nonempty, i.e. in this case the feasibility assumption from statement (b) is always satisfied. On the other hand, if the feasible set is empty, it is interesting to note that statement (a) still holds, whereas the assumption from statement (b) cannot be satisfied.

5 Sequential KKT conditions

Throughout this section, we assume that both f and g are continuously Fréchet-differentiable on X, and discuss the KKT conditions of the optimization problem (3). Recalling that K_Y is the nonnegative cone in Y, we denote by

$$K_Y^+ = \{ f \in Y^* \mid \langle f, y \rangle \ge 0 \ \forall y \in K_Y \}$$

its dual cone. This enables us to define the KKT conditions as follows.

Definition 5.1. A tuple $(x, \lambda) \in X \times K_Y^+$ is called a *KKT point* of (3) if

$$f'(x) + g'(x)^* \lambda = 0, \quad g(x) \le 0, \quad \text{and} \quad \langle \lambda, g(x) \rangle = 0. \tag{8}$$

We also call $x \in X$ a *KKT point* of (3) if (x, λ) is a KKT point for some λ.

From a practical perspective, when designing an algorithm for the solution of (3), we will expect the algorithm to generate a sequence which satisfies the KKT conditions in an asymptotic sense. Hence, it will be extremely important to discuss a sequential analogue of the KKT conditions.

Definition 5.2. We say that the *asymptotic KKT (or AKKT) conditions* hold in a feasible point $x \in X$ if there are sequences $x^k \to x$ and $(\lambda^k) \subset K_Y^+$ such that

$$f'(x^k) + g'(x^k)^* \lambda^k \to 0 \quad \text{and} \quad \big\langle \lambda^k, g_-(x^k) \big\rangle \to 0. \tag{9}$$

Asymptotic KKT-type conditions have previously been considered in the literature [1, 2, 7] for finite-dimensional optimization problems. Furthermore, in [7], it is shown that AKKT is a necessary optimality condition even in the absence of constraint qualifications. With little additional work, this result can be extended to our infinite-dimensional setting.

Theorem 5.3. *Suppose that (A1), (A2) hold, and that X is reflexive. Then every local solution $\bar{x}$ of* (3) *satisfies the AKKT conditions.*

Proof. By assumption, there is an $r > 0$ such that $\bar{x}$ is a global solution of (3) on $B_r(\bar{x})$. Now, for $k \in \mathbb{N}$, we consider the problem

$$\min \; f(x) + k \|g_+(x)\|_Z^2 + \|x - \bar{x}\|^2 \quad \text{s.t.} \quad x \in B_r(\bar{x}). \tag{10}$$

Since the above objective function is weakly lsc and $B_r(\bar{x})$ is weakly compact, this problem has a solution x^k. Due to $(x^k) \subset B_r(\bar{x})$, there is a $\mathcal{K} \subset \mathbb{N}$ such that $x^k \rightharpoonup_{\mathcal{K}} \bar{y}$ for some $\bar{y} \in B_r(\bar{x})$. Since x^k is a solution of (10), we have

$$f(x^k) + k \|g_+(x^k)\|_Z^2 + \|x^k - \bar{x}\|^2 \le f(\bar{x}) \tag{11}$$

for every k. Dividing by k and taking the limit $k \to_{\mathcal{K}} \infty$, we obtain from (A1) that $\|g_+(\bar{y})\|_Z = 0$, i.e. $\bar{y}$ is feasible. By (11), we also obtain $f(\bar{y}) + \|\bar{y} - \bar{x}\|^2 \le f(\bar{x})$. But $\bar{x}$ is the unique solution of

$$\min \; f(x) + \|x - \bar{x}\|^2 \quad \text{s.t.} \quad x \in B_r(\bar{x}), \; g(x) \le 0.$$

Hence, $\bar{y} = \bar{x}$ and (11) implies that $x^k \to_K \bar{x}$. In particular, we have $\|x^k - \bar{x}\| < r$ for sufficiently large $k \in \mathcal{K}$, and from (10) we obtain

$$f'(x^k) + 2k g'(x^k)^* g_+(x^k) + 2(x^k - \bar{x}) = 0.$$

Define $\lambda^k = 2k g_+(x^k)$. Then $f'(x^k) + g'(x^k)^* \lambda^k \to_{\mathcal{K}} 0$ and $\big\langle \lambda^k, g_-(x^k) \big\rangle = 0$. ☐

The above theorem also motivates our definition of the AKKT conditions. In particular, it justifies the formulation of the complementarity condition as $\big\langle \lambda^k, g_-(x^k) \big\rangle \to 0$, since the proof shows that (λ^k) needs not be bounded. Hence, the conditions

$$\min\{-g(x^k), \lambda^k\} \to_{\mathcal{K}} 0, \quad \big\langle \lambda^k, g(x^k) \big\rangle \to_{\mathcal{K}} 0, \quad \text{and} \quad \big\langle \lambda^k, g_-(x^k) \big\rangle \to_{\mathcal{K}} 0$$

are not equivalent. Note that the second of these conditions (which might appear as the most natural formulation of the complementarity condition) is often violated by practical algorithms [2].

In order to get the (clearly desirable) implication "AKKT ⇒ KKT", we will need a suitable constraint qualification. In the finite-dimensional setting, constraint qualifications such as MFCQ and CPLD [7, 23] have been used to enable this transition. However, in the infinite-dimensional setting, our choice of constraint qualification is much more restricted. For instance, we are not aware of any infinite-dimensional analogues of the (very amenable) CPLD condition. Hence, we have decided to employ the Zowe-Kurcyusz regularity condition [30], which is known to be equivalent to the Robinson condition [24] and to be a generalization of the finite-dimensional MFCQ. It should be noted, however, that any condition which guarantees "AKKT $\Rightarrow$ KKT" could be used in our analysis.

Definition 5.4. The *Zowe-Kurcyusz condition* holds in a feasible point $x \in X$ if

$$g'(x)X + \operatorname{cone}(K_Y + g(x)) = Y,$$

where cone is the conical hull in Y.

We note that the complete theory in this paper can be written down with $Y = Z$ only, so, formally, there seems to be no reason for introducing the imbedded space Y. One of the main reasons for the more general framework considered here with an additional space Y is that suitable constraint qualifications like the above Zowe-Kurcyusz condition are typically violated even in simple applications when formulated in Z, whereas we will see in Section 7 that this condition easily holds in suitable spaces Y. We therefore stress the importance of Definition 5.4 being defined in Y, and not in Z.

One of the most important consequences of the above condition is that the set of multipliers corresponding to a KKT point x is bounded [30]. From this point of view, it is natural to expect that the sequence (λ^k) from Definition 5.2 is bounded, provided the limit point x satisfies the Zowe-Kurcyusz condition.

Theorem 5.5. *Suppose that (A2) holds. Let $x \in X$ be a point which satisfies the AKKT conditions, and let (x^k), (λ^k) be the corresponding sequences from Definition 5.2.*

(a) *If x satisfies the Zowe-Kurcyusz condition, then (λ^k) is bounded in Y^*.*

(b) *If (A3), (A4) hold and (λ^k) is bounded in Y^*, then x is a KKT point.*

Proof. (a): In view of [30, Thm. 2.1], the Zowe-Kurcyusz condition implies that there is an $r > 0$ such that

$$B_r^Y \subset g'(x)B_1^X + (K_Y + g(x)) \cap B_1^Y,$$

where B_r^X and B_r^Y are the closed r-balls around zero in X and Y, respectively. By the AKKT conditions and (A2), there is a $k_0 \in \mathbb{N}$ such that

$$\|g'(x^k) - g'(x)\| \le r/4 \quad \text{and} \quad \|g(x^k) - g(x)\| \le r/4$$

for every $k \ge k_0$. Now, let $u \in B_r^Y$ and $k \ge k_0$. It follows that $-u = g'(x)w + z$ with $\|w\|_X \le 1$ and $z = z_1 + g(x)$, $\|z\|_Y \le 1$, $z_1 \in K_Y$. Furthermore, the AKKT conditions imply

$$\begin{aligned}\left\langle \lambda^k, z_1 + g(x^k) \right\rangle &= \left\langle \lambda^k, z_1 \right\rangle + \left\langle \lambda^k, g_+(x^k) \right\rangle - \left\langle \lambda^k, g_-(x^k) \right\rangle \\ &\ge - \left\langle \lambda^k, g_-(x^k) \right\rangle \\ &\to 0,\end{aligned}$$

i.e. $\langle \lambda^k, z_1 + g(x^k) \rangle$ is bounded from below. Using once again the AKKT conditions, we see that $\langle \lambda^k, g'(x^k)w \rangle$ is also bounded, and it follows that

$$\begin{aligned}\langle \lambda^k, u \rangle &= -\langle \lambda^k, g'(x)w \rangle - \langle \lambda^k, z_1 + g(x) \rangle \\ &\leq \frac{r}{4}\|\lambda^k\|_{Y^*} - \langle \lambda^k, g'(x^k)w \rangle + \frac{r}{4}\|\lambda^k\|_{Y^*} - \langle \lambda^k, z_1 + g(x^k) \rangle \\ &\leq \frac{r}{2}\|\lambda^k\|_{Y^*} + C\end{aligned}$$

for some constant $C > 0$. We conclude that

$$\|\lambda^k\|_{Y^*} = \sup_{\|u\| \leq r} \left\langle \lambda^k, \frac{1}{r}u \right\rangle \leq \frac{1}{r}\left(C + \frac{r}{2}\|\lambda^k\|_{Y^*} \right)$$

and, hence, $\|\lambda^k\|_{Y^*} \leq 2C/r$.

(b): Since (λ^k) is bounded in Y^* and the unit ball in Y^* is weak-* sequentially compact by (A4), there is a $\mathcal{K} \subset \mathbb{N}$ such that $\lambda^k \rightharpoonup^*_{\mathcal{K}} \lambda$ for some $\lambda \in Y^*$. Using $\lambda^k \in K_Y^+$ for all $k \in \mathbb{N}$, it follows that

$$\langle \lambda, y \rangle = \lim_{k \in \mathcal{K}} \langle \lambda^k, y \rangle \geq 0$$

for every $y \in K_Y$. In other words, $\lambda \in K_Y^+$. Hence, taking the limit in the AKKT conditions and using $g_-(x^k) \to g_-(x) = g(x)$ in Y, which is a consequence of (A3) and the feasibility of x, we see that (x, λ) satisfies the KKT conditions. □

The above theorem is a generalization of a well-known result for the MFCQ constraint qualification in finite dimensions. Recall that, for $Y = \mathbb{R}^m$ with the natural ordering, the Zowe-Kurcyusz condition is equivalent to MFCQ [30].

6 Convergence to KKT Points

We now discuss the convergence properties of Algorithm 3.1 from the perspective of KKT points. To this end, we make the following assumption.

Assumption 6.1. In Step 2 of Algorithm 3.1, we obtain x^{k+1} such that $L'_{\rho_k}(x^{k+1}, u^k) \to 0$ holds for $k \to \infty$.

The above is a very natural assumption which states that x^{k+1} is an (approximate) stationary point of the respective subproblem. Note that, from (4), we obtain the following formula for the derivative of L_{ρ_k} with respect to x:

$$\begin{aligned}L'_{\rho_k}(x^{k+1}, u^k) &= f'(x^{k+1}) + g'(x^{k+1})^*(u^k + \rho_k g(x^{k+1}))_+ \\ &= f'(x^{k+1}) + g'(x^{k+1})^* \lambda^{k+1}.\end{aligned}$$

Our further analysis is split into a discussion of feasibility and optimality. Regarding the feasibility aspect, note that we can measure the infeasibility of a point x by means of the function $\|g_+(x)\|_Z^2$. By standard projection theorems, this is a Fréchet-differentiable function and its derivative is given by $D\|g_+(x)\|_Z^2 = 2g'(x)^* g_+(x)$, cf. [4, Cor. 12.31]. This will be used in the proof of the following theorem.

Theorem 6.2. *Suppose that (A2) and Assumption 6.1 hold. If (x^k) is generated by Algorithm 3.1 and $\bar{x}$ is a limit point of (x^k), then $D\|g_+(\bar{x})\|_Z^2 = 0$.*

Proof. Let $\mathcal{K} \subset \mathbb{N}$ be such that $x^{k+1} \to_{\mathcal{K}} \bar{x}$. If (ρ_k) is bounded, then we can argue as in the proof of Theorem 4.2 (a) and conclude that $\bar{x}$ is feasible. Hence, there is nothing to prove. Now, assume that $\rho_k \to \infty$. By Assumption 6.1, we have

$$f'(x^{k+1}) + g'(x^{k+1})^*(u^k + \rho_k g(x^{k+1}))_+ \to 0.$$

Dividing by ρ_k and using the boundedness of (u^k) and $(f'(x^{k+1}))_{\mathcal{K}}$, it follows that

$$g'(x^{k+1})^* g_+(x^{k+1}) \to_{\mathcal{K}} 0.$$

This completes the proof. □

Similarly to Theorem 4.2, we remark that the above result does not fully use the fact that $L'_{\rho_k}(x^{k+1}, u^k) \to 0$ and remains valid if this sequence is only bounded.

We now turn to the optimality of limit points of Algorithm 3.1. To this end, recall that Assumption 6.1 implies that

$$f'(x^{k+1}) + g'(x^{k+1})^* \lambda^{k+1} \to 0, \tag{12}$$

which already suggests that the sequence of tuples (x^k, λ^k) satisfies AKKT for the optimization problem (3). In fact, the only missing ingredient is the asymptotic complementarity of g and λ. We deal with this issue in two steps. First, we consider the case where (ρ_k) is bounded. In this case, we even obtain the (exact) KKT conditions without any further assumptions.

Theorem 6.3. *Suppose that (A2) and Assumption 6.1 hold. Let (x^k) be generated by Algorithm 3.1 and assume that (ρ_k) is bounded. Then every limit point $\bar{x}$ of (x^k) is a KKT point of (3). The associated Lagrange multiplier $\bar{\lambda}$ belongs to Z.*

Proof. Let $\mathcal{K} \subset \mathbb{N}$ be such that $x^{k+1} \to_{\mathcal{K}} \bar{x}$. Without loss of generality, we assume that $\rho_k = \rho_0$ for all k. From Algorithm 3.1, it follows that $(\lambda^{k+1})_{\mathcal{K}}$ is bounded in Z and

$$\min\left\{-g(x^{k+1}), \frac{u^k}{\rho_0}\right\} \to 0 \quad \text{in } Z.$$

As in the proof of Theorem 4.2 (a), this implies $\|g_+(x^{k+1})\|_Z \to 0$. Furthermore, from Lemma 2.2, we get $\left(u^k, g(x^{k+1})\right)_Z \to_{\mathcal{K}} 0$. Using the definition of λ^{k+1}, we now obtain

$$\begin{aligned}
\left(\lambda^{k+1}, g(x^{k+1})\right)_Z &= \rho_0 \left(\left(\frac{u^k}{\rho_0} + g(x^{k+1}) \right)_+, g(x^{k+1}) \right)_Z \\
&= \rho_0 \left(\frac{u^k}{\rho_0} - \min\left\{-g(x^{k+1}), \frac{u^k}{\rho_0}\right\}, g(x^{k+1}) \right)_Z \\
&\to_{\mathcal{K}} 0.
\end{aligned}$$

Since $(\lambda^{k+1})_{\mathcal{K}}$ is bounded in Z, this also implies $\big(\lambda^{k+1}, g_-(x^{k+1})\big)_Z \to_{\mathcal{K}} 0$. Hence, recalling (12), the AKKT conditions hold in $\bar{x}$. Now, the claim essentially follows from Theorem 5.5 (b), the only difference here is that we are working in the Hilbert space Z instead of Y or Y^*, hence the two conditions (A3) and (A4) formally required in Theorem 5.5 (b) are automatically satisfied in the current Hilbert space situation. □

Some further remarks about the case of bounded multipliers are due. In this case, the multiplier sequence $(\lambda^k)_{\mathcal{K}}$ is also bounded in Z, and it does not make a difference whether we state the asymptotic complementarity of $g(x^k)$ and λ^k as

$$\min\{-g(x^k), \lambda^k\} \to_{\mathcal{K}} 0, \quad \big(\lambda^k, g(x^k)\big)_Z \to_{\mathcal{K}} 0, \quad \text{or} \quad \big(\lambda^k, g_-(x^k)\big)_Z \to_{\mathcal{K}} 0,$$

cf. the remarks after Theorem 5.3. However, this situation changes if we turn to the case where (ρ_k) is unbounded. Here, it is essential that we define the asymptotic KKT conditions exactly as we did in Definition 5.2.

Theorem 6.4. *Suppose that (A2), (A3) and Assumption 6.1 hold. Let (x^k) be generated by Algorithm 3.1 and let $\rho_k \to \infty$. Then every limit point $\bar{x}$ of (x^k) which is feasible satisfies AKKT for the optimization problem* (3).

Proof. Recalling (12), it suffices to show that

$$\big(\lambda^{k+1}, g_-(x^{k+1})\big) = \big((u^k + \rho_k g(x^{k+1}))_+, g_-(x^{k+1})\big) \to_{\mathcal{K}} 0$$

holds on some subset $\mathcal{K} \subset \mathbb{N}$ such that $x^{k+1} \to_{\mathcal{K}} \bar{x}$. To this end, let $v^k := (u^k + \rho_k g(x^{k+1}))_+ \cdot g_-(x^{k+1}) \in L^1(\Omega)$. Since $v^k \geq 0$, we show that $\int_\Omega v^k \to 0$ by using the dominated convergence theorem. Subsequencing if necessary, we may assume that $g(x^{k+1})$ converges pointwise to $g(\bar{x})$, and $|g(x^{k+1})|$ is bounded a.e. by an L^2-function. It follows that v^k is bounded a.e. by an L^1-function (recall that, if $g(x^{k+1})(t) \geq 0$, then $v^k(t) = 0$). Hence, we only need to show that $v^k \to 0$ pointwise. Let $t \in \Omega$ and distinguish two cases:
Case 1. $g(\bar{x})(t) < 0$. In this case, the pointwise convergence implies that $u^k(t) + \rho_k g(x^{k+1})(t) < 0$ for sufficiently large k and, hence, $v^k(t) = 0$ for all such k.
Case 2. $g(\bar{x})(t) = 0$. Then consider a fixed $k \in \mathcal{K}$. If $g(x^{k+1})(t) \geq 0$, it follows again from the definition of v^k that $v^k(t) = 0$. On the other hand, if $g(x^{k+1})(t) < 0$, it follows that $v^k(t) \leq u^k(t) \cdot |g(x^{k+1})(t)|$ (note that the right-hand side converges to zero if this subcase occurs infinitely many times).

Summarizing these two cases, the pointwise convergence $v^k(t) \to 0$ follows immediately. The assertion is therefore a consequence of the dominated convergence theorem. □

For a better overview, we now briefly summarize the two previous convergence theorems. To this end, let (x^k) be the sequence generated by Algorithm 3.1 and let $\bar{x}$ be a limit point of (x^k). For the sake of simplicity, we assume that (A2)-(A4) hold. Then Theorems 6.3, 6.4 and 5.5 imply that $\bar{x}$ is a KKT point if either

(a) the sequence (ρ_k) is bounded, or

(b) $\rho_k \to \infty$, $\bar{x}$ is feasible and the Zowe-Kurcyusz condition holds in $\bar{x}$.

Hence, for $\rho_k \to \infty$, the success of the algorithm crucially depends on the achievement of feasibility and the regularity of the constraint function g. Recall that, by Theorem 6.2, the limit point $\bar{x}$ is always a stationary point of the constraint violation $\|g_+(x)\|_Z^2$. Hence, situations in which $\bar{x}$ is infeasible are rare; in particular, this cannot occur for convex problems (unless, of course, the feasible set itself is empty).

We now turn to a convergence theorem under stronger assumptions. Note that convergence results using local [6, 11] and global [13] second-order conditions have been proven in the literature for algorithms similar to ours (some of these results deal with the finite-dimensional case only). Here, we prove a theorem which shows that our method converges globally for convex problems where the objective function is strongly convex. Note that the theorem does not require any constraint qualification.

Theorem 6.5. *Suppose that (A2) and Assumption 6.1 hold. Furthermore, assume that problem* (3) *has a KKT point* $(\bar{x}, \bar{\lambda})$ *in* $X \times Y^*$*, that* g *is convex, and that* f *is strongly convex. Then the sequence* (x^k) *from Algorithm 3.1 converges to* $\bar{x}$*.*

Proof. We first prove that (x^k) is bounded. Denoting by $c > 0$ the modulus of convexity of f, it is clear that the augmented Lagrangian $L_\rho(\cdot, u)$ is strongly convex with modulus at least c, independently of ρ and u. It follows that

$$c\|x^{k+1} - \bar{x}\|^2 \le \left\langle L'_{\rho_k}(x^{k+1}, u^k) - L'_{\rho_k}(\bar{x}, u^k), x^{k+1} - \bar{x} \right\rangle$$

for all k. Since $\left(L'_{\rho_k}(\bar{x}, u^k)\right)$ is bounded and $L'_{\rho_k}(x^{k+1}, u^k) \to 0$ by Assumption 6.1, it follows that $c\|x^{k+1} - \bar{x}\|^2 \le C\|x^{k+1} - \bar{x}\|$ for some $C > 0$, which implies the boundedness of (x^k).

Next, we prove that $g_+(x^{k+1}) \to 0$. Recalling the proof of Theorem 6.2, this is clear if (ρ_k) is bounded. On the other hand, if $\rho_k \to \infty$, we obtain $h'(x^{k+1}) \to 0$, where $h(x) := \|g_+(x)\|^2$. Since h is convex,

$$0 = h(\bar{x}) \ge h(x^{k+1}) + h'(x^{k+1})(\bar{x} - x^{k+1}).$$

Hence, $h(x^{k+1}) \to 0$ and this implies $g_+(x^{k+1}) \to 0$.

Our final ingredient is the assertion $\liminf_{k\to\infty} r^k \ge 0$, where (r^k) is the sequence given by $r^k = \left\langle \lambda^{k+1}, g(x^{k+1}) - g(\bar{x}) \right\rangle$. If (ρ_k) is bounded, then this assertion is clear from $g(\bar{x}) \le 0$ and the proof of Theorem 6.3. Now, let $\rho_k \to \infty$. Then

$$\begin{aligned} r^k &= \left\langle \lambda^{k+1}, g(x^{k+1}) - g(\bar{x}) \right\rangle \\ &\ge \frac{1}{\rho_k} \left\langle \lambda^{k+1}, u^k + \rho_k g(x^{k+1}) \right\rangle - \frac{1}{\rho_k} \left\langle \lambda^{k+1}, u^k \right\rangle \\ &= \frac{1}{\rho_k} \|\lambda^{k+1}\|_Z^2 - \frac{1}{\rho_k} \left(\lambda^{k+1}, u^k \right)_Z \\ &\ge -\frac{1}{4\rho_k} \|u^k\|_Z^2, \end{aligned}$$

where the last inequality follows from simple quadratic minimization. Hence, we also obtain $\liminf_{k\to\infty} r^k \ge 0$ in this case.

Now, write $\varepsilon^k = L'_{\rho_k}(x^{k+1}, u^k)(x^{k+1} - \bar{x})$ and recall that Assumption 6.1 implies $\varepsilon^k \to 0$. Using the strong monotonicity of f' and the convexity of g, we finally conclude that

$$\begin{aligned}
c\|x^{k+1} - \bar{x}\|^2 &\leq \left\langle f'(x^{k+1}) - f'(\bar{x}), x^{k+1} - \bar{x} \right\rangle \\
&= \left\langle g'(\bar{x})^* \bar{\lambda} - g'(x^{k+1})^* \lambda^{k+1}, x^{k+1} - \bar{x} \right\rangle + \varepsilon^k \\
&\leq \left\langle g'(\bar{x})^* \bar{\lambda}, x^{k+1} - \bar{x} \right\rangle - r^k + \varepsilon^k \\
&\leq \left\langle \bar{\lambda}, g(x^{k+1}) - g(\bar{x}) \right\rangle - r^k + \varepsilon^k \\
&= \left\langle \bar{\lambda}, g(x^{k+1}) \right\rangle - r^k + \varepsilon^k \\
&\leq \left\langle \bar{\lambda}, g_+(x^{k+1}) \right\rangle - r^k + \varepsilon^k.
\end{aligned}$$

This implies $\|x^{k+1} - \bar{x}\| \to 0$. □

The main purpose of Theorem 6.5 is to provide a sufficient condition for the existence of a strong limit point of a sequence (x^k) generated by Algorithm 3.1, as assumed in some of our convergence theorems. Other assumptions like a strong second order sufficiency condition might also guarantee this property.

7 Applications

We now give some applications and numerical results for Algorithm 3.1. To this end, we consider some standard problems from the literature. Apart from the first example, we place special emphasis on nonlinear and nonconvex problems since the appropriate treatment of these is one of the focal points of our method.

All our examples follow the general pattern that X, Y, Z are (infinite-dimensional) function spaces on some bounded domain $\Omega \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$. In each of the subsections, we first give a general overview about the problem in question and then present some numerical results on the unit square $\Omega = (0,1)^2$.

In practice, Algorithm 3.1 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

$$\lambda^0 = 0, \quad \rho_0 = 1, \quad u^{\max} = 10^6 e, \quad \gamma = 10, \quad \tau = 0.5$$

(where λ^0, $u^{\max}$, and $e := (1, \ldots, 1)^T$ are understood to be of appropriate dimension), together with a problem-dependent starting point x^0. The overall stopping criterion which we use for our algorithm is given by

$$\|\nabla f(x) + \nabla g(x)\lambda\|_\infty \leq 10^{-4} \quad \text{and} \quad \|\min\{-g(x), \lambda\}\|_\infty \leq 10^{-4},$$

i.e. it is an inexact KKT condition. Furthermore, in each outer iteration, we solve the corresponding subproblem in Step 2 by computing a point x^{k+1} which satisfies

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

We now turn to our three test problems.

7.1 The Obstacle Problem

We consider the well-known obstacle problem [16, 25]. To this end, let $\Omega \subseteq \mathbb{R}^d$ be a bounded domain, and let $X = Y = H_0^1(\Omega)$, $Z = L^2(\Omega)$. The obstacle problem considers the minimization problem

$$\min\ f(u) \quad \text{s.t.} \quad u \geq \psi, \tag{13}$$

where $f(u) = \|\nabla u\|_{L^2}^2$ and $\psi \in X$ is a fixed obstacle. In order to formally describe this problem within our framework (3), we make the obvious definition

$$g : X \to Y, \quad g(u) = \psi - u.$$

Using the Poincaré inequality, it is easy to see that f is strongly convex. Hence, the obstacle problem satisfies the requirements of Theorem 6.5, which implies that the augmented Lagrangian method is globally convergent. Furthermore, since $X = Y$, it follows that $g'(u) = -\operatorname{id}_X$ for every $u \in X$. Hence, the Zowe-Kurcyusz condition (cf. Definition 5.4) is trivially satisfied in every feasible point, which implies the boundedness of the dual iterates (λ^k), cf. Theorem 5.5.

In fact, the constraint function g satisfies much more than the Zowe-Kurcyusz condition. For every $u \in X$, the mapping $g'(u) = -\operatorname{id}_X$ is bijective. Hence, if a subsequence $(x^k)_\mathcal{K}$ converges to a KKT point $\bar{x}$ of (13) and $\bar{\lambda}$ is the corresponding multiplier, then we obtain

$$f'(x^k) - \lambda^k = f'(x^k) + g'(x^k)^* \lambda^k \to_\mathcal{K} 0.$$

In other words, we see that $\lambda^k \to_\mathcal{K} f'(\bar{x}) = \bar{\lambda}$, i.e. $(\lambda^k)_\mathcal{K}$ converges to the (unique) Lagrange multiplier corresponding to $\bar{x}$.

We now present some numerical results for $\Omega = (0,1)^2$ and the obstacle

$$\psi(x, y) = \max\left\{ 0.1 - 0.5 \left\| \begin{pmatrix} x - 0.5 \\ y - 0.5 \end{pmatrix} \right\|, 0 \right\},$$

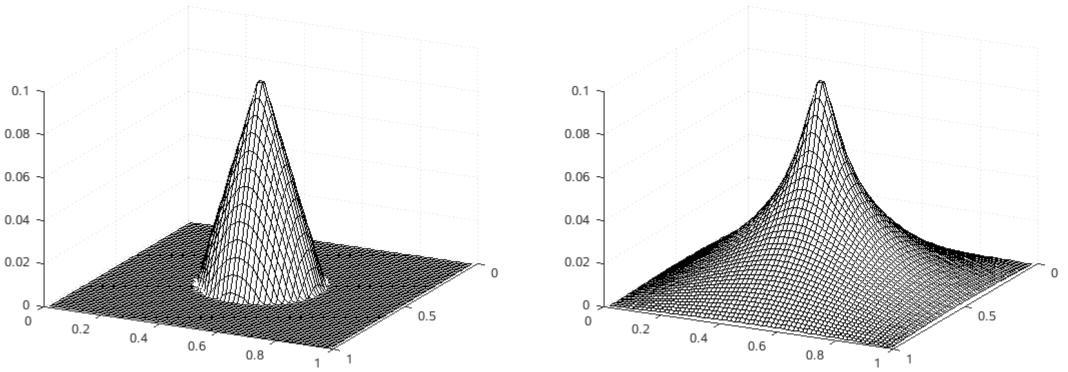
cf. Figure 1. For the solution process, we choose $n \in \mathbb{N}$ and discretize Ω by means of a standard grid which consists of n (interior) points per row or column, i.e. n^2 interior points in total. Furthermore, we use

$$f(u) = \|\nabla u\|_{L^2}^2 = -\left(\Delta u, u\right)_{L^2} \quad \text{for all } u \in X$$

and approximate the Laplace operator by a standard five-point finite difference scheme. The subproblems occurring in Algorithm 3.1 are unconstrained minimization problems which we solve by means of a globalized (semismooth) Newton method. The implementation details of this method are rather straightforward and, hence, we do not give them here.

The following table contains iteration numbers for some values of the discretization parameter n.

n	16	32	64	128	256
outer it.	4	4	3	3	3
inner it.	7	8	11	14	20



(a) Constraint function ψ (b) Solution $\bar{u}$

Figure 1: Numerical results for the obstacle problem with $n = 64$.

It is interesting to note that the outer iteration numbers stay approximately the same as n increases. Furthermore, we observed that the same holds for the final penalty parameter ρ, which is equal to 10 regardless of the dimension n. This suggests that our method is well-formed and works quite well for the obstacle problem.

7.2 The Obstacle Bratu Problem

Let us briefly consider the obstacle Bratu problem [9, 14], which we simply refer to as Bratu problem. This is a non-quadratic and nonconvex problem which differs from (13) in the choice of objective function. To this end, let

$$f(u) = \|\nabla u\|_{L^2}^2 - \alpha \int_\Omega e^{-u(x)} dx$$

for some fixed $\alpha > 0$. To ensure well-definedness of f, we set

$$X = Y = H_0^1(\Omega) \cap C(\bar{\Omega}), \quad \text{with} \quad \|u\|_X = \|u\|_{H_0^1(\Omega)} + \|u\|_{C(\bar{\Omega})}.$$

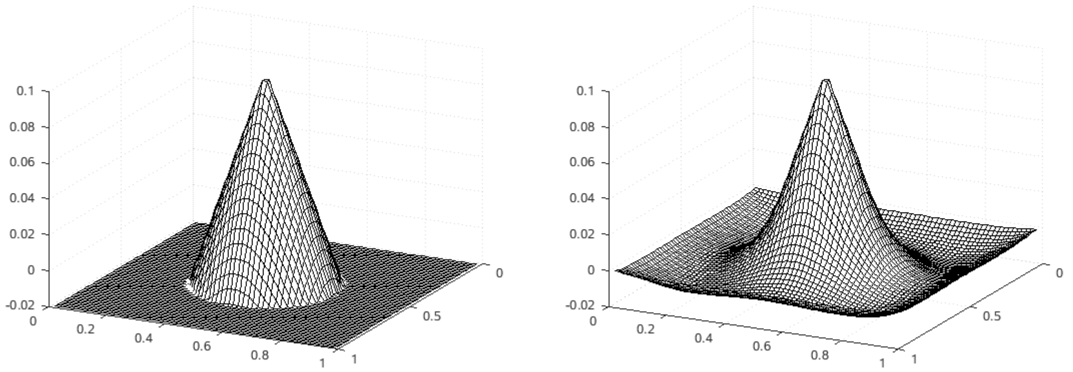
As before, $Z = L^2(\Omega)$ and we consider the minimization problem

$$\min \; f(u) \quad \text{s.t.} \quad u \geq \psi \tag{14}$$

for some fixed obstacle $\psi \in X$; that is, $g(u) = \psi - u$.

From a theoretical point of view, the Bratu problem is much more difficult than the obstacle problem from Section 7.1. While the constraint function is equally well-behaved, the objective function in (14) is neither quadratic nor convex. Hence, we cannot apply Theorem 6.5 or the theory from Section 4, wheras the KKT-like convergence results from Sections 5 and 6 still hold.

To analyse how our method behaves in practice, we again considered $\Omega = (0, 1)^2$ and implemented the Bratu problem using the same obstacle and a similar implementation



(a) Constraint function ψ

(b) Solution $\bar{u}$

Figure 2: Numerical results for the Bratu problem with $\alpha = 1$ and $n = 64$.

as we did for the standard obstacle problem. The resulting images are given in Figure 2, and some iteration numbers are given in the following table.

n	16	32	64	128	256
outer it.	4	4	3	3	3
inner it.	12	10	13	17	21

As with the obstacle problem, we note that both the inner iteration numbers and the final penalty parameters remain nearly constant as n increases. The final value of ρ is again equal to 10, regardless of the dimension n.

7.3 Optimal Control Problems

We now turn to a class of optimal control problems subject to a semilinear elliptic equation. Let $\Omega \subseteq \mathbb{R}^d$, $d = 2, 3$, be a bounded Lipschitz domain. The control problem we consider consists of minimizing the functional

$$J(y, u) := \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \tag{15}$$

subject to $y \in H^1_0(\Omega) \cap C(\bar{\Omega})$ and $u \in L^2(\Omega)$ satisfying the semilinear equation

$$-\Delta y + d(y) = u \quad \text{in } H^1_0(\Omega)^*$$

and the pointwise state constraints

$$y \geq y_c \quad \text{in } \Omega. \tag{16}$$

Here, α is a positive parameter, $y_d \in L^2(\Omega)$, and $y_c \in C(\bar{\Omega})$ with $y_c \leq 0$ 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 continuously differentiable and monotonically increasing.

Before we can apply the augmented Lagrangian method to (15), we need to formally eliminate the state equation coupling the variables y and u. Due to elliptic regularity results, this equation admits for each control $u \in L^2(\Omega)$ a uniquely determined weak solution $y \in H_0^1(\Omega) \cap C(\bar{\Omega})$. Moreover, the mapping $u \mapsto y$ is Fréchet differentiable in this setting [27, Thm. 4.15]. Let us denote this mapping by S. Using S, we can eliminate the state equation to obtain an optimization problem with inequality constraints:

$$\min \; J(S(u), u) \quad \text{s.t.} \quad S(u) \geq y_c. \tag{17}$$

We can now apply Algorithm 3.1 to this problem. The inequality $S(u) \geq y_c$ has to be understood in the sense of $C(\bar{\Omega})$, which necessitates the choice $Y = C(\bar{\Omega})$. Furthermore, we have $X = Z = L^2(\Omega)$. Assuming a linearized Slater condition, one can prove that the Zowe-Kurcyusz condition is fulfilled, and there exists a Lagrange multiplier $\lambda \in C(\bar{\Omega})^*$ to the inequality constraint $S(u) \geq y_c$, see, e.g., [27, Thm. 6.8].

The subproblems generated by Algorithm 3.1 are unconstrained optimization problems. By reintroducing the state variable y, we can write these subproblems as

$$\min \; J(y, u) + \frac{\rho_k}{2} \left\| \left(y_c - y + \frac{u^k}{\rho_k} \right)_+ \right\|^2 \quad \text{s.t.} \quad y = S(u). \tag{18}$$

Hence, we have transformed (15) into a sequence of optimal control problems which include the state equation but not the pointwise constraint (16). Recall that u^k is an iteration parameter and should not be confused with the control u.

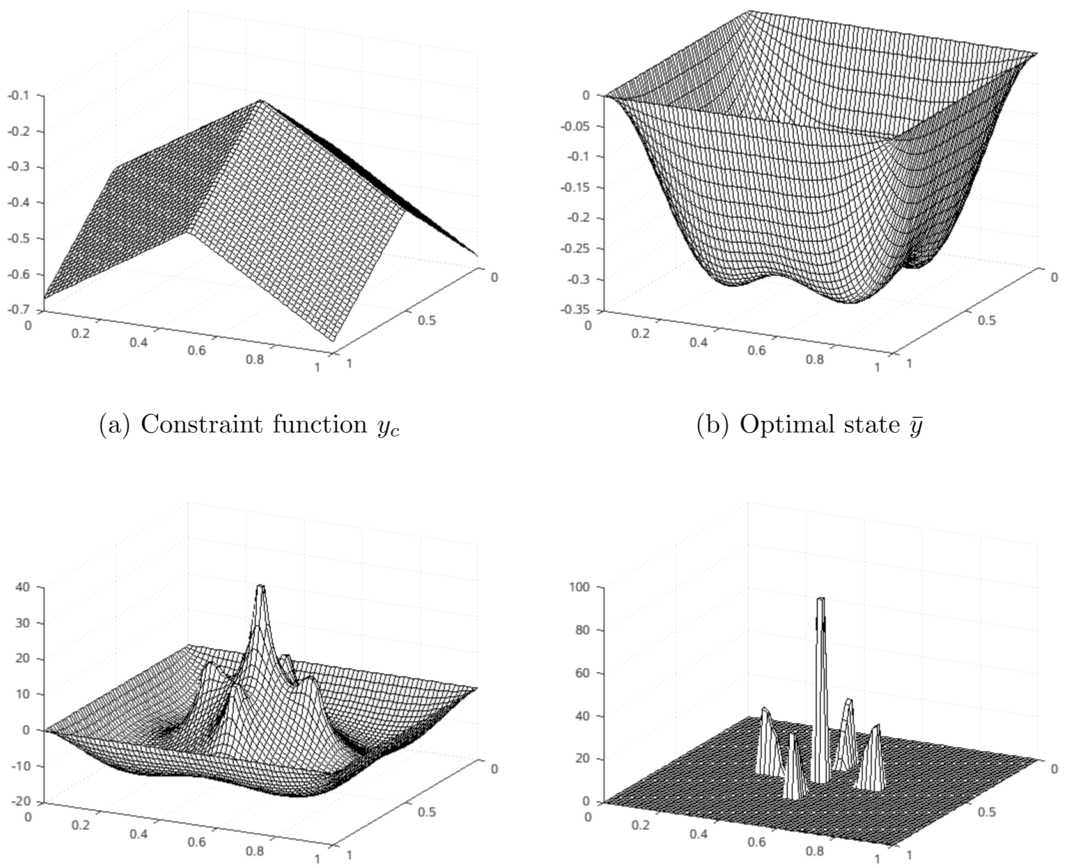
In the following, let us report about numerical results. As a test problem, we choose the example presented in [21], where $\Omega = (0, 1)^2$ and $d(y) = y^3$. Clearly, in this setting, (15) and its reformulation (17) are nonconvex problems. We solve the subproblems (18) by applying a semismooth Levenberg-Marquardt-type method to the corresponding KKT conditions (cf. [10, 28] for some similar methods). Some iteration numbers are given as follows:

n	16	32	64	128	256
outer it.	11	11	11	14	16
inner it.	268	573	701	2187	3459

As before, $n \in \mathbb{N}$ is the number of points per row or column of our discretization scheme. The state constraint y_c and the results of our method are given in Figure 3. 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

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

The latter is well-known to be the space of Radon measures on $\bar{\Omega}$, 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 (15) does not admit a multiplier in $L^2(\Omega)$ but only in $C(\bar{\Omega})^*$.



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

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

Figure 3: Numerical results for the optimal control problem with $n = 64$.

8 Final Remarks

We have presented an augmented Lagrangian method for the solution of optimization problems in Banach spaces, which is essentially a generalization of the modified augmented Lagrangian method from [7]. Furthermore, we have shown how the method can be applied to well-known problem classes, and the corresponding numerical results appear quite promising.

The main strength of our method is the ability to deal with very general classes of inequality constraints; in particular, inequality constraints with infinite-dimensional image space. Other notable features include desirable convergence properties for nonsmooth problems, the ability to find KKT points of arbitrary nonlinear (and nonconvex) problems, and a global convergence result which covers many prominent classes of convex problems. We believe the sum of these aspects to be a substantial contribution to the theory of augmented Lagrangian methods.

Another key concern in our work is the compatibility of the algorithm with suitable constraint qualifications. To deal with this matter properly, we investigated the well-known Zowe-Kurcyusz regularity condition [30], see also Robinson [24], and showed that this condition can be used to guarantee the boundedness of suitable multiplier sequences corresponding to asymptotic KKT conditions. While the main application of this result is clearly the boundedness of the multiplier sequence generated by the augmented Lagrangian method, we state explicitly that the underlying theory is independent of our specific algorithm. With the understanding that most iterative methods for constrained optimization usually satisfy the KKT conditions in an asymptotic sense, we hope that this aspect of our theory will facilitate similar research into other methods or find applications in other topics.

References

[1] R. Andreani, G. Haeser, and J. M. Martínez. On sequential optimality conditions for smooth constrained optimization. *Optimization*, 60(5):627–641, 2011.

[2] R. Andreani, J. M. Martínez, and B. F. Svaiter. A new sequential optimality condition for constrained optimization and algorithmic consequences. *SIAM J. Optim.*, 20(6):3533–3554, 2010.

[3] C. Baiocchi and A. Capelo. *Variational and Quasivariational Inequalities*. John Wiley & Sons, Inc., New York, 1984.

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

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

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

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

[8] C. V. Coffman, R. J. Duffin, and V. J. Mizel. Positivity of weak solutions of non-uniformly elliptic equations. *Ann. Mat. Pura Appl. (4)*, 104:209–238, 1975.

[9] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems. Vol. I.* Springer-Verlag, New York, 2003.

[10] J.-y. Fan and Y.-x. Yuan. On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. *Computing*, 74(1):23–39, 2005.

[11] 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.

[12] M. Fortin and R. Glowinski. *Augmented Lagrangian methods*, volume 15 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam, 1983. Applications to the numerical solution of boundary value problems, Translated from the French by B. Hunt and D. C. Spicer.

[13] 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.

[14] R. H. W. Hoppe and H. D. Mittelmann. A multi-grid continuation strategy for parameter-dependent variational inequalities. *J. Comput. Appl. Math.*, 26(1-2):35–46, 1989.

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

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

[17] K. Ito and K. Kunisch. Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces. *Nonlinear Anal.*, 41(5-6, Ser. A: Theory Methods):591–616, 2000.

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

[19] M. Marcus and V. J. Mizel. Every superposition operator mapping one Sobolev space into another is continuous. *J. Funct. Anal.*, 33(2):217–229, 1979.

[20] P. Meyer-Nieberg. *Banach Lattices*. Springer-Verlag, Berlin, 1991.

[21] 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.

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

[23] L. Qi and Z. Wei. On the constant positive linear dependence condition and its application to SQP methods. *SIAM J. Optim.*, 10(4):963–981, 2000.

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

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

[26] H. H. Schaefer. *Banach Lattices and Positive Operators.* Springer-Verlag, New York-Heidelberg, 1974.

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

[28] N. Yamashita and M. Fukushima. On the rate of convergence of the Levenberg-Marquardt method. In *Topics in Numerical Analysis*, volume 15 of *Comput. Suppl.*, pages 239–249. Springer, Vienna, 2001.

[29] K. Yosida. *Functional Analysis.* Springer-Verlag, Berlin, 1995.

[30] J. Zowe and S. Kurcyusz. Regularity and stability for the mathematical programming problem in Banach spaces. *Appl. Math. Optim.*, 5(1):49–62, 1979.