A strongly convergent method for nonsmooth convex minimization in Hilbert spaces

J. Y. Bello Cruz* A. N. Iusem†

November 22, 2010

Abstract

In this paper we propose a strongly convergent variant on the projected subgradient method for constrained convex minimization problems in Hilbert spaces. The advantage of the proposed method is that it converges strongly when the problem has solutions, without additional assumptions. The method also has the following desirable property: the sequence converges to the solution of the problem which lies closest to the initial iterate.

Keywords: Convex minimization, Projection method, Nonsmooth optimization, Strong convergence, Projected subgradient algorithm.
Mathematical Subject Classification (2008): 90C25, 65C25, 90C30.

1 Introduction

First we briefly describe our notation. Let C be a nonempty, closed and convex subset of a real Hilbert space $\mathcal{H}$ and $f : \mathcal{H} \to \mathbb{R}$ be a continuous, proper and convex function. The inner product in $\mathcal{H}$ is denoted by $\langle \cdot, \cdot \rangle$. The norm determined by the inner product is denoted by $\| \cdot \|$. For an element $x \in \mathcal{H}$, the orthogonal projection of x onto C is denoted by $P_C(x)$.
The problem under consideration is the following:

$$\min\ f(x) \qquad \text{s.t.} \quad x \in C. \tag{1}$$

We denote $f^* := \inf_{x\in C} f(x)$ and the solution set of this problem by $S^* := \{x \in C \ : \ f(x) = f^*\}$.

We remind that the subdifferential of f at x is the subset $\partial f(x)$ of $\mathcal{H}$ given by

$$\partial f(x) = \{u \in \mathcal{H} \ : \ f(y) \geq f(x) + \langle u, y - x\rangle \quad \forall\, y \in \mathcal{H}\}\,. \tag{2}$$

The elements of the subdifferential of f at x are called subgradients of f at x.

*Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus Samambaia, CEP 74001-970 GO, Goiânia, Brazil, e-mail: yunier@impa.br

†Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, CEP 22460-320 RJ, Rio de Janeiro, Brazil, e-mail: iusp@impa.br

1.1 The projected subgradient method

Many methods have been proposed to solve Problem (1), with the simplest being the projected subgradient method [17, 18, 2, 1].

The projected subgradient method has several advantages. Primarily, it is easy to implement (specifically, for optimization problems with relatively simple constraints), it uses little storage and readily exploits any sparsity or separable structure of ∂f or C. Furthermore, it is able to drop or add active constraints during the iterations. Thus, the projected subgradient method has been widely used for solving various cases of Problem (1).

Each iteration of the projected subgradient method, consists basically of two stages: starting from the iterate x^k, we move a certain length in the direction opposite to a subgradient of f at x^k, and next the resulting vector is projected onto C. This is not a trivial idea, since the function will not, in general, decrease along a direction selected in this way.

When $C = \mathcal{H}$ and f is differentiable this is just the steepest descent method [8]. For differentiable functions the projected subgradient method coincides with the projected gradient method and different variants rules are used to choose the stepsizes in order to ensure functional decrease at each iteration; see [13]. Its extension to the generalized convex case is studied in [21, 7, 14].

1.2 On strong convergence

1.2.1 The new method

In this paper, a modification of the projected subgradient method forcing strong convergence in Hilbert spaces is considered. The new method is related to the classical projected subgradient method and it uses a similar idea exposed in [6] and [19], with the same goal, upgrading weak convergence to strong one. Strong convergence is forced by combining a subgradient iteration with a simple projection step onto the intersection of C and two halfspaces, containing S^*.

Additionally, our algorithm has the distinctive feature that the limit of the generated sequence is the closest solution of the problem to the initial iterate x^0. This property is useful in many specific applications, e.g. in image reconstruction [12, 9, 16]. We emphasize that this feature is of interest also in finite dimension, differently from the strong versus weak convergence issue.

We mention that the proposed method requires in each iteration the computation of the exact projection onto the intersection of C with two halfspaces. The presence of the halfspaces does not entail any significant additional cost over the computation of the projection onto C itself. The computational cost of this projection is similar to the cost of a projected subgradient step.

In order to prove the convergence of the new method, we assume that f is finite valued, so that the effective domain is $\mathcal{H}$. Also we suppose that ∂f is bounded on bounded sets, which holds automatically in finite dimension, as a consequence of the maximal monotonicity of the subdifferential of lower semicontinuous convex functions. We also mention that these assumptions are required in the analysis of [2] for proving weak convergence of the projected subgradient method in infinite-dimensional Hilbert spaces.

1.2.2 Why strong convergence?

It is important to mention that many real-world problems in economics and engineering are modeled in infinite-dimensional spaces. These include optimal control and structural design problems, among others. It is clear that weak and strong convergence are only distinguishable in the infinite-dimensional setting. On the other hand, even when we have to solve infinite-dimensional problems, numerical implementations of algorithms are certainly applied to finite-dimensional approximations of the problems. Nevertheless, it is important to have convergence theory for the infinite-dimensional case, because it guarantees robustness and stability with respect to discretization schemes employed for obtaining finite-dimensional approximations of infinite-dimensional problems. This issue is closely related to the so-called Mesh Independence Principle [4, 3, 15]. This principle relies on infinite-dimensional convergence to predict the convergence properties of a discretized finite-dimensional method. Furthermore, the mesh independence provides theoretical justification for the design of refinement strategies.

Note that fine discretization is crucial for the obtained discrete solution to be an appropriate approximation to the true solution of the infinite-dimensional problem being solved. Many important real-world problems in economics and engineering are modeled in infinite-dimensional spaces. These include optimal control and structural design problems, and the problem of minimal area surface with obstacles, among others. We refer the reader to [11], where a variety of applications are described. A strong convergence principle for Fejér-monotone methods in Hilbert spaces was presented in [5].

The importance of strong convergence is also underlined in [10], where a convex function f is minimized via the proximal-point algorithm: It is shown that the rate of convergence of the value sequence $\{f(x^k)\}$ is better when $\{x^k\}$ converges strongly than when it converges weakly. Such properties have a direct impact when the algorithm is executed directly in the underlying infinite-dimensional space, as is the case, for instance, in optical signal processing, see [20].

1.3 Preliminary material

In this subsection we present some results that are used in the convergence analysis. First, we state a well known fact on orthogonal projections.

Proposition 1. *Let K be any nonempty closed and convex set in $\mathcal{H}$. For all $x, y \in \mathcal{H}$ and all $z \in K$, $\langle x - P_K(x), z - P_K(x)\rangle \leq 0$.*

Proof. Elementary. □

We continue with an elementary property of the solution set of Problem (1).

Proposition 2. *Assume that $f : \mathcal{H} \to \mathbb{R}$ is proper, finite-valued, convex and continuous. Then S^*, if nonempty, is a closed and convex set.*

Proof. Closedness of $S^* = \{x \in C \ : f(x) \leq f^*\}$ follows easily from the continuity of f. Convexity of S^* is elementary. □

2 The Algorithm

In this section we state the method, which generates a sequence strongly convergent to a point belonging to S^*, differently from the projected subgradient method, for which only weak convergence has been established. We assume that the optimal value f^* of Problem (1) is available.

Algorithm A

Initialization step. Take $x^0 \in C$, $\beta_0 := f(x^0) - f^*$ and $u^0 \in \partial f(x^0)$.

Iterative step. Given x^k and $u^k \in \partial f(x^k)$, define

$$\beta_k := f(x^k) - f^*, \tag{3}$$

$$H_k := \left\{x \in \mathcal{H} \, : \, \langle x - x^k, u^k \rangle + \beta_k \le 0\right\},$$

and

$$W_k := \left\{x \in \mathcal{H} \, : \, \langle x - x^k, x^0 - x^k \rangle \le 0\right\}.$$

Compute

$$x^{k+1} := P_{H_k \cap W_k \cap C}(x^0). \tag{4}$$

If $x^{k+1} = x^k$ then stop.

Before the formal analysis of the convergence properties of the algorithm, we make the following remarks:

1. Regarding the projection step, we shall prove that the set $H_k \cap W_k \cap C$ is nonempty, ensuring that the algorithm is well defined.
2. Concerning the complexity of the projection step in Algorithm A, the presence of the halfspaces does not entail any significant additional cost over the computation of the projection onto C itself. The computational cost of this projection is similar to the cost of a projected subgradient step. Even though we are working in an infinite-dimensional space, projection onto an intersection of two halfspaces amounts to solving, at most, a linear system of two equations with two unknowns. Thus, the cost of each iteration of Algorithm A is about the same as that of an iteration of the projected subgradient method.

Next we proceed to establish strong convergence of the sequence generated by Algorithm A.

3 Convergence analysis

First we establish some properties of the iterates.

Proposition 3. *For all $k \ge 0$ it holds that*

$$\|x^{k+1} - x^0\|^2 \ge \|x^k - x^0\|^2 + \|x^{k+1} - x^k\|^2, \tag{5}$$

and

$$\|x^{k+1} - x^k\| \ge \frac{f(x^k) - f^*}{\|u^k\|} \,. \tag{6}$$

Proof. Since $x^{k+1} \in W_k$,

$$0 \ge \langle x^{k+1} - x^k, x^0 - x^k \rangle = \frac{1}{2}\left(\|x^{k+1} - x^k\|^2 - \|x^{k+1} - x^0\|^2 + \|x^k - x^0\|^2\right),$$

which implies (5).

Now, since $\|x^k - P_{H_k}(x^k)\| \le \|z - x^k\|$ for all $z \in H_k$ and $x^{k+1} \in H_k$, we have that

$$\|x^{k+1} - x^k\| \ge \|x^k - P_{H_k}(x^k)\|. \tag{7}$$

Using (3) and the fact that $P_{H_k}(x^k) = x^k - \beta_k \dfrac{u^k}{\|u^k\|^2}$, we obtain from (7)

$$\|x^{k+1} - x^k\| \ge \|x^k - P_{H_k}(x^k)\| = \frac{\beta_k}{\|u^k\|} = \frac{f(x^k) - f^*}{\|u^k\|}.$$

□

Now we show feasibility of the sequence generated by Algorithm A and the validity of the stopping criterion.

Proposition 4. *If the Algorithm A generates an infinite sequence* $\{x^k\}$*, then* $\{x^k\} \subseteq C$*. Furthermore, if Algorithm A stops,* $x^k \in S^*$*.*

Proof. First suppose that Algorithm A generates an infinite sequence. By (4), we have $\{x^k\} \subseteq C$. If Algorithm A stops, then $x^{k+1} = x^k$. It follows from (6) that $f(x^k) - f^* \le 0$, so $x^k \in S^*$.

□

Next we prove optimality of the weak cluster points of $\{x^k\}$.

Theorem 1. *Suppose that Algorithm A generates an infinite sequence* $\{x^k\}$*. Then either* $\{x^k\}$ *is bounded and each of its weak cluster points belongs to* $S^* \neq \emptyset$*, or* $S^* = \emptyset$ *and* $\lim_{k\to\infty} \|x^k\| = \infty$*.*

Proof. If $\{x^k\}$ is bounded, we obtain from (5) that the sequence $\{\|x^k - x^0\|\}$ is nondecreasing and bounded, hence convergent. By (5) again, $0 \le \|x^{k+1} - x^k\|^2 \le \|x^{k+1} - x^0\|^2 - \|x^k - x^0\|^2$, and we conclude that

$$\lim_{k\to\infty} \|x^{k+1} - x^k\|^2 = 0. \tag{8}$$

By Proposition 4, $f(x^k) - f^* \ge 0$ for all k. So,

$$0 = \lim_{k\to\infty} \|x^{k+1} - x^k\| \ge \lim_{k\to\infty} \frac{f(x^k) - f^*}{\|u^k\|} \ge 0,$$

using (6) and (8). The boundedness of $\{u^k\}$ follows from the boundedness of $\{x^k\}$ and the assumption that ∂f is bounded on bounded sets. Thus,

$$\lim_{k\to\infty} f(x^k) - f^* = 0, \tag{9}$$

Since $\{x^k\}$ is bounded, there exists a weakly convergent subsequence. Let $\{x^{i_k}\}$ be any weakly convergent subsequence of $\{x^k\}$, and let $\bar{x} \in C$ be its weak limit. Since f is weakly lower semicontinuous we get, using (9),

$$f^* \le f(\bar{x}) \le \liminf_{k\to\infty} f(x^{i_k}) = \lim_{k\to\infty} f(x^k) = f^*.$$

Thus $f(\bar{x}) = f^*$, implying that $\bar{x} \in S^*$. We conclude that all weak cluster points of $\{x^k\}$ belongs to S^*.

Suppose now that $S^* = \emptyset$. Using the previous assertion in this theorem, we obtain that $\{x^k\}$ has no bounded subsequence, and consequently $\lim_{k\to\infty} \|x^k\| = \infty$. □

Next, we shall prove that when $S^* \neq \emptyset$ Algorithm A is well defined and the generated sequence $\{x^k\}$ converges strongly to a solution.

We assume from now on that S^* is nonempty. Having chosen the initial iterate x^0, define

$$V_0 := \{x \in \mathcal{H} : \langle z - x, x^0 - x \rangle \le 0 \quad \forall z \in S^*\}. \tag{10}$$

Next we show that the generated sequence $\{x^k\}$ is contained in V_0 and that the set $H_k \cap W_k \cap C$ always contains the solution set S^*.

Proposition 5. *If* $x^k \in V_0$ *then*

i) $S^* \subseteq H_k \cap W_k \cap C$*,* *ii)* x^{k+1} *is well defined and* $x^{k+1} \in V_0$.

Proof. (i) Since $\beta_k = f(x^k) - f^*$ by (3), we get, using (2),

$$\beta_k + \langle u^k, x^* - x^k \rangle = f(x^k) - f(x^*) + \langle u^k, x^* - x^k \rangle \leq 0, \tag{11}$$

for any $x^* \in S^*$. It follows from (11) that $S^* \subseteq H_k$.

Since $x^k \in V_0$, we have that $\langle x^* - x^k, x^0 - x^k \rangle \leq 0$ for all $x^* \in S^*$. By the definition of W_k, we obtain that $S^* \subseteq W_k$. Therefore, we conclude that $S^* \subseteq H_k \cap W_k \cap C$.

(ii) Since $S^* \subseteq H_k \cap W_k \cap C$ and S^* is nonempty, it follows that $H_k \cap W_k \cap C$ is nonempty. Thus, the next iterate x^{k+1} is well defined, in view of (4).

Using (4) and Proposition 1, we have that

$$\langle z - x^{k+1}, x^0 - x^{k+1} \rangle \leq 0 \quad \forall z \in H_k \cap W_k \cap C. \tag{12}$$

By (i), $S^* \subseteq H_k \cap W_k \cap C$ for all k. Since (12) holds for all $z \in S^*$, so $x^{k+1} \in V_0$ by (10).

□

Corollary 1. *Algorithm A is well defined,* $\{x^k\} \subseteq V_0$ *and* $S^* \subseteq H_k \cap W_k \cap C$ *for all* k.

Proof. It is enough to observe that $x^0 \in V_0$ and apply inductively Proposition 5.

□

Corollary 2. *The sequence* $\{x^k\}$ *generated by Algorithm A is bounded and each of its weak cluster points belong to* S^*.

Proof. If the solution set is nonempty, in view of (4) we have that $\|x^{k+1} - x^0\| \leq \|z - x^0\|$ for all $z \in H_k \cap W_k \cap C$. Since $S^* \subseteq H_k \cap W_k \cap C$ by Corollary 1, it follows that $\|x^{k+1} - x^0\| \leq \|x^* - x^0\|$ for all $x^* \in S^*$. Thus, $\{x^k\}$ is bounded, and by Theorem 1, all its weak cluster points belong to S^*. □

Finally, we are now ready to prove strong convergence of the sequence $\{x^k\}$ generated by Algorithm A to the solution which lies closest to x^0.

Theorem 2. *Assume that* $S^* \neq \emptyset$ *and let* $\{x^k\}$ *be a sequence generated by Algorithm A. Define* $x^* = P_{S^*}(x^0)$. *Then* $\{x^k\}$ *converges strongly to* x^*.

Proof. By Proposition 2, S^* is closed and convex. Therefore x^*, the orthogonal projection of x^0 onto S^*, exists. By the definition of x^{k+1}, we have that

$$\|x^{k+1} - x^0\| \leq \|z - x^0\| \qquad \forall\, z \in H_k \cap W_k \cap C. \tag{13}$$

Since $x^* \in S^* \subseteq H_k \cap W_k \cap C$ for all k, it follows from (13) that

$$\|x^k - x^0\| \leq \|x^* - x^0\|, \tag{14}$$

for all k. By Corollary 2, $\{x^k\}$ is bounded and each of its weak cluster points belongs to S^*. Let $\{x^{i_k}\}$ be any weakly convergent subsequence of $\{x^k\}$, and let $\hat{x} \in S^*$ be its weak limit. Observe that

$$\begin{aligned} \|x^{i_k} - x^*\|^2 &= \|x^{i_k} - x^0 - (x^* - x^0)\|^2 \\ &= \|x^{i_k} - x^0\|^2 + \|x^* - x^0\|^2 - 2\langle x^{i_k} - x^0, x^* - x^0 \rangle \\ &\leq 2\|x^* - x^0\|^2 - 2\langle x^{i_k} - x^0, x^* - x^0 \rangle, \end{aligned}$$

where the inequality follows from (14). By the weak convergence of $\{x^{i_k}\}$ to $\hat{x}$, we obtain

$$\limsup_{k\to\infty} \|x^{i_k} - x^*\|^2 \leq 2(\|x^* - x^0\|^2 - \langle \hat{x} - x^0, x^* - x^0\rangle). \tag{15}$$

Applying Proposition 1 with $K = S^*$, $x = x^0$ and $z = \hat{x} \in S^*$, and taking into account that x^* is the projection of x^0 onto S^*, we have that

$$\langle x^0 - x^*, \hat{x} - x^*\rangle \leq 0. \tag{16}$$

Now, using (16) we have

$$\begin{aligned} 0 &\geq -\langle \hat{x} - x^*, x^* - x^0\rangle = -\langle x^0 - x^*, x^* - x^0\rangle - \langle \hat{x} - x^0, x^* - x^0\rangle \\ &\geq \|x^* - x^0\|^2 - \langle \hat{x} - x^0, x^* - x^0\rangle. \end{aligned} \tag{17}$$

Combining (17) with (15), we conclude that $\{x^{i_k}\}$ converges strongly to x^*. Thus, we have shown that every weakly convergent subsequence of $\{x^k\}$ converges strongly to x^*. Hence, the whole sequence $\{x^k\}$ converges strongly to $x^* \in S^*$. □

Acknowledgment

Research for this paper was partially supported by PROCAD-nf - UFG/UnB/IMPA research and PRONEX - CNPq-FAPERJ - Optimization research. The first author thanks the Institute for Pure and Applied Mathematics (IMPA), at Rio de Janeiro, Brazil, where he was a visitor while working on this paper.

References

[1] Alber, Ya.I., Iusem, A.N. Extension of subgradient techniques for nonsmooth optimization in Banach spaces. *Set-Valued Analysis* **9** (2001) 315-335.

[2] Alber, Ya.I., Iusem, A.N., Solodov, M.V. On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. *Mathematical Programming* **81** (1998) 23-37.

[3] Allgower, E.L., Böhmer, K. Application of the mesh-independence principle to mesh refinement strategies. *SIAM Journal on Numerical Analysis* **24** (1987) 1335-1351.

[4] Allgower, E.L., Böhmer, K., Potra, F.-A., Rheinboldt, W.C. A mesh-independence principle for operator equations and their discretizations. *SIAM Journal on Numerical Analysis* **23** (1986) 160-169.

[5] Bauschke, H.H., Combettes, P.L. A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces. *Mathematics of Operations Research* **26** (2001) 248-264.

[6] Bello Cruz, J.Y., Iusem, A.N. A strongly convergent direct method for monotone variational inequalities in Hilbert spaces. *Numerical Functional Analysis and Optimization* **30** (2009) 23-36.

[7] Bello Cruz, J.Y., Lucambio Pérez, L.R. Convergence of a projected gradient method variant for quasiconvex objectives. *Nonlinear Analysis* **9** (2010) 2917-2922.

[8] Burachik, R., Graña Drummond, L.M., Iusem, A. N., Svaiter, B. F. Full convergence of the steepest descent method with inexact line searches. *Optimization* **32** (1995) 137-146.

[9] Gordon, R., Herman, G.T. Reconstruction of pictures from their projections. *Communications of the Association for Computing Machinery* **14** (1971) 759-768.

[10] Güler, O. New proximal point algorithms for convex minimization. *SIAM Journal on Optimization* **2** (1992) 649-664.

[11] Henry, J., Yvon, J-P. *System Modelling and Optimization.* Lecture Notes in Control and Information Sciences **197**. Springer, London (1994).

[12] Hudson, H.M., Larkin, R.S. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging* **13** (1994) 601-609.

[13] Iusem, A.N. On the convergence properties of the projected gradient method for convex optimization. *Computational and Applied Mathematics* **22** (2003) 37-52.

[14] Kiwiel, K.C., Murty, K. Convergence of the steepest descent method for minimizing quasiconvex functions. *Journal of Optimization Theory and Applications* **89** (1996) 221-226.

[15] Laumen, M. Newton's mesh independence principle for a class of optimal shape design problems. *SIAM Journal on Control and Optimization* **37** (1999) 1070-1088.

[16] Rockmore, A.J., Macovski, A. A maximum likelihood approach to transmission image reconstruction from projections. *IEEE Transactions on Nuclear Science* **24** (1977) 1929-1935.

[17] Polyak, B.T. Minimization of unsmooth functionals. *U.S.S.R. Computational Mathematics and Mathematical Physics* **9** (1969) 14-29.

[18] Shor, N.Z. *Minimization methods for nondifferentiable functions*. Springer, Berlin (1985).

[19] Solodov, M.V., Svaiter, B.F. Forcing strong convergence of proximal point iterations in a Hilbert space. *Mathematical Programming* **87** (2000) 189-202.

[20] Vanderlugt, A. *Optical Signal Processing*. Wiley, New York (1992).

[21] Wang, C., Xiu, N. Convergence of the gradient projection method for generalized convex minimization. *Computational Optimization and Applications* **16** (2000) 111-120.