Strong Convergence in Hilbert Spaces via Γ-Duality

M. Marques Alves* J. G. Melo†

Communicated by Alfredo N. Iusem

Abstract

We analyze a primal-dual pair of problems generated via a duality theory introduced by Svaiter. We propose a general algorithm and study its convergence properties. The focus is a general primal-dual principle for strong convergence of some classes of algorithms. In particular, we give a different viewpoint for the weak-to-strong principle of Bauschke and Combettes and unify many results concerning weak and strong convergence of subgradient type methods.

Key words: Γ-duality · Hilbert spaces · convex feasibility · strong convergence · subgradient method.

2000 Mathematics Subject Classification: 49N15 · 47N10 · 65K10.

1 Introduction

Infinite dimensional models arise in different fields of applied mathematics, including optimization [1, 2], partial differential equations [3], optimal control [4] and inverse problems [5]. One of the main difficulties when moving from finite to infinite dimensions is the lack of compactness of closed bounded sets, in the topology generated by the norm of the underlying normed space, which in particular implies that one cannot extract convergent subsequences of norm-bounded sequences. This fact imposes several limitations in the convergence analysis of algorithms and approximations methods in such a general setting. A remedy to overcome this drawback is to replace the topology generated by the norm by

*Universidade Federal de Santa Catarina, Departamento de Matemática, 88040-900, Florianópolis, Brazil (maicon@impa.br) (Corresponding author)

†Universidade Federal de Goiás, Instituto de Matemática e Estatística, 74001-970, Goiânia, Brazil (jefferson@mat.ufg.br)

a weak topology, generated by continuous linear functionals [3]. This leads to a notion of weak convergence for sequences and, for a special class of normed spaces, one are able to get compactness. Among such spaces, special attention is given for models in Hilbert spaces where weak convergence of sequences is completely characterized via the Riesz-Fréchet representation theorem by the inherent inner product [3, Theorem 5.5].

As expected, weak and strong convergence are only distinguishable in infinite dimensional spaces. A canonical example of a sequence which is weakly but not strongly convergent is given by considering a sequence formed by the elements of an orthonormal basis of a Hilbert space [3]. A more concrete example appears in the study of proximal point methods for finding zeroes of maximal monotone operators. Rockafellar [6] proved that the proximal point method is weakly convergent, under suitable conditions, and posed the question whether it converges strongly in infinite dimensional spaces. Although Rockafellar's question has been answered negatively by Güler [7], it has been posed the question whether weakly convergent methods (in particular proximal point methods) can be appropriately modified in order to guarantee strong convergence. The proximal point method was modified by Solodov and Svaiter in [8], obtaining in this way a strongly convergent method in infinite dimensional Hilbert spaces. It combines inexact proximal steps with projections onto the intersection of two half spaces generated by the current information at each step. In [9], Bauschke and Combettes proposed a weak-to-strong convergence principle for modifying (weakly convergent) Fejér-monotone type methods in order to obtain strongly convergent ones. Their abstract framework is based on a special class of operators and encompasses exact proximal point methods, constraint disintegration methods, subgradient methods, among others.

Duality theory is a powerful and very useful tool in constrained optimization, raised with the duality theory for linear programming and game theory [10]. Here, we are interested in the Γ-duality theory, which was proposed and studied in [11] and further developed and used in [12], where the author (among other results) formulated strongly convergent proximal point methods for finding zeroes of maximal monotone operators in Hilbert spaces.

In this paper, we go a step further and inspect additional applications of Γ-duality for obtaining strongly convergent methods in infinite dimensional Hilbert spaces. Precisely, we consider a general convex feasibility problem and, by means of Γ-duality, we propose and analyze a general algorithm for solving a variational formulation of this feasibility problem. Motivated by [11, 12], we also analyze two variants of the general algorithm. The first one has weak convergence properties, while the second one is strongly convergent. The main focus is to propose a weak-to-strong convergence principle for solving general feasibility problems. As indicated in [12], this will be done by enforcing asymptotic complementarity of the primal dual sequence generated by the algorithm. Since [9] also concerns a weak-to-strong convergence principle, their results will be compared to ours, and we present an example showing that general algorithm presented here cannot be obtained by the corresponding one presented in [9]. In particular, we discuss how subgradient type methods can be seen as a realization of one of the algorithms presented in this paper. We also show that our general setting encompasses a special case of a strongly convergent subgradient type method, recently proposed by Bello Cruz and Iusem [13], for solving a wide class of constrained optimization problems. Additionally, we characterize Fejér-monotone methods in the setting of Γ-duality. As we will see, this property is related to the increasing property of the dual function.

This paper is organized as follows. In Section 2, we present the results concerning Γ-duality that will be used in this work. We also recall in this section some basic results of the Bauschke-Combettes' weak-to-strong convergence principle. In Section 3, the first algorithm is proposed and analyzed. In Section 4, we propose the second algorithm of this paper, and relate it with a general algorithm proposed in [9]. We also state and prove a weak-to-strong convergence principle based on Γ-duality setting. Section 5 is devoted to show how subgradient type methods fit in the general setting presente here.

2 Preliminaries and Notations

We denote by $\mathcal{H}$ a (nontrivial) real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. The orthogonal projection onto a closed and convex set C is denoted by P_C. We denote by $\mathcal{B}(x^k)$ the set of weak cluster points of a sequence $\{x^k\}$. The extended-real number system is $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. The *convex conic hull* of a subset D of a vector space is denoted by $\operatorname{conv}(\operatorname{cone}) D$. The (Fréchet) derivative of a function $f : \mathcal{H} \to \mathbb{R}$ is denoted by ∇f.

2.1 Basic Facts on Γ-Duality

Let X be nonempty, $D \subset X$ and $f : X \to \overline{\mathbb{R}}$. Consider the following constrained optimization problem

$$\begin{cases} \min f(x) \\ \text{s.t. } \ x \in D. \end{cases} \tag{1}$$

Let Γ be a set of extended-real valued functions on X. The dual of (1) in the sense of Γ-duality [11, 12] is

$$\begin{cases} \max \varphi(\xi) \\ \text{s.t. } \ \xi \in \Gamma(D), \end{cases} \tag{2}$$

where $\varphi : \Gamma \to \overline{\mathbb{R}}$,

$$\varphi(\xi) = \inf_{x \in X} \ f(x) + \xi(x) \tag{3}$$

is the dual function of f and

$$\Gamma(D) = \{\xi \in \Gamma : \xi(x) \leq 0, \forall x \in D\}$$

is the dual feasible set.

As pointed out by [11, 12], the function $(x, \xi) \mapsto f(x)+\xi(x)$ does not contain information of the primal feasible set D. Instead, it is incorporated into the dual feasible set $\Gamma(D)$.

A compreensive account of this duality theory is presented and studied in [11] and further developed and exploited in [12]. In this Section, we will only present the results of Γ-duality that will be needed in this paper.

Definition 2.1. (General KKT Conditions)

A pair $(\bar{x}, \bar{\xi}) \in X \times \Gamma$ is said to satisfy the KKT conditions if and only if

a) $\bar{x} \in \arg\min_{x \in X} f(x) + \bar{\xi}(x)$ *(Lagrangian condition);*

b) $\bar{x} \in D$ *(primal feasibility);*

c) $\bar{\xi} \in \Gamma(D)$ *(dual feasibility);*

d) $\bar{\xi}(\bar{x}) = 0$ *(complementarity).*

The following proposition relates the KKT conditions with primal-dual solutions; see [12, Proposition 4.3].

Proposition 2.1. *Let $(\bar{x}, \bar{\xi}) \in X \times \Gamma$ be such that $f(\bar{x}) + \bar{\xi}(\bar{x})$ is finite. Then $(\bar{x}, \bar{\xi})$ satisfies the general KKT conditions if and only if*

a) $\bar{x}$ and $\bar{\xi}$ are, respectively, solutions of the primal and dual problems;

b) There is no duality gap and the primal optimal value is finite.

As a consequence, $\varphi(\bar{\xi}) = f(\bar{x})$.

From now on, the primal space X is a real Hilbert space, and denoted by $\mathcal{H}$. The next proposition and its corollary are related to [12, Lemma 6.4], where continuous affine functions are used to define the primal feasible set D.

Let $C \subset \mathcal{H}$ be nonempty, Γ be the class of continuously differentiable convex functions on $\mathcal{H}$, and $f, g_i \in \Gamma$, $i = 1, 2, \cdots, m$. Suppose that $C \subset D$, where $D := \{x \in \mathcal{H} : g_i(x) \leq 0, \ i = 1, \cdots, m\}$, and let $\Sigma := \text{conv}(\text{cone})\{g_1, \cdots, g_m\}$. It follows directly from the definition of Σ and $\Gamma(D)$ that $\Sigma \subset \Gamma(D)$.

Proposition 2.2. *If the primal problem* (1)*, with f and D as above, has a solution $\bar{x} \in D$ such that $(\bar{x}, \bar{\mu})$ satisfies KKT conditions* (*in the usual sense*) *for some $\bar{\mu} \in \mathbb{R}^m_+$, then the dual problem* (2) *has a solution $\bar{\xi} \in \Sigma$ satisfying*

$$-\infty < \varphi(\bar{\xi}) = \max_{\xi \in \Sigma} \varphi(\xi) = \max_{\xi \in \Gamma(D)} \varphi(\xi) = f(\bar{x}) < \infty.$$

Additionally, $\bar{\xi} \in \Gamma(C)$.

Proof. Consider $\bar{x} \in D$ solution of (1) and $\bar{\mu} \in \mathbb{R}^m_+$ such that $(\bar{x}, \bar{\mu})$ satisfies the KKT conditions (see, for instance, [2, Proposition 26.18]). Therefore

a) $\nabla f(\bar{x}) + \sum_{i=1}^m \bar{\mu}_i \nabla g_i(\bar{x}) = 0\,;$

b) $\bar{\mu}_i \geq 0,\ i = 1, 2, \cdots, m\,;$

c) $\bar{\mu}_i\, g_i(\bar{x}) = 0,\ i = 1, 2, \cdots, m\,.$

Defining $\bar{\xi} := \sum_{i=1}^m \bar{\mu}_i\, g_i$, we claim that the pair $(\bar{x}, \bar{\xi})$ satisfies the KKT conditions in the sense of Γ-duality (see Definition 2.1). Indeed, from b) and convexity of $f, g_i, i = 1, \cdots, m$, it follows that $f(\cdot) + \bar{\xi}(\cdot)$ is convex, and from a) we obtain that $\bar{x} \in \operatorname{argmin}_{x \in \mathcal{H}} f(x) + \bar{\xi}(x)$. Note that $f(\bar{x}) + \bar{\xi}(\bar{x})$ is finite.

By definition of $\bar{\xi}$ and c), we obtain $\bar{\xi} \in \Sigma$ and $\bar{\xi}(\bar{x}) = 0$. Since $\Sigma \subset \Gamma(D)$, it follows that $\bar{\xi} \in \Gamma(D)$ and so the claim holds. Therefore, from Proposition 2.1, $(\bar{x}, \bar{\xi})$ is a solution of the primal dual problems (1)-(2) and there is no duality gap. Since $\bar{\xi} \in \Sigma \subset \Gamma(D)$, it follows that

$$\varphi(\bar{\xi}) \leq \sup_{\xi \in \Sigma} \varphi(\xi) \leq \sup_{\xi \in \Gamma(D)} \varphi(\xi) = \varphi(\bar{\xi}) = f(\bar{x}).$$

Finally, the last assertion of the proposition follows from the fact that $C \subset D$ implies $\Gamma(D) \subset \Gamma(C)$. □

In the next corollary, Γ denotes the set of affine continuous functions on $\mathcal{H}$, i.e., any element ξ of Γ is of the form $\xi(x) = \langle u, x \rangle + b$, where $u \in \mathcal{H}$ and $b \in \mathbb{R}$. Note that $\xi(x) = \langle \nabla \xi, x \rangle + \xi(0)$, where $\nabla \xi$ does not depend on the point x.

Corollary 2.1. *Let $x^0 \in \mathcal{H}$ and $\{g_i\}_{i=1}^m \subset \Gamma$. Let $\bar{x}$ be the solution of* (1) *with $f(x) = (1/2)\|x - x^0\|^2$ and $D = \{x \in \mathcal{H} \,:\, g_i(x) \leq 0,\ i = 1, 2, \cdots, m\}$.*

Let $\bar{\xi}$ be given by Proposition 2.2. Then

1. $\bar{x} = x^0 - \nabla\, \bar{\xi}\,;$

2. $\bar{\xi}(x) = \langle x^0 - \bar{x}, x - \bar{x} \rangle,\ \forall x \in \mathcal{H}$. *In particular,* $\bar{\xi}(\bar{x}) = 0\,;$

3. $\bar{\xi} = \arg\max_{\xi \in \Sigma} \varphi(\xi)$, *where* $\Sigma = \operatorname{conv}(\operatorname{cone})\, \{g_1, \cdots, g_m\}$.

Proof. First, since g_i, $i = 1, \cdots, n$, is affine, canonical application of Farkas's Lemma [14] ensures the existence of $\bar{\mu} \in \mathbb{R}^m_+$ such that $(\bar{x}, \bar{\mu})$ satisfies the KKT conditions (in the usual sense). Therefore, by Propositions 2.2 and 2.1, we see that $\bar{\xi}(\bar{x}) = 0$ and $\bar{x} = \arg\min_{x \in \mathcal{H}} f(x) + \bar{\xi}(x)$. In particular, item 1 holds. Now, we obtain

$$\bar{\xi}(x) = \langle \nabla \bar{\xi}, x \rangle + \bar{\xi}(0) = \langle x^0 - \bar{x}, x \rangle + \bar{\xi}(0). \tag{4}$$

Since $\bar{\xi}(\bar{x}) = 0$, it follows from (4) that $\bar{\xi}(x) = \langle x^0 - \bar{x}, x - \bar{x} \rangle$, proving item 2. For proving item 3, we observe that, by Proposition 2.2, $\bar{\xi} \in \arg\max_{\xi \in \Sigma} \varphi(\xi)$.

Therefore, we just need to show that $\bar{\xi}$ is the unique affine function maximizing φ over Σ. Indeed, let η be an affine function with such a property. By Propositions 2.2 and 2.1, $(\bar{x}, \eta)$ satisfies the general KKT conditions, and similar to the proof of item 1, we have $\bar{x} = x^0 - \nabla\eta$. Using that η is affine and satisfies $\eta(\bar{x}) = 0$, we obtain $\eta = \bar{\xi}$. Thus, item 3 holds, concluding the proof of the corollary. □

2.2 Basic Facts on the Weak-to-Strong Convergence Principle of Bauschke and Combettes

In [9], a weak-to-strong convergence principle was introduced in order to force strong convergence of weakly convergent Fejér-monotone methods in infinite dimensional Hilbert spaces. Strongly convergent variants of subgradient and proximal point methods were proposed and analyzed for solving optimization problems, convex feasibility and monotone inclusion problems.

In what follows we summarize some of the main convergence results obtained in [9], which will be useful in this paper. We adopt the same basic notation of [9].

We denote by $\mathrm{Fix}(T)$ the set of fixed points of an operator $T : \mathcal{H} \to \mathcal{H}$ and by P_C the orthogonal projector onto a nonempty closed and convex set C. We recall the following characterization of P_C:

$$z = P_C(x) \iff \langle x - z, y - z\rangle \le 0, \quad \forall y \in C. \tag{5}$$

For each $x, y \in \mathcal{H}$ define

$$H(x, y) := \{z \in \mathcal{H} : \langle z - y, x - y\rangle \le 0\}.$$

Note that $H(x, x) = \mathcal{H}$ and, if $x \neq y$, $H(x, y)$ is a closed affine half space for which $y = P_{H(x,y)}(x)$. We also denote by $Q(x, y, z)$ the (orthogonal) projection of x onto $H(x, y) \cap H(y, z)$, i.e.,

$$Q(x, y, z) := P_{H(x,y)\cap H(y,z)}(x).$$

Definition 2.2. *Let $\mathcal{T}$ be the class of operators $T : \mathcal{H} \to \mathcal{H}$ satisfying*

$$\mathrm{dom}(T) = \mathcal{H}, \text{ and } \mathrm{Fix}(T) \subset H(x, Tx), \ \forall x \in \mathcal{H}.$$

See [9, Proposition 2.3] for examples of operators in the family $\mathcal{T}$. Next, we present two algorithms which were proposed and analyzed in [9]. The first one generates weakly convergent sequences, while the second one has strong convergence properties.

Algorithm 2.1.

0) Let $\varepsilon \in]0, 1]$ and $x^0 \in \mathcal{H}$. Set $k = 0$;

1) Select $T_k \in \mathcal{T}$;

2) Set $x^{k+1} = x^k + (2-\varepsilon)(T_k\, x^k - x^k)$, $k := k+1$ and go to 1.

Algorithm 2.2.

0) Let $x^0 \in \mathcal{H}$. Set $k = 0$;

1) Select $T_k \in \mathcal{T}$. If $H(x^0, x^k) \cap H(x^k, T_k\, x^k) = \emptyset$ stops, otherwise,

2) Set $x^{k+1} = Q(x^0, x^k, T_k\, x^k)$, $k := k+1$ and go to 1.

Remark 2.1. Let $C := \bigcap_{k\geq 0} Fix(T_k)$. It was proved in [9, Proposition 3.4] that, if $C \neq \emptyset$, then Algorithm 2.2 generates an infinite sequence $\{x^k\}$ such that

$$C \subset H(x^0, x^k) \cap H(x^k, T_k x^k), \qquad \forall k \in \mathbb{N}.$$

Next, we summarize the main properties of Algorithms 2.1 and 2.2, which will be useful in this paper. For details see [9, Theorem 2.9] and [9, Theorem 3.5].

Proposition 2.3. *Let $\{x^k\}$ be a sequence generated by Algorithm 2.1. Suppose that $C := \bigcap_{k\geq 0} \mathrm{Fix}(T_k) \neq \emptyset$. Then $\{x^k\}$ is bounded and it holds*

$$x^k \rightharpoonup x \in C \iff \mathcal{B}(x^k) \subset C.$$

Proposition 2.4. *Let $\{x^k\}$ be a sequence generated by Algorithm 2.2 and suppose that $C := \bigcap_{k\geq 0} \mathrm{Fix}(T_k) \neq \emptyset$. Then $\{x^k\}$ is bounded and*

$$x^k \to P_C(x^0) \iff \mathcal{B}(x^k) \subset C.$$

The weak-to-strong convergence principle of [9] depends on the concept of *coherent* sequence of operators in $\mathcal{T}$.

Definition 2.3. *A sequence $\{T_k\} \subset \mathcal{T}$ is* coherent *if and only if, for every bounded sequence $\{y^k\} \subset \mathcal{H}$, there holds*

$$\begin{cases} \sum_{k=0}^{\infty} \|y^{k+1} - y^k\|^2 < \infty \\ \sum_{k=0}^{\infty} \|y^k - T_k\, y^k\|^2 < \infty \end{cases} \Rightarrow \mathcal{B}(y^k) \subset \bigcap_{k\geq 0} \mathrm{Fix}(T_k).$$

Next, we present the weak-to-strong convergence principle [9, Theorem 4.2].

Theorem 2.3. *Let $\{T_k\} \subset \mathcal{T}$ be coherent and let $C := \bigcap_{k\geq 0} \mathrm{Fix}(T_k)$. Then:*

i) If $C \neq \emptyset$, then every sequence generated by Algorithm 2.1 converges weakly to a point in C;

ii) (Trichotomy) Let $\{x^k\}$ be a sequence generated by Algorithm 2.2. One and only one of the following alternatives holds:

(a) $C \neq \emptyset$ and $x^k \to P_C(x^0)$;

(b) $C = \emptyset$ and $\|x^k\| \to \infty$;

(c) $C = \emptyset$ and the algorithm terminates.

3 Fejér-Monotone Algorithms and Γ-Duality

Throughout this section Γ denotes the set of continuous affine functions on $\mathcal{H}$. Recall that any element ξ of Γ is of the form $\xi(x) = \langle \nabla\xi, x\rangle + \xi(0)$, where $\nabla\xi$ does not depend on the point x.

Let $C \subset \mathcal{H}$ be nonempty, closed and convex. We consider the following problem

$$\text{Find } x \in \mathcal{H} \text{ such that } x \in C. \tag{6}$$

A variational formulation of (6) consists in finding in C the closest point of a given point $x^0 \in \mathcal{H}$. This is equivalent to solve the following (primal) optimization problem

$$\begin{cases} \min \frac{1}{2} \|x - x^0\|^2 \\ \text{s.t. } \ x \in C. \end{cases} \tag{7}$$

Problem (6) (or (7)) is fundamental in many areas of applied mathematics, including convex optimization, fixed point theory, monotone operator theory, among others; see, for instance, [2, 8, 9] and references therein.

A wide class of approximation methods for solving (7) are known as *Fejér-monotone* algorithms [9]. Recall that a sequence $\{x^k\}$ in $\mathcal{H}$ is *Fejér-monotone* with respect to (w.r.t.) $S \subset \mathcal{H}$ if and only if

$$\|x^{k+1} - x\| \leq \|x^k - x\|, \qquad \forall x \in S, \ \forall k \in \mathbb{N}. \tag{8}$$

A very useful result to prove (weak) convergence of sequences in Hilbert spaces is given in the following lemma [9].

Lemma 3.1 (Browder)**.** *Let $C \subset \mathcal{H}$ be nonempty, closed and convex. Suppose that $\{x^k\}$ is Fejér-monotone w.r.t. C. Then*
$x^k \rightharpoonup x \in C$ if and only if $\mathcal{B}(x^k) \subset C$.

In this Section, we describe Fejér-monotone methods in terms of Γ-duality theory. Firstly, we prove some technical results concerning the dual problem of (7) (in the sense of Γ-duality). Recall that the dual problem of (7) is

$$\begin{cases} \max \varphi(\xi) \\ \text{s.t. } \ \xi \in \Gamma(C), \end{cases}$$

where $\Gamma(C) = \{\gamma \in \Gamma \,:\, \gamma(x) \leq 0, \ \forall x \in C\}$ is the dual feasible set and

$$\varphi(\xi) = \inf_{x\in\mathcal{H}} \frac{1}{2} \|x - x^0\|^2 + \xi(x) \tag{9}$$

is the dual function.

The proof of the following result follows by direct calculations.

Lemma 3.2. *Let $x^0 \in \mathcal{H}$ and $\xi \in \Gamma$. Then $x^0 - \nabla\xi$ is the unique solution of*

$$\min_{x\in\mathcal{H}} \frac{1}{2}\|x - x^0\|^2 + \xi(x).$$

In the following proposition, we analyze the increasing directions of the dual function φ. It will be useful to characterize Fejér-monotone properties of the algorithms presented here (see Proposition 3.3). In [12, Lema 6.3], a related result concerning the problem of finding zeroes of maximal monotone operators is presented.

Proposition 3.1. *Let $\xi, \gamma \in \Gamma$, $t > 0$ and $z := \arg\min_{x\in\mathcal{H}} (1/2)\|x - x^0\|^2 + \xi(x)$. It holds*

$$\varphi(\xi + t\gamma) = \varphi(\xi) + t\left(\gamma(z) - \frac{t}{2}\|\nabla\gamma\|^2\right).$$

Proof. Let $\hat{z} := \arg\min\limits_{x\in\mathcal{H}}(1/2)\|x - x^0\|^2 + (\xi + t\gamma)(x)$. Therefore

$$\varphi(\xi + t\gamma) = \frac{1}{2}\|\hat{z} - x^0\|^2 + \xi(\hat{z}) + t\gamma(\hat{z}).$$

Using Lemma 3.2 and the fact $\nabla(\xi + t\gamma) = \nabla\xi + t\nabla\gamma$, we obtain

$$\begin{aligned}
\varphi(\xi + t\gamma) \quad &= \quad \tfrac{1}{2}\|x^0 - (\nabla\xi + t\nabla\gamma) - x^0\|^2 + (\xi + t\gamma)(x^0 - \nabla\xi - t\nabla\gamma) \\
&= \quad \tfrac{1}{2}\|z - x^0\|^2 + \tfrac{1}{2}t^2\|\nabla\gamma\|^2 + t\langle\nabla\xi, \nabla\gamma\rangle \\
&\quad + \quad \xi(z - t\nabla\gamma) + t\gamma(z - t\nabla\gamma).
\end{aligned} \tag{10}$$

Since ξ and γ are affine functions, it follows that

$$\xi(z - t\nabla\gamma) + t\gamma(z - t\nabla\gamma) = \xi(z) - t\langle\nabla\xi, \nabla\gamma\rangle + t\gamma(z) - t^2\langle\nabla\gamma, \nabla\gamma\rangle.$$

Combining this with (10) we obtain

$$\begin{aligned}
\varphi(\xi + t\gamma) \quad &= \quad (1/2)\|z - x^0\|^2 + \xi(z) + t\gamma(z) - \tfrac{1}{2}t^2\|\nabla\gamma\|^2 \\
&= \quad \varphi(\xi) + t(\gamma(z) - \tfrac{1}{2}t\|\nabla\gamma\|^2),
\end{aligned}$$

concluding the proof of the proposition. □

Remark 3.1. Proposition 3.1 shows that to increase the value of the dual function $\varphi(\xi)$ in the direction of γ, it should hold $\gamma(z) > \frac{t}{2}\|\nabla\gamma\|^2$, where $t > 0$, and $z = x^0 - \nabla\xi$.

Now we analyze Fejér-monotone algorithms in terms of Γ-duality. The next Proposition is related to [9, Proposition 2.7], where Fejér-monotone sequences are described in terms of iterations for operators in the family $\mathcal{T}$.

Proposition 3.2. *Let $\{x^k\} \subset \mathcal{H}$ be Fejér-monotone w.r.t. C. Consider*

$$\gamma^k(x) = \left\langle x - \left(\frac{x^k + x^{k+1}}{2}\right), \frac{x^k - x^{k+1}}{2} \right\rangle, \qquad k \in \mathbb{N}, x \in \mathcal{H}.$$

Set $\xi^0 = 0$ and $\xi^{k+1} = \xi^k + 2\,\gamma^k$. Then γ^k is dual feasible w.r.t. C, i.e., $\gamma^k \in \Gamma(C)$, $\forall k \in \mathbb{N}$. Moreover,

$$x^k = \arg\min_{x \in \mathcal{H}} \frac{1}{2}\|x - x^0\|^2 + \xi^k(x). \tag{11}$$

Proof. First, observe that for all $k \in \mathbb{N}$, γ^k is an affine function, i.e., $\gamma^k \in \Gamma$. Let $x \in C$. Since $\{x^k\}$ is Fejér-monotone w.r.t. C, we have

$$\|x^{k+1} - x\| \leq \|x^k - x\|, \;\; \forall k \in \mathbb{N}. \tag{12}$$

Therefore, $\forall k \in \mathbb{N}$,

$$\begin{aligned} \gamma^k(x) &= \left\langle x - \left(\tfrac{x^k + x^{k+1}}{2}\right), \tfrac{x^k - x^{k+1}}{2} \right\rangle \\ &= \tfrac{1}{4}\|x^{k+1} - x^k\|^2 + \tfrac{1}{2}\langle x^{k+1} - x^k, x^k - x\rangle \\ &= \tfrac{1}{4}\|x^{k+1} - x\|^2 - \tfrac{1}{4}\|x^k - x\|^2 \\ &\leq 0. \end{aligned}$$

Now, let us prove the last assertion of the proposition by induction on $k \in \mathbb{N}$. Since $\xi^0 = 0$, it follows that (11) holds for $k = 0$. Suppose (11) holds for some $k \in \mathbb{N}$. Using Lemma 3.2 we have $x^k = x^0 - \nabla\xi^k$. Direct calculations yields $\nabla\gamma^k = (1/2)(x^k - x^{k+1})$, implying $x^{k+1} = x^k - 2\nabla\gamma^k$. Thus,

$$x^{k+1} = (x^0 - \nabla\xi^k) - 2\nabla\gamma^k = x^0 - \nabla(\xi^k + 2\gamma^k) = x^0 - \nabla\xi^{k+1},$$

where in the last equality we used $\xi^{k+1} = \xi^k + 2\gamma^k$. Therefore, $x^{k+1} = x^0 - \nabla\xi^{k+1}$ and Lemma 3.2 implies that (11) holds also for $k+1$, concluding the proof. □

Corollary 3.1. *Let $\{T_k\}$ be a family of operators in $\mathcal{T}$ and $C := \bigcap_{k\in\mathbb{N}} \mathrm{Fix}(T_k)$. Let $\{x^k\}$ be a sequence generated by Algorithm 2.1 and γ^k, ξ^k be defined as in Proposition 3.2. Then,*

$$x^k = \arg\min_{x \in \mathcal{H}} \frac{1}{2}\|x - x^0\|^2 + \xi^k(x), \qquad \forall k \in \mathbb{N}.$$

Proof. Since by [9, Theorem 2.9], $\{x^k\}$ is Fejér-monotone w.r.t. C, the result follows as a direct consequence of Proposition 3.2.

□

Motivated by Algorithm 1 of [12] and by Proposition 3.2, we propose the following algorithm for solving (7) (or 6).

Algorithm A:

0) Let $\xi^0 = 0$, and $x^0 \in \mathcal{H}$. Set $k = 0$;

1) (Subproblem and Stopping Criterion)

a) Find $x^k = \arg\min_{x\in\mathcal{H}} \frac{1}{2}\|x - x^0\|^2 + \xi^k(x)$,

b) if $x^k \in C$ and $\xi^k(x^k) = 0$ then stops, otherwise,

2) Choose $\gamma^k \in \Gamma(C)$, $t_k > 0$ and define $\xi^{k+1} = \xi^k + t_k\gamma^k$;
Set $k := k + 1$ and go to **1**.

Some important remarks are in order.

Remark 3.2. Since $\xi^{k+1} = \xi^k + t_k\gamma^k$ and $\xi^0, \gamma^k \in \Gamma(C)$ *(*$\forall k \in \mathbb{N}$*)*, it follows that $\xi^k \in \Gamma(C)$, $\forall k \in \mathbb{N}$. In particular, using Proposition 2.1 one can conclude that, if the stopping criterion of Algorithm A is satisfied at some iteration k, then x^k is a primal solution of (7).

Remark 3.3. Combining Proposition 3.2 and Corollary 3.1, we obtain that Algorithm 2.1 is a special case of Algorithm A.

In what follows, we analyze Fejér-monotone properties of Algorithm A.

Proposition 3.3. *Let* $\{x^k\}$ *and* $\{\xi^k\}$ *be generated by Algorithm A. The following conditions are equivalent.*

1. $\varphi(\xi^{k+1}) \geq \varphi(\xi^k)$, $\forall k \in \mathbb{N}$*;*

2. $\gamma^k(x^k) \geq \frac{t_k}{2}\|\nabla\gamma^k\|^2$, $\forall k \in \mathbb{N}$*.*

Moreover, if these conditions are satisfied, then $\{x^k\}$ *is Fejér-monotone w.r.t.* C*.*

Proof. The equivalence between items 1 and 2 follows from Proposition 3.1. Assuming that item 2 holds, let us prove the Fejér-monotone property. By Lemma 3.2, we have $x^k = x^0 - \nabla\xi^k$, $\forall k \in \mathbb{N}$. Since $\xi^{k+1} = \xi^k + t_k\gamma^k$, it follows that

$$x^{k+1} = x^0 - \nabla\xi^{k+1} = x^0 - \nabla\xi^k - t_k\nabla\gamma^k = x^k - t_k\nabla\gamma^k.$$

Therefore, for every $x \in \mathcal{H}$ we get

$$\|x^{k+1} - x\|^2 = \|x^k - x\|^2 - 2t_k\langle\nabla\gamma^k, x^k - x\rangle + t_k^2\|\nabla\gamma^k\|^2.$$

Using the above identity and the fact that γ^k is an affine function, we obtain

$$\|x^{k+1} - x\|^2 = \|x^k - x\|^2 + 2t_k \left(\gamma^k(x) - \gamma^k(x^k) + \frac{t_k}{2}\|\nabla\gamma^k\|^2 \right). \tag{13}$$

Now, note that identity (13) and the assumption in item 2 yield

$$\|x^{k+1} - x\|^2 = \|x^k - x\|^2 + 2t_k\, \gamma^k(x)\,. \tag{14}$$

To finish the proof note that $\gamma^k(x) \leq 0$ for every $x \in C$.

□

Corollary 3.2. *Let $\{x^k\}$, $\{\xi^k\}$ and $\{\gamma^k\}$ be generated by Algorithm A. Suppose that one of the following conditions holds:*

1. *$\{x^k\}$ is Fejér-monotone w.r.t. C ;*
2. *$\varphi(\xi^{k+1}) \geq \varphi(\xi^k)$, $\forall k \in \mathbb{N}$;*
3. *$\gamma^k(x^k) \geq \frac{t_k}{2}\|\nabla\gamma^k\|^2$, $\forall k \in \mathbb{N}$.*

Then, $x^k \rightharpoonup x \in C$ if and only if $\mathcal{B}(x^k) \subset C$.

Proof. The proof follows directly by combining Proposition 3.3 and Lemma 3.1.

□

4 A Weak-to-Strong Convergence Principle in the Setting of Γ-Duality

In this Section, we use the same notation as in the previous one. In the setting of Γ-duality, a strongly convergent modification of a proximal point type method for finding zeroes of maximal monotone operators, was proposed and studied in [12, Algorithm 2]. In the following, we will generalize this algorithm for solving problem (7).

Algorithm B:

0) Let $x^0 \in \mathcal{H}$ and $\xi^0 = 0$. Set $k = 0$;

1) Find $x^k = \arg\min_{x\in\mathcal{H}} \frac{1}{2}\|x - x^0\|^2 + \xi^k(x)$;

if $x^k \in C$ stops, otherwise,

2) Choose $\gamma^k \in \Gamma(C)$. Let $\Sigma_k := \text{conv(cone)}\{\xi^k, \gamma^k\}$, and define

$$\xi^{k+1} := \arg\max_{\xi\in\Sigma_k} \varphi(\xi);$$

set $k := k + 1$ and go to **1**.

The following proposition is related to [12, Theorem 6.5], where the problem of finding a zero of a maximal monotone operator is considered.

Proposition 4.1. *The Algorithm B is well-defined, and if $x^k \in C$, then x^k is a solution of* (7)*. If $\{x^k\}$ and $\{\xi^k\}$ are generated by Algorithm B, then the following conditions hold.*

a) ξ^k *is dual feasible, i.e.,* $\xi^k \in \Gamma(C), \ \ \forall k \in \mathbb{N}$ *;*

b) $\xi^k(x) = \langle x^0 - x^k, x - x^k \rangle, \ \ \forall k \in \mathbb{N}$ *;*

c) $\xi^k(x^k) = 0, \ \ \forall k \in \mathbb{N}$ *.*

In particular, $\xi^k(x^k) \to 0$, i.e., asymptotic complementarity for $\{\xi^k(x^k)\}$ holds.

Proof. Let us proceed by induction on $k \in \mathbb{N}$. For $k = 0$, items a), b) and c) are trivially verified. Now, assume that items a), b) and c) hold for $k \in \mathbb{N}$ and let $\xi^k, \gamma^k \in \Gamma(C)$ be generated by Algorithm B. Define $g_1 := \xi^k$, $g_2 := \gamma^k$ and $D := \{x \in \mathcal{H} \, : \, g_i(x) \leq 0, \ i = 1, 2\}$. Since $g_1, g_2 \in \Gamma(C)$, it follows that $C \subset D$, which in turn implies $D \neq \emptyset$. Considering $\bar{x}$ and $\bar{\xi}$ given by Corollary 2.1, we obtain from items 3 and 1 (of Corollary 2.1) that $\xi^{k+1} = \bar{\xi}$ and $\bar{x} = x^0 - \nabla \xi^{k+1}$. On the other hand, since $x^0 = \arg\min_{x \in \mathcal{H}} (1/2)\|x - x^0\|^2 + \xi^0(x)$ and the function $x \mapsto (1/2)\|x - x^0\|^2 + \xi^{k+1}(x)$ is strongly convex, x^{k+1} is well defined and has an explicit expression, namely $x^{k+1} = x^0 - \nabla \xi^{k+1}$. Therefore, $x^{k+1} = \bar{x}$, and by item 2 we conclude that items b) and c) hold for ξ^{k+1}. Finally, item a) follows from the last assertion of Proposition 2.2. To see that, if Algorithm B stops at some iteration k, then x^k is a solution, observe that, in such a situation, $x^k \in C$ and by items a) and c), ξ^k is dual feasible and $\xi^k(x^k) = 0$, implying that (x^k, ξ^k) satisfies the general KKT conditions. In particular, by Proposition 2.1, x^k is a primal solution. $\square$

We next show that Algorithm 2.2 falls within the framework of Algorithm B, when applied for solving problem (7) with $C \neq \emptyset$, and that the converse also holds for a special choice of γ^k (see Proposition 4.3).

Proposition 4.2. *Let $\{x^k\}$ be generated by Algorithm 2.2 and assume that $C := \bigcap_{k \geq 0} \mathrm{Fix}(T_k) \neq \emptyset$. Define γ^k, $\xi^k \in \Gamma$ as $\gamma^k(x) := \langle x^k - T_k \, x^k, x - T_k \, x^k \rangle$ and $\xi^k(x) = \langle x^0 - x^k, x - x^k \rangle$. Then γ^k, $\xi^k \in \Gamma(C)$,*

$$x^k = \arg\min_{x \in \mathcal{H}} \frac{1}{2}\|x - x^0\|^2 + \xi^k(x) \ \ \text{and} \ \ \xi^{k+1} = \arg\max_{\xi \in \Sigma_k} \varphi(\xi),$$

where $\Sigma_k := \mathrm{conv}(\mathrm{cone})\{\xi^k, \gamma^k\}$.

As a consequence, Algorithm 2.2 is a special case of Algorithm B, when applied for solving problem (7) *with $C \neq \emptyset$.*

Proof. Let $k \in \mathbb{N}$. From Remark 2.1, we obtain $C \subset H(x^0, x^k) \cap H(x^k, T_k x^k)$, which implies $\gamma^k, \xi^k \in \Gamma(C)$. Using the definition of ξ^k and Lemma 3.2 we obtain $x^k = \arg\min_{x \in \mathcal{H}} (1/2)\|x - x^0\|^2 + \xi^k(x)$. Define $g_1 := \xi^k$, $g_2 := \gamma^k$ and $D := \{x \in \mathcal{H} : g_i(x) \le 0,\ i = 1, 2\}$. Since $g_1, g_2 \in \Gamma(C)$, it follows that $C \subset D$, which in turn implies $D \ne \emptyset$.

Let $\bar{x}$ and $\bar{\xi}$ as in Corollary 2.1. Since $x^{k+1} = Q(x^0, x^k, T_k x^k)$, which means that x^{k+1} is the solution of (7), we obtain $x^{k+1} = \bar{x}$. Now, by items 2 and 3 of Corollary 2.1, $\xi^{k+1} = \bar{\xi}$ and $\xi^{k+1} = \arg\max_{\xi \in \Sigma_k} \varphi(\xi)$, respectively. □

Recall that in this section Γ denotes the set of continuous affine functions on $\mathcal{H}$.

Lemma 4.1. *Let $\gamma \in \Gamma$ be such that $S := \{x \in \mathcal{H} : \gamma(x) \le 0\}$ is nonempty and let $x^0 \in \mathcal{H}$. If $x^0 \notin S$, then*

$$S = H(x^0, P_S(x^0)).$$

Proof. Let $y \in S$. Using (5) we get $\langle x^0 - P_S(x^0), y - P_S(x^0)\rangle \le 0$, which means that $y \in H(x^0, P_S(x^0))$. Thus, $S \subset H(x^0, P_S(x^0))$.

Take now $y \in H(x^0, P_S(x^0))$ and suppose that $\gamma(x) = \langle u, x\rangle - b$, with $u \in \mathcal{H}$ and $b \in \mathbb{R}$. If $u = 0$, then $S = \mathcal{H}$ and so the desired result follows trivially. Assume that $u \ne 0$. Note that in this case

$$P_S(x^0) = x^0 - t\,u\,, \qquad \langle u, P_S(x^0)\rangle = b, \tag{15}$$

where $t = (1/\|u\|^2)(\langle u, x^0\rangle - b)$. Using the first identity in (15) and the fact that $y \in H(x^0, P_S(x^0))$ we obtain

$$\langle tu, y - P_S(x^0)\rangle = \langle x^0 - P_S(x^0), y - P_S(x^0)\rangle \le 0. \tag{16}$$

Since $x^0 \notin S$, it follows that $t > 0$. Direct use of (16) gives $\langle u, y - P_S(x^0)\rangle \le 0$, which combined with the second identity in (15), yields $\langle y, u\rangle \le b$, i.e., $y \in S$. Thus, $H(x^0, P_S(x^0)) \subset S$, which conclude the proof of the lemma.

□

Proposition 4.3. *Let $\{x^k\}, \{\gamma^k\}$ be generated by Algorithm B and assume that $\gamma^k(x^k) > 0$, $\forall k \in \mathbb{N}$. For any $k \in \mathbb{N}$, let $C_k := \{x \in \mathcal{H} : \gamma^k(x) \le 0\}$ and define $T_k = P_{C_k}$. Then*

$$x^{k+1} = Q(x^0, x^k, T_k\, x^k) \ \ (\forall k \in \mathbb{N}).$$

Therefore, if $\gamma^k(x^k) > 0$ $(\forall k \in \mathbb{N})$, then Algorithm B is a special case of Algorithm 2.2.

Proof. Let $k \in \mathbb{N}$. Since $\gamma^k \in \Gamma(C)$, we have $C \subset C_k$, which implies that $C_k \ne \emptyset$ and, in particular, that $\{T_k\}$ is well-defined. From the projection characterization given in (5), we obtain $P_{C_k} \in \mathcal{T}$. Thus, $\{T_k\} \subset \mathcal{T}$. Since $x^k \notin C_k$, it follows from Lemma 4.1 that $C_k = H(x^k, T_k x^k)$. Using the same reasoning as in Proposition 4.2, we may conclude that $x^{k+1} = \bar{x}$ (see Corollary 2.1). Hence,

$$x^{k+1} = \bar{x} = \arg\min_{x \in D} \frac{1}{2}\|x - x^0\|^2, \tag{17}$$

where $D = \{x \in \mathcal{H} \, : \, \xi^k(x) \leq 0 \;\text{ and }\; \gamma^k(x) \leq 0\}$. Therefore, observing that $D = H(x^0, x^k) \cap H(x^k, T_k x^k)$, it follows from (17) that $x^{k+1} = Q(x^0, x^k, T_k\, x^k)$. $\square$

Remark 4.1. In Section 5, we show that a subgradient type method can be seen as a special case of Algorithm A. In such a situation, γ^k satisfies the property $\gamma^k(x^k) > 0, \quad \forall\; k \in \mathbb{N}$. Therefore, by Proposition 4.3, in such a situation, Algorithm B is a special case of Algorithm 2.2. See also Remark 5.1.

We have seen that attempting to solve (7), Algorithm A generates a sequence with weak convergence properties (see Corollary 3.2), while Algorithm B is motivated by forcing complementarity iteratively, i.e., $\xi^k(x^k) = 0, \;\; \forall k \in \mathbb{N}$. As a consequence, Algorithm B has strong convergence properties, as we show in Corollary 4.1. In the following, we propose a general algorithm, which contains as special cases Algorithms A and B. We will discuss under what hypothesis this algorithm generates a strongly convergent sequence for the solution of (7).

Γ-Algorithm:

0) Let $x^0 \in \mathcal{H}$ and $\xi^0 = 0$. Set $k = 0$;

1) Find $x^k = \arg\min_{x \in \mathcal{H}} \frac{1}{2}\|x - x^0\|^2 + \xi^k(x)$;

2) Choose $\xi^{k+1} \in \Gamma(C)$, set $k := k + 1$ and go to **1**.

Note that Algorithm A is a special case of Γ-Algorithm. Indeed, this follows from the fact that $\xi^k, \gamma^k \in \Gamma(C)$ implies $\xi^{k+1} = \xi^k + t_k\, \gamma^k \in \Gamma(C)$ (see Remark 3.2).

We now state and prove one of our main results. It was pointed out by Svaiter (without a proof) in [12, Section 6.2], for the special case of the problem of finding zeroes of maximal monotone operators.

Theorem 4.1. (Dual strong convergence principle) *Consider the sequences* $\{x^k\}$ *and* $\{\xi^k\}$ *generated by* Γ*-Algorithm. Suppose that* asymptotic complementarity *holds, i.e.,*

$$\xi^k(x^k) \to 0, \;\; k \to \infty.$$

Then,

i) $\{x^k\}$ *is bounded;*

ii) $x^k \to P_C(x^0)$ *if and only if* $\mathcal{B}(x^k) \subset C$ *.*

Proof. It follows from the definition of x^k, from the definition of the dual function φ, from the dual feasibility of ξ^k and from the weak duality property that $(\forall k \in \mathbb{N})$

$$\|x^k - x^0\|^2 + 2\,\xi^k(x^k) = 2\,\varphi(\xi^k) \le \|x - x^0\|^2, \;\; \forall\, x \in C.$$

By assumption we have $\xi^k(x^k) \to 0, \;\; k \to \infty$. Thus, $\{x^k\}$ is bounded and

$$\limsup_{k\to\infty} \|x^k - x^0\|^2 + 2\,\xi^k(x^k) \le \|P_C(x^0) - x^0\|^2. \tag{18}$$

The "if" implication in *ii*) is trivial. For the "only if" implication, assume that $\mathcal{B}(x^k) \subset C$.

Next, since $\{x^k\}$ is bounded, we obtain $\mathcal{B}(x^k) \neq \emptyset$. Let $\bar{x} \in \mathcal{B}(x^k)$ and take $\{x^{k_j}\}$ such that $x^{k_j} \rightharpoonup \bar{x}, \;\; j \to \infty$.

Since $\xi^{k_j}(x^{k_j}) \to 0, \;\; j \to \infty$, we have, in particular, that

$$\limsup_{j\to\infty} \|x^{k_j} - x^0\|^2 \le \limsup_{j\to\infty} \|x^{k_j} - x^0\|^2 + 2\,\xi^{k_j}(x^{k_j}). \tag{19}$$

Using the weak lower semi-continuity of $\|\cdot\|$, (19) and (18) we obtain

$$\|\bar{x} - x^0\|^2 \le \liminf_{j\to\infty} \|x^{k_j} - x^0\|^2 \le \limsup_{j\to\infty} \|x^{k_j} - x^0\|^2 \le \|P_C(x^0) - x^0\|^2\,. \tag{20}$$

The inclusion $\mathcal{B}(x^k) \subset C$ implies $\bar{x} \in C$. Thus, it follows from (20) that

$$\bar{x} = P_C(x^0) \;\text{ and }\; \|x^{k_j} - x^0\| \to \|P_C(x^0) - x^0\|, \;\; j \to \infty\,.$$

Therefore, since $\{x^{k_j}\}$ is arbitrary, it follows that $\{x^k\}$ converges weakly to $P_C(x^0)$ and $\|x^k - x^0\| \to \|P_C(x^0) - x^0\|, \;\; k \to \infty$. However,

$$\|x^k - P_C(x^0)\|^2 = \|x^k - x^0\|^2 - \|P_C(x^0) - x^0\|^2 + 2\langle x^k - P_C(x^0), x^0 - P_C(x^0)\rangle.$$

Thus, $x^k \to P_C(x^0)$. □

Corollary 4.1. *Let* $\{x^k\}$ *and* $\{\xi^k\}$ *be generated by Algorithm B. Then,*

$$x^k \to P_C(x^0) \;\text{ if and only if }\; \mathcal{B}(x^k) \subset C.$$

Although Algorithm 2.2 is very general, including many well known algorithms in its setting [9], it is a special case of Γ-Algorithm. This is due to the fact that Γ-Algorithm requires dual feasibility of ξ^k and asymptotic complementarity, while Algorithm 2.2 uses projection onto the intersection of two special half spaces, implying, in particular, complementarity iteratively, i.e., $\xi^k(x^k) = 0$, $\forall k \in \mathbb{N}$.

The following example is for illustrative purposes only.

Example 4.1. Let $0 \neq x^0 \in \mathcal{H}$ and $C = \{x \in H : \langle x^0, x\rangle \leq 0\}$. Consider the following problem

$$\begin{cases} \min \frac{1}{2}\|x - x^0\|^2 \\ \text{s.t. } \; x \in C. \end{cases} \tag{21}$$

Let $\{\alpha_k\} \subset \;]0, 1[\;$ be a sequence converging to 0, such that $\alpha_k < \alpha_{k+1}$ for infinite indexes k (for example, define $\alpha_k = 1/(k + 1)$ if k is odd, otherwise $\alpha_k = 2/(k+1)$). Consider Γ-Algorithm with the following updating rule for ξ^k:

$$\xi^k(x) = \langle (1 - \alpha_k)x^0, x\rangle.$$

It follows from Lemma 3.2 and this updating rule that

$$x^k = x^0 - \nabla\xi^k = x^0 - (1 - \alpha_k)x^0 = \alpha_k x^0, \tag{22}$$

which implies that $\{x^k\}$ converges strongly to $0 = P_C(x^0)$, solution of the primal problem (21). If $x \in C$, then

$$\xi^k(x) = \langle (1 - \alpha_k)x^0, x\rangle = (1 - \alpha_k)\langle x^0, x\rangle \leq 0,$$

which means that ξ^k is dual feasible. Using (22) we obtain

$$\xi^k(x^k) = (1 - \alpha_k)\alpha_k\langle x^0, x^0\rangle > 0, \quad \forall k \in \mathbb{N}.$$

Therefore the complementarity is achieved only asymptotically, $\lim \xi^k(x^k) = 0$. We claim that it is not possible to find a family of operators $T_k \in \mathcal{T}$ (see Definition 2.2), such that

$$x^{k+1} = Q(x^0, x^k, T_k x^k),$$

that is, $\{x^k\}$ can not be generated within the setting of [9]. Indeed, if this is the case, then $x^{k+1} \in H(x^0, x^k)$ for all $k \in \mathbb{N}$. But this means that

$$\langle x^{k+1} - x^k, x^0 - x^k\rangle \leq 0 \quad \forall k \in \mathbb{N},$$

which, by (22), implies that

$$(\alpha_{k+1} - \alpha_k)(1 - \alpha_k)\|x^0\|^2 \leq 0, \qquad \forall \; k \in \mathbb{N}.$$

Since $x^0 \neq 0$ and $\{\alpha_k\} \subset \;]0, 1[$, the last inequality implies $\alpha_{k+1} \leq \alpha_k$, $\forall k \in \mathbb{N}$, contradicting the way that $\{\alpha_k\}$ was chosen.

5 Subgradient Type Methods in the Setting of Γ-Duality

In this Section, we show how subgradient type methods fit in the setting of Γ-duality. In particular, we discuss how an algorithm proposed in [13] for solving a constrained optimization problem can be seen as a special case of Γ-Algorithm.

5.1 Subgradient Method

Let $f : \mathcal{H} \to \mathbb{R}$ be a convex function, and $C := \{x \in \mathcal{H} : f(x) \leq 0\}$. Consider the problem of finding a point $x \in C$. A well known approach for solving this problem is to apply a subgradient algorithm stated as follows,

Subgradient Method

0) Let $\{\alpha_k\} \subset [\epsilon, 2-\epsilon]$, where $\epsilon > 0$ and let y^0. Set $k = 0$;

1) If $f(y^k) \leq 0$ stops, otherwise obtain a subgradient $v^k \in \partial f(y^k)$;

2) Let $\mu_k := \frac{\alpha_k f(y^k)}{\|v^k\|^2}$ and define $y^{k+1} = y^k - \mu_k v^k$;
Set $k = k+1$ and go to **1**.

Next proposition shows how the subgradient method described above fit in the setting of Γ-duality.

Proposition 5.1. *The subgradient method is a special case of Algorithm A.*

Proof. Assume that $\{y^j\}_{j=0}^{j=k}$ is generated by the subgradient method. Let $\{x^j\}_{j=0}^{j=k}$ and $\{\xi^j\}_{j=0}^{j=k}$ generated by Algorithm A, with $\xi^0 = 0$, $x^0 = y^0$, $\lambda_j := \mu_j$ for $j = 0, \cdots, k-1$, and the affine functions γ^j defined as

$$\gamma^j(x) = \langle v^j, x - y^j \rangle + f(y^j), \qquad \forall\, j = 0, \cdots, k-1.$$

From the subgradient inequality $\langle v^j, x - y^j \rangle + f(y^j) \leq f(x)$, it follows that, if $x \in C$ (i.e., $f(x) \leq 0$), then $\gamma^j(x) \leq 0$, implying $\gamma^j \in \Gamma(C)$ for $j = 0, \cdots, k-1$.

Let us show by induction that $y^j = x^j$ for $j = 0, \cdots, k$. By assumption, $y^0 = x^0$. Now, using recursively the updating rules $y^{j+1} = y^j - \mu_j v^j$ and $\xi^{j+1} = \xi^j + \lambda_j \gamma^j$, we obtain

$$y^{j+1} = y^0 - \sum_{i=0}^{j} \mu_i v^i, \quad \text{and} \quad \xi^{j+1} = \xi^0 + \sum_{i=0}^{j} \lambda_i \gamma^i.$$

Since $\nabla \gamma^i = v^i$, it follows that

$$\nabla \xi^{j+1} = \nabla \xi^0 + \sum_{i=0}^{j} \lambda_i v^i.$$

Therefore, using that $\xi^0 = 0$, $y^0 = x^0$, and $\lambda_i = \mu_i$ for $i = 0, \cdots, k-1$, we obtain

$$y^{j+1} = y^0 - \sum_{i=0}^{j} \mu_i v^i = x^0 - \nabla \xi^{j+1} = x^{j+1},$$

where the last equality follows by Lemma 3.2. Thus $x^i = y^i$, for $i = 0, \cdots, k$, concluding the proof. □

The Fejér-monotonicity of the subgradient method w.r.t. C can be verified in the Γ-duality setting as follows. Let γ^j as in the proof of Proposition 5.1. If $f(x^j) > 0$ then, using that $\nabla\gamma^j = v^j$, $\gamma^j = f(x^j)$ and $\lambda_j = \frac{\alpha_j f(x^j)}{\|v^j\|^2}$, it follows that

$$\lambda_j \|\nabla\gamma^j\|^2 = \alpha_j \gamma^j(x^j),$$

with $\alpha_j \in \,]0, 2[$. Hence, condition (2) of Proposition 3.3 is satisfied, which implies the Fejér-monotonicity of the subgradient method w.r.t. C.

Consider now the following optimization problem

$$\begin{cases} \min \ f(x) \\ \text{s.t.} \ \ x \in \mathcal{H}\,. \end{cases} \tag{23}$$

Assuming that the optimal solution set $S^* \neq \emptyset$ and the optimal value f^* is known a priori, the subgradient method discussed above can be easily generalized to solve this problem by considering

$$\mu_k := \frac{\alpha_k(f(y^k) - f^*)}{\|v^k\|^2}, \quad \text{and} \quad C := S^*.$$

Therefore, using $\gamma^k(x) = \langle v^k, x - y^k\rangle + f(y^k) - f^*$, we may, similarly to the proof of Proposition 5.1, fit the subgradient method (with the steplenth above) in the setting of Algorithm A to solve (23).

A variant of the subgradient method, which has strong convergence properties, was proposed in [13] to solve the following constrained optimization problem

$$\begin{cases} \min \ f(x) \\ \text{s.t.} \ \ x \in C, \end{cases} \tag{24}$$

where $C \subset \mathcal{H}$ is a closed and convex set with a simple structure. Precisely, it is assumed to be easy to project any point $x \in \mathcal{H}$ onto C. Next, we recall this algorithm and show that it is a special case of Γ-Algorithm when C is polyhedron.

Algorithm 5.1.

0) *Let* $x^0 \in C$*, set* $k := 0$*;*

1) *If* $f(x^k) = f^*$ *stops, otherwise let* $v^k \in \partial f(x^k)$*;*

2) *Set*

$$H_k := \{x \in \mathcal{H} : f(x^k) - f^* + \langle v^k, x - x^k\rangle \leq 0\}$$

and

$$W_k := \{x \in \mathcal{H} : \langle x - x^k, x^0 - x^k\rangle \leq 0\};$$

3) *Let* $x^{k+1} := P_{Z_k}(x^0)$, *where* $Z_k := C \cap H_k \cap W_k$; *set* $k := k+1$ *and go to* **1**.

Observe that (24) is equivalent to the convex feasibility problem:

$$\text{find } \ x \in S := C \cap \{x \in \mathcal{H} : f(x) \leq f^*\}.$$

We may consider the more stringent problem: given $x^0 \in \mathcal{H}$,

$$\begin{cases} \min \ \frac{1}{2}\|x - x^0\|^2 \\ \text{s.t. } \ x \in S. \end{cases} \tag{25}$$

Assume that C is a polyhedron set, that is, there exist affine functions g_i for $i = 1, \cdots, n$, such that

$$C := \{x \in \mathcal{H} : g_i(x) \leq 0, \ i = 1, \cdots, n\}.$$

The proof of the following proposition is similar to the one of Proposition 4.2, we show it for the sake of completeness.

Proposition 5.2. *Let* $\{x^k\}$ *be generated by Algorithm 5.1 for solving* (24) *with* C *a polyhedron, as above. Then* $\{x^k\}$ *is also generated by* Γ*-Algorithm with* $\xi^0 = 0$, *and updating* ξ^k *as*

$$\xi^{k+1} = \arg\max_{\xi \in \Sigma_k} \varphi(\xi),$$

where $\Sigma_k := \text{co(cone)}\{\xi^k, \gamma^k, g_1, \cdots, g_n\}$, *and* $\gamma^k(x) = \langle v^k, x - x^k\rangle + f(x^k) - f^*$.

Proof. Let $\{\xi^k\}$ be generated by Γ-Algorithm, and $\{x^k\}$ be generated by Algorithm 5.1. Let us show that $\xi^k(x) = \langle x - x^k, x^0 - x^k\rangle$ for any k. By assumption, this result is true for $k = 0$. Assume that it holds for some $k \in \mathbb{N}$, and consider the following set $D := \{x \in \mathcal{H} : \xi^k(x) \leq 0, \gamma^k(x) \leq 0, g_i(x) \leq 0, \ i = 1, \cdots, n\}$. Since by induction hypothesis $\xi^k(x) = \langle x - x^k, x^0 - x^k\rangle$, it follows that $D = Z_k$ and thus, $D \neq \emptyset$, because $x^{k+1} = P_{Z_k}(x^0) \in Z_k$. Let $\bar{x}$ and $\bar{\xi}$ as in Corollary 2.1. By definition, $\bar{x} = P_D(x^0)$, and thus $x^{k+1} = \bar{x}$. From items 2 and 3 of Corollary 2.1, $\bar{\xi} = \arg\max_{\xi \in \Sigma_k} \varphi(\xi)$ and $\bar{\xi}(x) = \langle x - \bar{x}, x^0 - \bar{x}\rangle$, implying from the updating rule of ξ^{k+1} that $\xi^{k+1} = \bar{\xi}$. Combining these results, we have $\xi^{k+1}(x) = \langle x - x^{k+1}, x^0 - x^{k+1}\rangle$. Therefore $\xi^k(x) = \langle x - x^k, x^0 - x^k\rangle$ for any k. As a consequence, $\nabla\xi^k = x^0 - x^k$ for any $k \in \mathbb{N}$, which implies, by Lemma 3.2, that $x^k = \arg\min_{x \in \mathcal{H}}(1/2)\|x - x^0\|^2 + \xi^k(x)$ for any $k \in \mathbb{N}$. Hence, $\{x^k\}$ is also generated by Γ-Algorithm, as desired. □

Remark 5.1. We mention that Algorithm 5.1 applied to solving problem (23), ($C = \mathcal{H}$), is a particular case of Algorithm B, taking γ^k as defined above. In particular $\gamma^k(x^k) > 0$, and so by Proposition 4.3, in such a situation, Algorithm B can be seen as a realization of Algorithm 2.2.

6 Concluding Remarks

In this article, based on a duality theory introduced and studied by B .F. Svaiter in [11], and further developed and used in [12], we propose a general algorithm for finding a point in a closed and convex subset of a Hilbert space. We analyze two variants of the algorithm which have different properties. The first one has weak convergence properties and the second one, under some assumptions, generates a sequence that converges strongly to a solution. One of our main focus is to propose a weak-to-strong convergence principle based on the Svaiter's Γ-duality. We show that updating the dual sequence keeping dual feasibility, and forcing asymptotic complementarity [12] of the primal sequence guarantee strong convergence, assuming that all weak accumulation points of the sequence belong to the convex set. We relate the obtained results with the weak-to-strong convergence principle of Bauschke and Combettes and discuss how subgradient type methods can be fit in the setting presented here.

Acknowledgments

This work was completed while the first author was visiting the Institute of Mathematics and Statistics at the University of Goias, whose warm hospitality is gratefully acknowledged. The authors thank Dr. Benar Fux Svaiter for bringing this problem to our attention, for sending us a copy of [11] and [12] as well as for the criticism concerning the first version of this manuscript. The authors also thank the two anonymous referees and the editor for their valuable comments.

M. Marques Alves is partially supported by CNPq grants $305414/2011-9$ and $479729/2011-5$ and PRONEX-Optimization. J. G. Melo is Partially supported by Procad/NF and by PRONEX-Optimization.

References

1. Ekeland, I., Temam, R.: Convex Analysis and Variational Problems. North Holland, Amsterdan (1976)

2. Bauschke, H. H., Combettes, P. L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)

3. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer, New York (2011)

4. Vinter, R.: Optimal Control. Modern Birkhäuser Classics, Secaucus (2000)

5. Engl, H. W., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer Academic Publishers, Dordrecht (1996)

6. Rockafellar, R. T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. 14, 877–898 (1976)

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

8. Solodov, M. V., Svaiter, B. F.: Forcing strong convergence of proximal point iterations in a Hilbert space. Math. Program. 87, 189–202 (2000)

9. Bauschke, H. H.; Combettes, P. L.: A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. Math. Oper. Res. 26, 248–264 (2001)

10. Dantzig, G. B., Linear Programming and Extensions. Princeton University Press, New Jersey (1963)

11. Svaiter, B. F.: Uma nova abordagem para dualidade em programação não linear e aplicações. Ph.D. Thesis, IMPA (1994)

12. Svaiter, B. F.: A new duality theory for mathematical programming. Optimization 60, 1209–1231 (2011)

13. Bello Cruz, J. Y., Iusem, A. N.: A strongly convergent method for nonsmooth convex minimization in Hilbert spaces. Numer. Funct. Anal. Optim. 32, 1009–1018 (2011)

14. Fan, Ky.: On systems of linear inequalities. Linear inequalities and related systems. Annals of Math. Studies 38, Princeton University Press, New Jersey 99–156 (1956)