Deutsche Forschungsgemeinschaft
Priority Programme 1962

A Generalized Proximal-Point Method for Convex Optimization Problems in Hilbert Spaces

Christian Kanzow, Daniel Steck

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

Preprint Number SPP1962-002

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/

A Generalized Proximal-Point Method for Convex Optimization Problems in Hilbert Spaces*

Christian Kanzow† Daniel Steck†

December 8, 2016

Abstract

We deal with a generalization of the proximal-point method and the closely related Tikhonov regularization method for convex optimization problems. The prime motivation behind this is the well-known connection between the classical proximal-point and augmented Lagrangian methods, and the emergence of modified augmented Lagrangian methods in recent years. Our discussion includes a formal proof of a corresponding connection between the generalized proximal-point method and the modified augmented Lagrange approach in infinite dimensions. Several examples and counterexamples illustrate the convergence properties of the generalized proximal-point method and indicate that the corresponding assumptions are sharp.

1 Introduction

Let X be a Hilbert space and $f : X \to \overline{\mathbb{R}}$ a proper, weakly lower semicontinous (lsc.), convex function. For brevity, we write $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. Consider the optimization problem

$$\min_{x \in X} \; f(x). \tag{1}$$

We denote by $S \subseteq X$ its solution set (possibly empty) and by $f_{\min} \in \mathbb{R} \cup \{-\infty\}$ the infimum of f. Some standard methods for the solution of (1) include the proximal-point method [4, 11, 19, 20, 23] and the Tikhonov regularization method [10, 24]. It is particularly the former method which has been found to have a rich theoretical background, including a connection to the classical augmented Lagrangian method (or *method of multipliers*), cf. [5, 13, 22].

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

†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.

The augmented Lagrangian method is one of the standard techniques for the solution of constrained optimization methods. It is therefore also the basis of some standard software packages like LANCELOT [9] and ALGENCAN [7]. The well-known relation between the augmented Lagrangian method and the proximal-point technique [13, 22] gives some deeper insight in the convergence properties of augmented Lagrangian methods, especially with respect to the dual variables.

In recent years, modified versions of the augmented Lagrangian method have surfaced [1, 6, 7] which seek to alleviate some of the weaknesses of the classical method (and are the basis of the previously mentioned ALGENCAN software package). Furthermore, these modified methods have been extended to more general classes of problems such as variational inequalities [2], quasi-variational inequalities [14, 16], generalized Nash equilibrium problems [15], and optimization problems in Banach spaces [17]. The convergence properties of the modified augmented Lagrangian methods seem to be stronger than those of their classical counterparts, at least with respect to the primal variables.

Motivated by these developments, we are naturally inclined to consider the relation to proximal-point-type algorithms for these new methods. It turns out that this connection is given by a generalized proximal-point method. This generalized proximal-point method is, of course, also of its own interest, and encompasses both the classical proximal-point and the Tikhonov regularization method. The convergence results obtained for the new method cover (directly or indirectly) the convergence properties of the standard proximal-point method and partially extend the convergence theory of the Tikhonov regularization technique. Finally, we note that our theoretical framework also includes modified proximal-point type methods such as the one presented in [12].

This paper is organized as follows. In Section 2, we formally introduce our method and prove corresponding convergence theorems. Section 3 contains some examples and counterexamples related to the convergence results which show the sharpness of our assertions. In Section 4, we give a simple proof of the connection between our method and the modified augmented Lagrangian method from [7]. This proof generalizes the corresponding result from [13] by considering Hilbert spaces with conical constraints. We conclude with some final remarks in Section 5.

Notation: The scalar product in the Hilbert space X is denoted by $\langle \cdot, \cdot \rangle$, the induced norm is written as $\| \cdot \|$. For a nonempty, closed, and convex set $C \subset X$ and a given point $u \in X$, the symbol $P_C(u)$ denotes the projection of u onto C with respect to the induced norm. Convex and strongly convex functions are defined in the usual way, see, for example, [4].

2 The Generalized Proximal-Point Method

Recall that we are dealing with the optimization problem (1). For a given $\gamma > 0$, we consider the proximal point mapping

$$\operatorname{prox}_{\gamma f}(u) = \operatorname*{argmin}_{x \in X} \left\{ f(x) + \frac{1}{2\gamma} \|x - u\|^2 \right\}. \tag{2}$$

Note that the right-hand side is always strongly convex, which implies the well-definedness of the prox-operator. The classical proximal-point method [4], which we simply refer to as PPM, is defined by means of the recurrence

$$x^{k+1} = \operatorname{prox}_{\gamma_k f}(x^k), \tag{3}$$

where $\gamma_k > 0$ is some parameter. A similar method is the Tikhonov regularization method, which is given by

$$x^{k+1} = \operatorname*{argmin}_{x \in X} \left\{ f(x) + \frac{1}{2\gamma_k}\|x - u\|^2 \right\} = \operatorname{prox}_{\gamma_k f}(u) \tag{4}$$

for some fixed $u \in X$ (the classical Tikhonov method corresponds to $u = 0$). The method we consider is a generalization of both these methods which is obtained by replacing u in (4) with an arbitrary (bounded) sequence $(u^k) \subseteq X$. Note that, from a theoretical point of view, this encompasses (3) because the sequence (x^k) generated by the proximal-point method is known to be weakly convergent (and, hence, bounded) if the solution set is nonempty [11]. Note that [12] also considers an iterative scheme where the fixed u in (4) is replaced by a suitable sequence (u^k), but the sequence there is constructed by the algorithm with the motivation to improve the rate of convergence, whereas here the (bounded) sequence (u^k) is provided by the user with the idea to get suitable convergence results related to a class of modified augmented Lagrangian methods.

We formally state our generalized proximal-point method in the following algorithm.

Algorithm 2.1. *(Generalized Proximal-Point Method)*

(S.0) Let $x^1 \in X$, $B \subseteq X$ a bounded set, and set $k := 1$.

(S.1) If x^k is a solution of the optimization problem (1): STOP.

(S.2) Choose $u^k \in B$, $\gamma_k > 0$, and let $x^{k+1} := \operatorname{prox}_{\gamma_k f}(u^k)$, i.e. x^{k+1} solves

$$\min_{x \in X} \; f(x) + \frac{1}{2\gamma_k}\|x - u^k\|^2. \tag{5}$$

(S.3) Set $k \leftarrow k + 1$, and go to (S.1).

Note that the choice of u^k in (S.2) may not depend on the current iterate x^k, in which case we view Algorithm 2.1 as a generalization of the Tikhonov regularization approach. On the other hand, a (possibly more natural) choice of u^k depending on x^k brings Algorithm 2.1 closer to the classical proximal-point method. The convergence properties of Algorithm 2.1 obviously depend on the particular choice of the sequence (u^k).

Since the objective function in (5) is strongly convex, this implies the well-definedness of the sequence (x^k) and, hence, of the overall algorithm. Note that, formally speaking, Algorithm 2.1 is not really an iterative method, since the definition of x^{k+1} does not necessarily depend on x^k (unless u^k depends on x^k; furthermore, in practice, one might use x^k as an initial approximation for the solution of (5)). Nevertheless, the formulation as an iterative procedure simplifies our formal treatment of the method. The subsequent convergence analysis assumes implicitly that our generalized proximal-point method generates an infinite sequence.

We now turn to a detailed convergence analysis for Algorithm 2.1. To this end, we begin with a technical lemma that will be one of our main tools.

Lemma 2.2. *Let (x^k) be generated by Algorithm 2.1, and $x \in X$. Then*

$$\begin{aligned} f(x^{k+1}) - f(x) &\leq \frac{\|x - u^k\|^2 - \|x^{k+1} - u^k\|^2}{2\gamma_k} \qquad (6) \\ &\leq \frac{\|x - u^k\|^2}{2\gamma_k} \qquad (7) \end{aligned}$$

for every $k \geq 1$. Furthermore, if S is nonempty, then (x^k) is bounded.

Proof. The second inequality is obvious. To verify the first inequality, note that the definition of x^{k+1} implies that

$$f(x^{k+1}) + \frac{1}{2\gamma_k}\|x^{k+1} - u^k\|^2 \leq f(x) + \frac{1}{2\gamma_k}\|x - u^k\|^2$$

for all $k \geq 1$. Reordering gives the desired statement.

If S is nonempty, then (6) holds for $x = x^* \in S$. Since $f(x^*) \leq f(x^{k+1})$, it follows that the left-hand side of (6) is nonnegative. We therefore obtain $\|x^{k+1} - u^k\| \leq \|x^* - u^k\|$ for all $k \geq 1$. Using the boundedness of (u^k), we conclude that (x^k) is also bounded. □

As the proof shows, the assertion of Lemma 2.2 is essentially a trivial reformulation of the definition of x^{k+1}. Recall that x^{k+1} does not really depend on x^k; hence, roughly speaking, we cannot expect x^{k+1} to satisfy more than the inequality (6). This point will be emphasized in Section 3, where we give some examples which show that our upcoming convergence theorems can, in general, not be strengthened.

Theorem 2.3. *Let (x^k) be generated by Algorithm 2.1, and let $\gamma_k \to \infty$. Then $f(x^k) \to f_{\min}$ and every weak limit point of (x^k) is a solution of the optimization problem (1). Furthermore, if S is nonempty and $x^* \in S$, then*

$$f(x^{k+1}) - f_{\min} = \mathcal{O}\left(\frac{\|x^{k+1} - x^*\|}{\gamma_k}\right) = \mathcal{O}\left(\frac{1}{\gamma_k}\right).$$

Proof. The boundedness of (u^k), the assumption $\gamma_k \to \infty$, and the inequality (7) together yield

$$\limsup_{k\to\infty} f(x^{k+1}) \leq f(x)$$

for every $x \in X$. This implies $f(x^{k+1}) \to f_{\min}$. The second assertion follows from the weak lower semicontinuity of f, and the third assertion follows by applying (6) with $x = x^* \in S$ and using the triangle inequality. The final equation is a consequence of the boundedness of (x^k), cf. Lemma 2.2. □

The above is our main convergence theorem and includes a host of assertions on the sequence (x^k) and the corresponding sequence of function values $(f(x^k))$. Note that, if S is a singleton (i.e. the original problem has a unique solution x^*), then $x^k \rightharpoonup x^*$ as a consequence of Lemma 2.2 and Theorem 2.3. These results also imply that the sequence (x^k) is unbounded if the solution set S is empty, since its boundedness would imply the existence of a weakly convergent subsequence whose limit point would then be a solution of (1). We further note that Theorem 2.3 talks about weak limit points; in fact, without any further assumptions, the sequence (x^k) cannot be expected to have a strong limit point. This follows (indirectly) from the observation that Algorithm 2.1 includes the classical proximal-point algorithm for which strong convergence cannot be expected, cf. [11]. In addition, we also provide an explicit counterexample in Section 3.

Furthermore, we note that Theorem 2.3 requires that $\gamma_k \to \infty$ (it is fairly trivial to construct examples which show that this is necessary). Some further examples illustrating the sharpness of the theorem will be given in Section 3.

We now turn to the case where f is strongly convex. This allows us to prove strong convergence of the iterates x^k and an improved rate of convergence.

Theorem 2.4. *Let f be strongly convex, and let $\gamma_k \to \infty$. Then (x^k) converges (strongly) to the unique element $x^* \in S$. Moreover, we have*

$$\|x^{k+1} - x^*\| = \mathcal{O}\left(\frac{1}{\gamma_k}\right) \quad \text{and} \quad f(x^{k+1}) - f(x^*) = \mathcal{O}\left(\frac{1}{\gamma_k^2}\right).$$

Proof. The strong convexity implies that there is a constant $c_1 > 0$ such that

$$\begin{aligned} c_1\|x^k - x^*\|^2 &\leq \frac{f(x^k) + f(x^*)}{2} - f\left(\frac{x^k + x^*}{2}\right) \\ &\leq \frac{f(x^k) + f(x^*)}{2} - f(x^*) \\ &= \frac{f(x^k) - f(x^*)}{2}. \end{aligned}$$

Hence, by Theorem 2.3, we obtain $x^k \to x^*$. Moreover, by Theorem 2.3,

$$f(x^{k+1}) - f(x^*) = \mathcal{O}\left(\frac{\|x^{k+1} - x^*\|}{\gamma_k}\right).$$

Hence, there is a constant $c_2 > 0$ such that

$$c_1\|x^{k+1} - x^*\|^2 \leq \frac{1}{2}\left(f(x^{k+1}) - f(x^*)\right) \leq c_2 \frac{\|x^{k+1} - x^*\|}{\gamma_k}.$$

This immediately gives the desired convergence rate estimates. □

This concludes our general convergence analysis of Algorithm 2.1. We now give one final result, mainly for the sake of completeness, which emphasizes the connection between our method and the Tikhonov regularization method. Recall that S is the solution set of (1) and, hence, closed and convex.

Theorem 2.5. *Let (x^k) be generated by Algorithm 2.1, let $\gamma_k \to \infty$, and let $u^k \to u^*$ for some $u^* \in X$. If S is nonempty, then $x^k \to P_S(u^*)$.*

Proof. First note that $f(x^*) \leq f(x^{k+1})$ holds for an arbitrary solution $x^* \in S$. Hence, from (6), we obtain $\|x^{k+1} - u^k\| \leq \|x^* - u^k\|$ for all $x^* \in S$ and all $k \geq 1$. The assumed convergence of u^k then implies

$$\begin{aligned}
\limsup_{k\to\infty} \|x^{k+1} - u^*\| &\leq \limsup_{k\to\infty} \big(\|x^{k+1} - u^k\| + \|u^k - u^*\|\big) \\
&= \limsup_{k\to\infty} \|x^{k+1} - u^k\| \\
&\leq \limsup_{k\to\infty} \|x^* - u^k\| \\
&= \|x^* - u^*\|.
\end{aligned}$$

By Lemma 2.2 and Theorem 2.3, there is a subset $K \subseteq \mathbb{N}$ such that $x^{k+1} \rightharpoonup_K x$ for some $x \in S$. For the particular solution $x^* := P_S(u^*)$, it then follows that

$$\|x - u^*\| \leq \limsup_{k\in K} \|x^{k+1} - u^*\| \leq \|x^* - u^*\|,$$

where the first inequality exploits the weak lower semicontinuity of the norm. Hence, $x = x^*$ and, therefore, $\|x^{k+1} - u^*\| \to_K \|x - u^*\|$, which implies $x^{k+1} \to_K x = x^*$. Since this holds for every weakly convergent subsequence of (x^k), we conclude that $x^k \to x^*$. $\square$

Note that Theorem 2.5 recovers the well-known result that the iterates generated by the classical Tikhonov regularization method (where $u^k = 0$ for all $k \geq 1$) converges to the minimum-norm solution of (1) if the solution set S is nonempty. We also stress a difference between the (generalized) proximal-point method and the (generalized) Tikhonov regularization technique in the infinite-dimensional Hilbert setting: The former generates a sequence with, usually, weak convergence properties, whereas the latter computes a strongly convergent sequence (under the assumptions of Theorem 2.5).

Finally, we note that the assertions made by Theorem 2.5 are strongly dependent on the assumption that the algorithm generates an infinite sequence. For instance, if we consider an optimization problem where S is not a singleton (e.g. Example 3.1 in the upcoming section), then (x^k) may converge to the projection of u^* onto S, but this does not exclude the possibility that some of the iterates x^k themselves already lie in S. In this case, depending on the stopping criterion, Algorithm 2.1 might terminate with a point that is a solution, but is not equal to $P_S(u^*)$.

3 Examples and Counterexamples

We now give some examples which illustrate the convergence assertions of the theorems from Section 2. In particular, we wish to show that the convergence results can, in general, not be strengthened.

Recall that Theorem 2.3 asserts the convergence $f(x^k) \to f_{\min}$ and a convergence rate if f attains its minimum. Hence, it is natural to ask whether (x^k) converges (possibly weakly) to an element $x^* \in S$. As we remarked in Section 2, the weak convergence follows trivially if S is a singleton. However, if S has at least two elements x^* and y^*, it is easy to see that (x^k) may not be convergent. The following is a concrete counterexample.

Example 3.1. *Consider the convex function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) := \max\{0, x^2-1\}$. The solution set of the corresponding optimization problem* (1) *is given by $S = [-1, 1]$. Now, consider Algorithm 2.1 with $\gamma_k := k$ and the alternating sequence $u^{2k} := 1$, $u^{2k+1} := -1$ for all k. Then x^{2k+1} is the unique solution of*

$$\min_x \ \max\{0, x^2 - 1\} + \frac{1}{4k}(x-1)^2,$$

which is given by $x^{2k+1} = 1$ since this number minimizes both terms separately, whereas x^{2k} is the solution of

$$\min_x \ \max\{0, x^2 - 1\} + \frac{1}{2(2k-1)}(x+1)^2$$

and is therefore given by $x^{2k} = -1$ for similar reasons. Hence, we eventually get the alternating sequence $(-1, 1, -1, 1, \ldots)$. $\Diamond$

In practice, Algorithm 2.1 applied to Example 3.1 would have stopped after the first iteration since x^1 is already a solution of the underlying minimization problem. However, to illustrate certain convergence properties, it is useful to consider what happens if the method generates an infinite sequence. In fact, one could easily modify the example to obtain a sequence $(x^k) \subseteq \mathbb{R} \setminus [-1, 1]$ with the same accumulation points 1 and -1.

The following example shows that the convergence rate asserted by Theorem 2.3 holds only if S is nonempty.

Example 3.2. *Let $f : \mathbb{R} \to \mathbb{R}$ be the convex function given by*

$$f(x) = \begin{cases} x^{-\alpha} & \text{if } x > 0 \\ \infty & \text{otherwise} \end{cases}$$

for some constant $\alpha > 0$. Hence $f_{\min} = 0$, but the minimum is not attained. If $u^k = 0$ for all k, then x^{k+1} is the global minimum of $x^{-\alpha} + x^2/(2\gamma_k)$ on $(0, \infty)$. A straightforward calculation shows that $x^{k+1} = (\alpha\gamma_k)^{\frac{1}{\alpha+2}}$ and, hence,

$$f(x^{k+1}) = \mathcal{O}\left(\frac{1}{\gamma_k^d}\right) \quad \text{with} \quad d = \frac{\alpha}{\alpha+2}.$$

In particular, we see that d becomes arbitrarily small if $\alpha \downarrow 0$. This shows that the convergence rate from Theorem 2.3 only holds if f attains its minimum. (Note that the constant hidden in $\mathcal{O}$ depends on α, but is independent of k for any fixed α.) $\Diamond$

We now turn to the discussion of weak convergence vs. strong convergence (again in the context of Theorem 2.3). Clearly, we need an example which is infinite-dimensional and, in view of Theorem 2.4, not strongly convex.

Example 3.3. *Let $f : \ell^2 \to \mathbb{R}$ be given by*

$$f(x) = \sum_{i=1}^{\infty} \frac{1}{2i} x_i^2,$$

and let $u^k = e^k$ be the sequence of unity vectors. Note that f is well-defined and convex (not strongly convex). The minimization problem (5) *can be solved analytically. The solution is given by*

$$x^{k+1} = \frac{k}{k+\gamma_k} e^k.$$

Now, let $\gamma_k = k$ for all k. Then $x^{k+1} = \frac{1}{2} e^k$ and $f(x^{k+1}) = \frac{1}{8k}$. This shows that (x^k) converges only weakly to $x^* = 0$. Furthermore, the convergence rate from Theorem 2.3 can, in general, not be strengthened. ♢

This concludes our set of examples for Theorem 2.3. Hence, we now turn to Theorem 2.4, which deals with the strongly convex case. The following (fairly trivial) example shows that the convergence rates stated in the theorem can, in general, not be strengthened.

Example 3.4. *Let $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$. An easy calculation shows that, for any $u^k \in \mathbb{R}$ and $\gamma_k > 0$, the solution of the subproblem* (5) *is given by*

$$x^{k+1} = \frac{u^k}{2\gamma_k + 1}.$$

If (u^k) is bounded, it follows that

$$\|x^{k+1} - x^*\| = \mathcal{O}\left(\frac{1}{\gamma_k}\right) \quad \text{and} \quad f(x^{k+1}) - f(x^*) = \mathcal{O}\left(\frac{1}{\gamma_k^2}\right),$$

as asserted by Theorem 2.4. ♢

4 The Augmented Lagrangian Method

We now turn to a discussion of augmented Lagrangian-type methods and their relationship with the (generalized) proximal-point method. To this end, we consider an optimization problem of the form

$$\min\ F(x) \quad \text{s.t.} \quad g(x) \in K,\ x \in C, \tag{8}$$

where

- $F : X \to \overline{\mathbb{R}}$ is proper and convex,
- X is a Banach space and $C \subseteq X$ a nonempty, closed, convex set,
- Y is a Hilbert space and $K \subseteq Y$ a nonempty, closed, convex cone,
- $g : X \to Y$ is $\mathcal{C}^1$ and concave with respect to the order induced by K.

We denote by K^+ and K° the dual and polar cones of K, respectively.

The function g occurring in (8) is occasionally called a *concave operator*. For some results and further reading about this generalized form of concavity (and convexity), we refer the reader to [8, 18]. Here, we only mention that many well-known results for real-valued convex functions remain true for their operator counterparts. For instance, the composition of an increasing convex operator and a convex operator is always convex (this implies the convexity of the Lagrangian and augmented Lagrangian, see below). On the other hand, differentiability and subdifferentiability are, in general, much more peculiar to analyse. For instance, the composition

$$\lambda \circ h, \quad \text{where} \quad \lambda \in K^+, \quad h : X \to Y \text{ convex},$$

is convex, but generally does not admit a chain rule involving the subdifferential of h. Additional assumptions such as regular subdifferentiability [3, 25] are needed for this. To circumvent such issues and prevent our analysis from becoming overly technical, we decided to simply assume that our constraint function g is of class $\mathcal{C}^1$.

We now turn to a description of the augmented Lagrangian (or multiplier-penalty) method. The augmented Lagrangian of the problem (8) is given by

$$L_\gamma(x, u) = F(x) + \frac{\gamma}{2} \left\| P_{K^\circ} \left(g(x) + \frac{u}{\gamma} \right) \right\|^2,$$

where P is the projection operator; note that this is only a partial augmented Lagrangian since the potentially easy constraints $x \in C$ are not included in our definition of L_γ. The *modified augmented Lagrangian method* (cf. [7] for a finite-dimensional version) consists of the iterative procedure

$$x^{k+1} \in \operatorname*{argmin}_{x \in C} L_{\gamma_k}(x, u^k) \quad \text{and} \quad \lambda^{k+1} = P_{K^\circ}\big(u^k + \gamma_k g(x^{k+1})\big), \tag{9}$$

where $(u^k) \subseteq Y$ is a bounded sequence and (γ_k) is a sequence of real numbers, typically converging to $+\infty$. The *standard augmented Lagrangian method* with the Hestenes-Stiefel-Rockafellar update of the multiplier λ^{k+1} corresponds to the special case $u^k := \lambda^k$, cf. [5, 21]. In this standard method, however, the sequence (u^k) is not necessarily bounded.

To establish the connection of the iterative scheme (9) with our generalized proximal-point method, we denote by $q : Y \to \overline{\mathbb{R}}$ the (Lagrange-) dual function of the problem, i.e.

$$q(\lambda) = \inf_{x \in C} L(x, \lambda),$$

where $L(x, \lambda) := F(x) + \langle \lambda, g(x) \rangle$ is the usual Lagrange function. Then the dual problem is given by

$$\max \ q(\lambda) \quad \text{s.t.} \quad \lambda \in K^\circ,$$

which can be written as a minimization problem

$$\min \ \tilde{q}(\lambda) \quad \text{s.t.} \quad \lambda \in K^\circ, \tag{10}$$

by defining $\tilde{q} := -q$. Recall that $\tilde{q}$ is a convex function and that, in this section, the minimization problem (10) plays the role of the optimization problem (1).

Theorem 4.1. *Let $u \in Y$ and $\gamma > 0$. Furthermore, let $\bar{x}$ be a minimum of $L_\gamma(\,\cdot\,, u)$, $\bar{\lambda} = P_{K^\circ}(u + \gamma g(\bar{x}))$, and $\bar{\mu} = \operatorname{prox}_{\gamma\tilde{q}}(u)$. Then $\bar{\mu} = \bar{\lambda}$, and $\bar{x}$ is a point where the infimum $q(\bar{\lambda})$ is attained.*

Proof. We first claim that $(\bar{x}, \bar{\lambda})$ is a saddle point of the convex-concave function

$$h : X \times K^\circ \to \mathbb{R}, \quad h(x, \lambda) = L(x, \lambda) - \frac{1}{2\gamma}\|\lambda - u\|^2.$$

To verify this saddle-point property, note that the definition of $\bar{x}$ implies that

$$0 \in \partial_x L_\gamma(\bar{x}, u) = \partial F(\bar{x}) + g'(\bar{x})^* \bar{\lambda} = \partial_x h(\bar{x}, \bar{\lambda}),$$

where we used [4, Prop. 12.31] to differentiate the projection operator. Hence, $\bar{x}$ is a minimizer of the convex function $h(\,\cdot\,, \bar{\lambda})$. On the other hand, $h(\bar{x}, \,\cdot\,)$ is a quadratic function of the form

$$h(\bar{x}, \lambda) = \langle \lambda, g(\bar{x}) \rangle - \frac{1}{2\gamma}\|\lambda - u\|^2 + c,$$

where c is a constant independent of λ. Since

$$\langle \lambda, g(\bar{x}) \rangle - \frac{1}{2\gamma}\|\lambda - u\|^2 = -\frac{1}{2\gamma}\|\lambda - u - \gamma g(\bar{x})\|^2 + \tilde{c},$$

where $\tilde{c}$ is again independent of λ, we see that the (unique) maximizer of $h(\bar{x}, \,\cdot\,)$ on K° is $\bar{\lambda} = P_{K^\circ}(u + \gamma g(\bar{x}))$. This proves the saddle-point property of $(\bar{x}, \bar{\lambda})$.

A standard saddle point theorem, see, e.g., [3, Prop. 2.105], implies that $(\bar{x}, \bar{\lambda})$ satisfies

$$h(\bar{x}, \bar{\lambda}) = \max_{\lambda \in K^\circ} h(\bar{x}, \lambda) = \max_{\lambda \in K^\circ} \min_{x \in X} h(x, \lambda). \tag{11}$$

On the other hand, $\bar{\mu}$ is characterized by

$$\begin{aligned}
\bar{\mu} &= \operatorname{prox}_{\gamma\tilde{q}}(u) \\
&= \operatorname*{argmin}_{\lambda \in K^\circ} \left\{ \tilde{q}(\lambda) + \frac{1}{2\gamma}\|\lambda - u\|^2 \right\} \\
&= \operatorname*{argmax}_{\lambda \in K^\circ} \left\{ q(\lambda) - \frac{1}{2\gamma}\|\lambda - u\|^2 \right\} \\
&= \operatorname*{argmax}_{\lambda \in K^\circ} \left\{ \inf_{x \in X} L(x, \lambda) - \frac{1}{2\gamma}\|\lambda - u\|^2 \right\} \\
&= \operatorname*{argmax}_{\lambda \in K^\circ} \left\{ \inf_{x \in X} h(x, \lambda) \right\},
\end{aligned}$$

so that $\bar{\mu}$ is also the solution of

$$\max_{\lambda \in K^\circ} \min_{x \in X} h(x, \lambda).$$

Using (11), the uniqueness of $\bar{\mu}$ implies $\bar{\mu} = \bar{\lambda}$, and the statement follows. □

Using an induction argument, Theorem 4.1 implies the following: if we initialize the standard augmented Lagrangian method (where $u^k = \lambda^k$ for all $k \geq 1$) with the same λ^1 as the classical proximal-point method applied to the corresponding dual problem, then both methods generate the same sequence (λ^k). This is the known relation between these two methods. In addition, Theorem 4.1 also shows that the same connection holds between the modified augmented Lagrangian method (with an arbitrary, bounded sequence (u^k)) and the generalized proximal-point method from Algorithm 2.1. Hence, all convergence results from the previous section hold for the generalized proximal-point method applied to the dual problem and, therefore, also yield convergence and rate-of-convergence results for the corresponding modified augmented Lagrangian method applied to the underlying primal problem.

5 Final Remarks

We have considered a generalization of the well-known proximal-point and Tikhonov regularization methods for convex optimization problems. Among other results, we have proved that the new method is essentially equivalent to the modified augmented Lagrangian algorithm from [7] (or, more precisely, a generalization of that method to 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] R. Andreani, E. G. Birgin, J. M. Martínez, and M. L. Schuverdt. Augmented Lagrangian methods under the constant positive linear dependence constraint qualification. *Math. Program.*, 111(1-2, Ser. B):5–32, 2008.

[3] V. Barbu and T. Precupanu. *Convexity and Optimization in Banach Spaces.* Springer Monographs in Mathematics. Springer, Dordrecht, fourth edition, 2012.

[4] 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, New York, 2011.

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

[6] E. G. Birgin, C. A. Floudas, and J. M. Martínez. Global minimization using an augmented Lagrangian method with variable lower-level constraints. *Math. Program.*, 125(1, Ser. A):139–162, 2010.

[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] J. M. Borwein. Continuity and differentiability properties of convex operators. *Proc. London Math. Soc. (3)*, 44(3):420–444, 1982.

[9] A. R. Conn, N. I. M. Gould, and P. L. Toint. *LANCELOT*. Springer-Verlag, Berlin, 1992. A Fortran package for large-scale nonlinear optimization (release A).

[10] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[11] O. Güler. On the convergence of the proximal point algorithm for convex minimization. *SIAM J. Control Optim.*, 29(2):403–419, 1991.

[12] O. Güler. New proximal point algorithms for convex minimization. *SIAM J. Optim.*, 2(4):649–664, 1992.

[13] A. N. Iusem. Augmented Lagrangian methods and proximal point methods for convex optimization. *Investigaci'on Operativa*, pages 11–50, 1999.

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

[15] C. Kanzow and D. Steck. Augmented Lagrangian methods for the solution of generalized Nash equilibrium problems. *SIAM J. Optim.*, 26(4):2034–2058, 2016.

[16] C. Kanzow and D. Steck. Augmented Lagrangian and exact penalty methods for quasi-variational inequalities. *Technical Report*, Institute of Mathematics, University of Würzburg, June 2016.

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

[18] A. G. Kusraev and S. S. Kutateladze. *Subdifferentials: Theory and Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995.

[19] B. Martinet. Régularisation d'inéquations variationnelles par approximations successives. *Rev. Française Informat. Recherche Opérationnelle*, 4(Ser. R-3):154–158, 1970.

[20] B. Martinet. Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox. *C. R. Acad. Sci. Paris Sér. A-B*, 274:A163–A165, 1972.

[21] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operations Research. Springer-Verlag, New York, 1999.

[22] R. T. Rockafellar. A dual approach to solving nonlinear programming problems by unconstrained optimization. *Math. Programming*, 5:354–373, 1973.

[23] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.

[24] A. N. Tikhonov and V. A. Morozov. Methods for the regularization of ill-posed problems. *Vychisl. Metody i Programmirovanie*, (35):3–34, 1981.

[25] J. Zowe. Subdifferentiability of convex functions with values in an ordered vector space. *Math. Scand.*, 34:69–83, 1974.