J Optim Theory Appl (2011) 150:360–378
DOI 10.1007/s10957-011-9837-z

Averaged Mappings and the Gradient-Projection Algorithm

Hong-Kun Xu

Published online: 12 April 2011
© Springer Science+Business Media, LLC 2011

Abstract It is well known that the gradient-projection algorithm (GPA) plays an important role in solving constrained convex minimization problems. In this article, we first provide an alternative averaged mapping approach to the GPA. This approach is operator-oriented in nature. Since, in general, in infinite-dimensional Hilbert spaces, GPA has only weak convergence, we provide two modifications of GPA so that strong convergence is guaranteed. Regularization is also applied to find the minimum-norm solution of the minimization problem under investigation.

Keywords Averaged mapping · Gradient-projection algorithm · Constrained convex minimization · Maximal monotone operator · Relaxed gradient-projection algorithm · Regularization · Minimum-norm

1 Introduction

The gradient-projection (or projected-gradient) algorithm is a powerful tool for solving constrained convex optimization problems and has extensively been studied (see [1–4] and the references therein). It has recently been applied to solve split feasibility problems which find applications in image reconstructions and the intensity modulated radiation therapy (see [5–12]).

Communicated by J.-C. Yao.

The author would like to thank the referees for their helpful comments on this manuscript. He was supported in part by NSC 97-2628-M-110-003-MY3.

H.-K. Xu (✉)
Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung 804, Taiwan
e-mail: xuhk@math.nsysu.edu.tw

H.-K. Xu
Department of Mathematics, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

Let H be a real Hilbert space and C a nonempty closed and convex subset of H. Let $f : H \to \mathbb{R}$ be a convex and continuously Fréchet differentiable functional, and consider the problem of minimizing f over the constraint set C (assuming the existence of minimizers). The gradient-projection algorithm (GPA) generates a sequence $\{x_n\}_{n=0}^{\infty}$ determined by the gradient of f and the metric projection onto C. It is known [1] that if f has a Lipschitz continuous and strongly monotone gradient, then the sequence $\{x_n\}_{n=0}^{\infty}$ can be strongly convergent to a minimizer of f in C. If the gradient of f is only assumed to be Lipschitz continuous, then $\{x_n\}_{n=0}^{\infty}$ can only be weakly convergent if H is infinite-dimensional (a counterexample will be presented in Sect. 5).

Since the Lipschitz continuity of the gradient of f implies that it is indeed inverse strongly monotone (ism), its complement can be an averaged mapping. Consequently, the GPA can be rewritten as the composite of a projection and an averaged mapping, which is again an averaged mapping. This shows that averaged mappings play an important role in the gradient-projection algorithm. Recall that a mapping T is nonexpansive iff it is Lipschitz with Lipschitz constant not more than one and that a mapping is an averaged mapping iff it can be expressed as a proper convex combination of the identity mapping and a nonexpansive mapping. An averaged mapping with a fixed point is asymptotically regular and its Picard iterates at each point converge weakly to a fixed point of the mapping. This convergence property is quite helpful. As a matter of fact, this is the core of our idea in this present paper. In other words, we will use averaged mappings to study the convergence analysis of the GPA, which is therefore an operator-oriented approach.

Regularization, in particular, the traditional Tikhonov regularization, is usually used to solve ill-posed optimization problems. The advantage of a regularization method is its possible strong convergence to the minimum-norm solution of the optimization problem under investigation. The disadvantage is however its implicity, and hence explicit iterative methods seem more attractive, with which we are also concerned in this paper.

The organization of this paper is as follows. In Sect. 2, we introduce the gradient-projection algorithm and its convergence theorems already obtained in the existing literature. In Sect. 3, we introduce averaged mappings and maximal monotone operators, and their properties. We also discuss the relationship between averaged mappings and inverse strongly monotone operators. In Sect. 4, we present our averaged mapping approach to the GPA and the relaxed GPA. In Sect. 5, we first construct a counterexample, which shows that the GPA does not converge in norm in an infinite-dimensional space; we then provide two strongly convergent modifications of it. Section 6 is devoted to regularization; in particular, we provide an iterative algorithm which generates a sequence that converges in norm to the minimum-norm solution of the minimization problem under investigation. Finally, we conclude in Sect. 7.

2 Preliminaries

Consider the following constrained convex minimization problem:

$$\underset{x\in C}{\text{minimize}}\ f(x), \tag{1}$$

where C is a closed and convex subset of a Hilbert space H and $f: C \to \mathbb{R}$ is a real-valued convex function. If f is Fréchet differentiable, then the gradient-projection algorithm (GPA) generates a sequence $\{x_n\}_{n=0}^{\infty}$ according to the recursive formula

$$x_{n+1} := \mathrm{Proj}_C(x_n - \gamma \nabla f(x_n)), \quad n \geq 0, \tag{2}$$

or more generally,

$$x_{n+1} := \mathrm{Proj}_C(x_n - \gamma_n \nabla f(x_n)), \quad n \geq 0, \tag{3}$$

where, in both (2) and (3), the initial guess x_0 is taken from C arbitrarily, the parameters γ or γ_n are positive real numbers, and Proj_C is the metric projection from H onto C. The convergence of the algorithms (2) and (3) depends on the behavior of the gradient ∇f. As a matter of fact, it is known that, if ∇f is strongly monotone and Lipschitz continuous, namely, there are constants $\alpha > 0$ and $L > 0$ such that

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2, \quad x, y \in C \tag{4}$$

and

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad x, y \in C, \tag{5}$$

then, for $0 < \gamma < 2\alpha/L^2$, the operator

$$T := \mathrm{Proj}_C(I - \gamma \nabla f) \tag{6}$$

is a contraction; hence, the sequence $\{x_n\}_{n=0}^{\infty}$ defined by the GPA (2) converges in norm to the unique solution of (1). More generally, if the sequence $\{\gamma_n\}_{n=0}^{\infty}$ is chosen to satisfy the property

$$0 < \liminf_{n\to\infty} \gamma_n \leq \limsup_{n\to\infty} \gamma_n < \frac{2\alpha}{L^2}, \tag{7}$$

then the sequence $\{x_n\}_{n=0}^{\infty}$ defined by the GPA (3) converges in norm to the unique minimizer of (1).

However, if the gradient ∇f fails to be strongly monotone, the operator T defined in (6) may fail to be contractive; consequently, the sequence $\{x_n\}_{n=0}^{\infty}$ generated by the algorithm (2) may fail to converge strongly (see Sect. 4). The following states that, if the Lipschitz condition (5) holds, then the algorithms (2) and (3) can still converge in the weak topology.

Theorem 2.1 *Assume that problem* (1) *is consistent* (*i.e.*, (1) *is solvable*) *and the gradient* ∇f *satisfies the Lipschitz condition* (5). *Let* $\{\gamma_n\}_{n=0}^{\infty}$ *satisfy*

$$0 < \liminf_{n\to\infty} \gamma_n \leq \limsup_{n\to\infty} \gamma_n < \frac{2}{L}. \tag{8}$$

Then the GPA (3) *converges weakly to a minimizer of* (1).

The proof of Theorem 2.1, given in the current existing literature, heavily depends on the function f; see Levitin and Polyak [1]. One of the aims of this paper is to give an alternative operator-oriented approach to the GPA (3), namely, an averaged mapping approach. This will be done in Sect. 4.

3 Averaged Mappings and Monotone Operators

In this section, we introduce the concepts of projections, nonexpansive mappings, averaged mappings, and monotone operators.

Assume that H is a Hilbert space and K a nonempty closed and convex subset of H. Recall that the (nearest point or metric) projection from H onto K, denoted Proj_K, assigns, to each $x \in H$, the unique point $\mathrm{Proj}_K x \in K$ with the property

$$\|x - \mathrm{Proj}_K x\| = \inf\{\|x - y\| : y \in K\}.$$

Some useful properties of projections are gathered in the proposition below.

Proposition 3.1 *Given $x \in H$ and $z \in K$, we have*:

(i) $z = \mathrm{Proj}_K x$ *iff* $\langle x - z, y - z\rangle \le 0, \quad y \in K$;
(ii) $\langle x - y, \mathrm{Proj}_K x - \mathrm{Proj}_K y\rangle \ge \|\mathrm{Proj}_K x - \mathrm{Proj}_K y\|^2, \quad x, y \in H$;
(iii) $\|x - \mathrm{Proj}_K x\|^2 \le \|x - y\|^2 - \|y - \mathrm{Proj}_K x\|^2, \quad x \in H, y \in K$.

Definition 3.1 A mapping $T : H \to H$ is said to be

(a) nonexpansive, iff $\|Tx - Ty\| \le \|x - y\|, \quad x, y \in H$;
(b) firmly nonexpansive, iff $2T - I$ is nonexpansive, or equivalently,
$\langle x - y, Tx - Ty\rangle \ge \|Tx - Ty\|^2, \quad x, y \in H$.

Alternatively, T is firmly nonexpansive, iff T can be expressed as

$$T = \frac{1}{2}(I + S),$$

where $S : H \to H$ is nonexpansive. Projections are firmly nonexpansive.

Definition 3.2 A mapping $T : H \to H$ is said to be an averaged mapping, iff it can be written as the average of the identity I and a nonexpansive mapping; that is,

$$T = (1 - \alpha)I + \alpha S, \tag{9}$$

where α is a number in $]0, 1[$ and $S : H \to H$ is nonexpansive. More precisely, when (9) holds, we say that T is α-averaged. Thus firmly nonexpansive mappings (in particular, projections) are $(1/2)$-averaged maps.

Proposition 3.2 ([6, 13]) *Let the operators $S, T, V : H \to H$ be given.*

(i) *If $T = (1 - \alpha)S + \alpha V$ for some $\alpha \in]0, 1[$ and if S is averaged and V is nonexpansive, then S is averaged.*

(ii) *T is firmly nonexpansive, iff the complement $I - T$ is firmly nonexpansive.*
(iii) *If $T = (1-\alpha)S + \alpha V$ for some $\alpha \in]0,1[$, S is firmly nonexpansive and V is nonexpansive, then T is averaged.*
(iv) *The composite of finitely many averaged mappings is averaged. That is, if each of the mappings $\{T_i\}_{i=1}^N$ is averaged, then so is the composite $T_1 \cdots T_N$. In particular, if T_1 is α_1-averaged and T_2 is α_2-averaged, where $\alpha_1, \alpha_2 \in]0,1[$, then the composite T_1T_2 is α-averaged, where $\alpha = \alpha_1 + \alpha_2 - \alpha_1\alpha_2$.*
(v) *If the mappings $\{T_i\}_{i=1}^N$ are averaged and have a common fixed point, then*

$$\bigcap_{i=1}^{N} Fix(T_i) = Fix(T_1 \cdots T_N).$$

Here the notation $Fix(T) \equiv Fix\ T$ denotes the set of fixed points of the mapping T; that is, $Fix\ T := \{x \in H : Tx = x\}$.

Averaged mappings are useful in the convergence analysis, due to the following result.

Proposition 3.3 ([14]) *Let $T : H \to H$ be an averaged mapping. Assume that T has a bounded orbit, i.e., $\{T^n x_0\}_{n=0}^\infty$ is bounded for some $x_0 \in H$. Then we have:*

(i) *T is asymptotically regular, that is, $\lim_{n\to\infty} \|T^{n+1}x - T^n x\| = 0$ for all $x \in H$;*
(ii) *for any $x \in H$, the sequence $\{T^n x\}_{n=0}^\infty$ converges weakly to a fixed point of T.*

The so-called demiclosedness principle for nonexpansive mappings will often be used.

Lemma 3.1 (Demiclosedness Principle) ([14]) *Let C be a closed and convex subset of a Hilbert space H and let $T : C \to C$ be a nonexpansive mapping with $Fix\ T \neq \emptyset$. If $\{x_n\}_{n=1}^\infty$ is a sequence in C weakly converging to x and if $\{(I-T)x_n\}_{n=1}^\infty$ converges strongly to y, then $(I-T)x = y$. In particular, if $y = 0$, then $x \in Fix\ T$.*

We next introduce monotonicity of nonlinear operators. Given is a nonlinear operator A with domain $D(A)$ and range $R(A)$ in a Hilbert space H.

Definition 3.3 (See [15] for comprehensive theory of monotone operators.)

(i) A is monotone iff, for all $x, y \in D(A)$,

$$\langle x - y, Ax - Ay \rangle \geq 0.$$

(ii) Given is a number $\beta > 0$. T is said to be *β-strongly monotone*, iff

$$\langle x - y, Ax - Ay \rangle \geq \beta \|x - y\|^2, \quad x, y \in H.$$

(iii) Given is a number $\nu > 0$. T is said to be *ν-inverse strongly monotone* (ν-ism), iff

$$\langle x - y, Ax - Ay \rangle \geq \nu \|Ax - Ay\|^2, \quad x, y \in H.$$

It is easily seen that, if T is nonexpansive, then $I-T$ is monotone. It is also easily seen that a projection Proj_K is a one-ism.

Inverse strongly (also referred to as co-coercive) monotone operators have widely been applied to solve practical problems in various fields; for instance, in traffic assignment problems (see [16, 17]).

The following proposition gathers some results on the relationship between averaged mappings and inverse strongly monotone operators.

Proposition 3.4 ([6, 18]) *Let $T : H \to H$ be given. We have*:

(i) *T is nonexpansive, iff the complement $I-T$ is $(1/2)$-ism*;
(ii) *if T is ν-ism, then for $\gamma > 0$, γT is (ν/γ)-ism*;
(iii) *T is averaged, iff the complement $I-T$ is ν-ism for some $\nu > 1/2$; indeed, for $\alpha \in]0,1[$, T is α-averaged, iff $I-T$ is $(1/2\alpha)$-ism.*

The following elementary result on real sequences is quite well known [19].

Lemma 3.2 *Assume that $\{a_n\}_{n=0}^{\infty}$ is a sequence of non-negative real numbers such that*

$$a_{n+1} \le (1-\gamma_n)a_n + \gamma_n\delta_n + \beta_n, \quad n \ge 0,$$

where $\{\gamma_n\}_{n=0}^{\infty}$ and $\{\beta_n\}_{n=0}^{\infty}$ are sequences in $]0,1[$ and $\{\delta_n\}_{n=0}^{\infty}$ is a sequence in $\mathbb{R}$ such that

(i) $\sum_{n=0}^{\infty}\gamma_n = \infty$;
(ii) *either* $\limsup_{n\to\infty}\delta_n \le 0$ *or* $\sum_{n=0}^{\infty}\gamma_n|\delta_n| < \infty$;
(iii) $\sum_{n=0}^{\infty}\beta_n < \infty$.

Then $\lim_{n\to\infty} a_n = 0$.

We adopt the following notation:

- $x_n \to x$ means that $x_n \to x$ strongly;
- $x_n \rightharpoonup x$ means that $x_n \to x$ weakly;
- $\omega_w(x_n) := \{x : \exists x_{n_j} \rightharpoonup x\}$ is the weak ω-limit set of the sequence $\{x_n\}_{n=1}^{\infty}$.

4 The Gradient-Projection Algorithm

In this section we first give an operator-oriented proof of the weak convergence of the gradient-projection algorithm. First we need a technical lemma whose proof is an immediate consequence of Opial's property [20] of a Hilbert space and is hence omitted.

Lemma 4.1 *Let K be a nonempty closed and convex subset of a real Hilbert space H. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence in H satisfying the properties*:

(i) $\lim_{n\to\infty}\|x_n - x\|$ *exists for each $x \in K$; and*
(ii) $\omega_w(x_n) \subset K$.

Then $\{x_n\}_{n=1}^{\infty}$ *is weakly convergent to a point in* K.

Remark 4.1 Recall that a sequence $\{x_n\}_{n=1}^{\infty}$ in H is said to be (a) Féjer-monotone with respect to K [21] if $\|x_{n+1} - x\| \le \|x_n - x\|$ for all n and $x \in K$; or, more generally, (b) quasi-Féjer-monotone with respect to K [13] if there is a summable sequence $\{\varepsilon_n\}_{n=1}^{\infty}$ of non-negative numbers such that $\|x_{n+1} - x\| \le \|x_n - x\| + \varepsilon_n$ for all n and $x \in K$. It is known that the quasi-Féjer-monotonicity condition (b) implies condition (i) in Lemma 4.1. So the quasi-Féjer-monotonicity of $\{x_n\}_{n=1}^{\infty}$ w.r.t. K and condition (ii) imply the weak convergence of $\{x_n\}_{n=1}^{\infty}$ [21, 22].

Next result is our averaged mapping approach to the gradient-projection algorithm.

Theorem 4.1 *Assume that the minimization problem* (1) *is consistent and let* S *denote its solution set. Assume that the gradient* ∇f *satisfies the Lipschitz condition* (5). *Let the sequence of parameters* $\{\gamma_n\}_{n=0}^{\infty}$ *satisfy the condition* (8). *Then the sequence* $\{x_n\}_{n=0}^{\infty}$ *generated by the GPA* (3) *converges weakly to a minimizer of* (1).

Proof First observe that $x^* \in C$ solves the minimization problem (1) if and only if x solves the fixed-point equation

$$x^* = \mathrm{Proj}_C(I - \gamma\nabla f)x^*, \tag{10}$$

where $\gamma > 0$ is any fixed positive number.

For the sake of simplicity, we may assume that (due to condition (8))

$$0 < a \le \gamma_n \le b < \frac{2}{L}$$

for all n, where a, b are constants.

Note that the Lipschitz condition (5) implies that the gradient ∇f is $(1/L)$-ism [23], which then implies that $\gamma\nabla f$ is $(1/\gamma L)$-ism. So by Proposition 3.4(iii), $I - \gamma\nabla f$ is $(\gamma L/2)$-averaged. Now since the projection Proj_C is $(1/2)$-averaged, we see from Proposition 3.2(iv) that the composite $\mathrm{Proj}_C(I - \gamma\nabla f)$ is $((2+\gamma L)/4)$-averaged for $0 < \gamma < 2/L$. Hence we have that, for each n, $\mathrm{Proj}_C(I - \gamma_n\nabla f)$ is $((2+\gamma_n L)/4)$-averaged. Therefore, we can write

$$\mathrm{Proj}_C(I - \gamma_n\nabla f) = \frac{2-\gamma_n L}{4} I + \frac{2+\gamma_n L}{4} T_n = (1-\beta_n)I + \beta_n T_n, \tag{11}$$

where T_n is nonexpansive and $\beta_n = (2+\gamma_n L)/4 \in [a_1, b_1] \subset (0,1)$, where $a_1 = (2+aL)/4$ and $b_1 = (2+bL)/4 < 1$. Then we can rewrite (3) as

$$x_{n+1} = \mathrm{Proj}_C(I - \gamma_n\nabla f)x_n = (1-\beta_n)x_n + \beta_n T_n x_n. \tag{12}$$

For any $x^* \in S$, noticing that $T_n x^* = x^*$, we have

$$\begin{aligned} \|x_{n+1} - x^*\|^2 &= (1-\beta_n)\|x_n - x^*\|^2 + \beta_n\|T_n x_n - x^*\|^2 - \beta_n(1-\beta_n)\|x_n - T_n x_n\|^2 \\ &\le \|x_n - x^*\|^2 - \beta_n(1-\beta_n)\|x_n - T_n x_n\|^2. \end{aligned} \tag{13}$$

It follows that the real non-negative sequence $\{\|x_n - x^*\|\}_{n=0}^{\infty}$ is nonincreasing. Hence,

$$\lim_{n\to\infty} \|x_n - x^*\| \quad \text{exists for all } x^* \in S. \tag{14}$$

From (13), we get

$$\|x_n - T_n x_n\|^2 \le \frac{1}{a_1(1-b_1)}\big(\|x_n - x^*\|^2 - \|x_{n+1} - x^*\|^2\big).$$

Hence,

$$\lim_{n\to\infty} \|x_n - T_n x_n\| = 0. \tag{15}$$

But $\|x_{n+1} - x_n\| \le b_1\|x_n - T_n x_n\|$, so we get

$$\lim_{n\to\infty} \|x_{n+1} - x_n\| = 0. \tag{16}$$

Next we prove that

$$\omega_w(x_n) \subset S. \tag{17}$$

Suppose $\hat{x} \in \omega_w(x_n)$ and $\{x_{n_j}\}_{j=1}^{\infty}$ is a subsequence of $\{x_n\}_{n=1}^{\infty}$ such that $x_{n_j} \rightharpoonup \hat{x}$; thus, $x_{n_j+1} \rightharpoonup \hat{x}$ by (16). We may assume that $\gamma_{n_j} \to \gamma$; then we have $0 < \gamma < 2/L$. Set $T := \mathrm{Proj}_C(I - \gamma\nabla f)$; then T is nonexpansive. Since $x_{n_j+1} = \mathrm{Proj}_C(I - \gamma_{n_j}\nabla f)x_{n_j}$ and $x_{n_j+1} - x_{n_j} \to 0$ in norm, we get

$$\begin{aligned}
\|x_{n_j} - Tx_{n_j}\| &\le \|x_{n_j+1} - x_{n_j}\| + \|\mathrm{Proj}_C(I - \gamma_{n_j}\nabla f)x_{n_j} - \mathrm{Proj}_C(I - \gamma\nabla f)x_{n_j}\| \\
&\le \|x_{n_j+1} - x_{n_j}\| + |\gamma_{n_j} - \gamma|\|\nabla f(x_{n_j})\| \\
&\le \|x_{n_j+1} - x_{n_j}\| + M|\gamma_{n_j} - \gamma| \to 0.
\end{aligned}$$

It then follows from Lemma 3.1 that $\hat{x} \in Fix\ T$. But $Fix\ T = S$, we therefore have $\hat{x} \in S$.

Finally, by virtue of (14) and (17), we can apply Lemma 4.1 to S to see that $\{x_n\}_{n=1}^{\infty}$ converges weakly to a point of S. □

Next we prove the convergence of a relaxed gradient-projection algorithm.

Theorem 4.2 *Assume that the minimization problem* (1) *is consistent and let S denote its solution set. Assume that the gradient ∇f satisfies the Lipschitz condition* (5). *Let a sequence $\{x_n\}_{n=0}^{\infty}$ be generated by the following relaxed gradient-projection algorithm*:

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n \mathrm{Proj}_C(x_n - \gamma_n\nabla f(x_n)), \quad n = 0, 1, 2, \ldots. \tag{18}$$

Assume that the sequences of parameters $\{\gamma_n\}_{n=0}^{\infty}$ and $\{\alpha_n\}_{n=0}^{\infty}$ satisfy condition (8) *and the following conditions*:

$$0 < \alpha_n < \frac{4}{2+\gamma_n L} \quad (\forall n) \quad \textit{and} \quad 0 < \liminf_{n\to\infty}\alpha_n \le \limsup_{n\to\infty}\alpha_n < \frac{4}{2 + L\cdot\limsup_{n\to\infty}\gamma_n}. \tag{19}$$

Then the sequence $\{x_n\}_{n=0}^{\infty}$ *converges weakly to a minimizer of* (1).

Proof Setting

$$V_n := \text{Proj}_C(I - \gamma_n \nabla f),$$

we can rewrite x_{n+1} as

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n V_n x_n = (1-\alpha_n\beta_n)x_n + \alpha_n\beta_n T_n x_n,$$

where T_n is as defined in (11) and $\beta_n = (2+\gamma_n L)/4$. It is easy to find from (19) that

$$0 < \liminf_{n\to\infty} \alpha_n\beta_n \le \limsup_{n\to\infty} \alpha_n\beta_n < 1. \tag{20}$$

Repeating the proof of Theorem 4.1, we see that relations (14)–(17) remain valid. Therefore, an application of Lemma 4.1 yields that $\{x_n\}_{n=0}^{\infty}$ converges weakly to a point of S. □

If, in Theorem 4.2, we choose the parameters λ_n to be independent of n, then we can allow the sequence $\{\alpha_n\}$ to tend to either zero or $(2+\gamma L)/4$.

Theorem 4.3 *Assume that the minimization problem* (1) *is consistent and let* S *denote its solution set. Assume that the gradient* ∇f *satisfies the Lipschitz condition* (5). *Let a sequence* $\{x_n\}_{n=0}^{\infty}$ *be generated by the following relaxed gradient-projection algorithm*:

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n \text{Proj}_C(x_n - \gamma\nabla f(x_n)), \quad n = 0, 1, 2, \ldots, \tag{21}$$

where $0 < \gamma < 2/L$ *and* $0 \le \alpha_n \le (2+\gamma L)/4$ *for all* n. *Assume, in addition, the sequence* $\{\alpha_n\}_{n=0}^{\infty}$ *satisfies the condition*

$$\sum_{n=1}^{\infty} \alpha_n \left(\frac{4}{2+\gamma L} - \alpha_n\right) = \infty. \tag{22}$$

Then the sequence $\{x_n\}_{n=0}^{\infty}$ *converges weakly to a minimizer of* (1).

To prove Theorem 4.3, we need the following classic result of convergence of Mann's iterative method for nonexpansive mappings in the setting of Hilbert spaces.

Lemma 4.2 [24] *Let* T *be a nonexpansive self-mapping of a closed and convex subset* C *of a Hilbert space with a fixed point. Let* $\{\alpha_n\}_{n=0}^{\infty}$ *be a sequence in* $[0, 1]$ *such that*

$$\sum_{n=0}^{\infty} \alpha_n(1-\alpha_n) = \infty. \tag{23}$$

Then, initializing with $x_0 \in C$, *the sequence* $\{x_n\}_{n=1}^{\infty}$ *generated by Mann's iteration*

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n T x_n, \quad n \ge 0 \tag{24}$$

converges weakly to a fixed point of T.

Proof of Theorem 4.3 Since $0 < \gamma < 2/L$, $\mathrm{Proj}_C(I - \gamma\nabla f)$ is $((2 + \gamma L)/4)$-averaged. Hence there are $\beta \in (0, 1)$ and a nonexpansive T such that

$$\mathrm{Proj}_C(I - \gamma\nabla f) = (1 - \beta)I + \beta T.$$

Consequently, the algorithm (21) is rewritten as

$$x_{n+1} = (1 - \hat{\alpha}_n)x_n + \hat{\alpha}_n T x_n, \quad n = 0, 1, 2, \dots, \tag{25}$$

where $\hat{\alpha}_n = \beta\alpha_n \in [0, 1]$. Since it is easily seen that condition (22) is equivalent to the condition

$$\sum_{n=0}^{\infty} \hat{\alpha}_n(1 - \hat{\alpha}_n) = \infty,$$

we can apply Lemma 4.2 to (25) to conclude that $\{x_n\}_{n=0}^{\infty}$ converges weakly to a fixed point of T. □

Remark 4.2 From (19) and (22) we see that the parameters α_n in both (18) and (21) can be taken to be bigger than one.

5 Strong Convergence

The gradient-projection algorithm has, in general, weak convergence only, unless the underlying Hilbert space is finite-dimensional. Indeed, basing on Hundal [25], we can construct a counterexample as follows.

Example 5.1 In the space $H = l^2$, Hundal [25] constructed two closed and convex subsets K and L such that (see also [26–28])

(i) $K \cap L \neq \emptyset$;
(ii) the sequence $\{x_n\}_{n=0}^{\infty}$ generated by alternating projections,

$$x_n = (\mathrm{Proj}_L \circ \mathrm{Proj}_K)^n x_0, \quad n \geq 0 \tag{26}$$

with $x_0 \in L$, converges weakly, but not strongly. (Hundal's counterexample settles in the negative the question whether alternating projections onto closed convex subsets of a Hilbert space can have strong convergence, which remained open for nearly 40 years.)

Now consider the convex minimization problem

$$\min_{x\in L} f(x) := \frac{1}{2}\|x - \mathrm{Proj}_K x\|^2.$$

The gradient of f is

$$\nabla f(x) = (I - \mathrm{Proj}_K)x, \quad x \in l^2.$$

Note that ∇f is firmly nonexpansive and thus is Lipschitzian with Lipschitz constant one. Taking an initial guess $x_0 \in L$ and $\gamma = 1$, we see that the gradient-projection algorithm (2) generates a sequence $\{x_n\}_{n=0}^{\infty}$ which coincides with the sequence $\{x_n\}_{n=0}^{\infty}$ given in (26). Therefore, the gradient-projection algorithm generates weakly (not strongly) convergent sequences in general in infinite-dimensional spaces.

This gives naturally rise to a question how to appropriately modify the gradient-projection algorithm so as to have strong convergence. Below we include two such modifications, one is simply a convex combination of a contraction with the point that is generated by the projected-gradient algorithm, and the other involves additional projections. Both modifications are adaptations of those modifications [19, 29–31] for Rockafellar's proximal point algorithm [32] which has only weak convergence in infinite-dimensional Hilbert spaces [26, 33]. Moreover, the first modification is of viscosity nature [34, 35].

Theorem 5.2 *Assume that the minimization problem* (1) *is consistent and the gradient* ∇f *satisfies the Lipschitz condition* (5). *Let* $h : C \to C$ *be a* ρ*-contraction with* $\rho \in [0, 1[$. *Let a sequence* $\{x_n\}_{n=0}^{\infty}$ *be generated by the following hybrid gradient-projection algorithm*:

$$x_{n+1} = \theta_n h(x_n) + (1 - \theta_n)\mathrm{Proj}_C(x_n - \gamma_n \nabla f(x_n)), \quad n = 0, 1, 2, \dots. \tag{27}$$

Assume that the sequence $\{\gamma_n\}_{n=0}^{\infty}$ *satisfies the condition* (8) *and, in addition, that the following conditions are satisfied for* $\{\gamma_n\}_{n=0}^{\infty}$ *and* $\{\theta_n\}_{n=0}^{\infty} \subset [0, 1]$:

(i) $\theta_n \to 0$;
(ii) $\sum_{n=0}^{\infty} \theta_n = \infty$;
(iii) $\sum_{n=0}^{\infty} |\theta_{n+1} - \theta_n| < \infty$;
(iv) $\sum_{n=0}^{\infty} |\gamma_{n+1} - \gamma_n| < \infty$.

Then the sequence $\{x_n\}_{n=0}^{\infty}$ *converges in norm to a minimizer of* (1) *which is also the unique solution of the variational inequality* (*VI*)

$$x^* \in S, \quad \langle (I - h)x^*, x - x^* \rangle \geq 0, \quad x \in S. \tag{28}$$

In other words, x^* *is the unique fixed point of the contraction* $\mathrm{Proj}_S h$, $x^* = (\mathrm{Proj}_S h)x^*$.

Proof (1°) The sequence $\{x_n\}_{n=0}^{\infty}$ is bounded. Indeed, we have, for $\bar{x} \in S$,

$$\begin{aligned}\|x_{n+1} - \bar{x}\| &= \|\theta_n[h(x_n) - h(\bar{x}) + h(\bar{x}) - \bar{x}] + (1 - \theta_n)(V_n x_n - \bar{x})\| \\ &\leq (1 - (1 - \rho)\theta_n)\|x_n - \bar{x}\| + \theta_n \|h(\bar{x}) - \bar{x}\| \\ &\leq \max\{\|x_n - \bar{x}\|, \|h(\bar{x}) - \bar{x}\|/(1 - \rho)\}.\end{aligned}$$

So, an induction argument shows that

$$\|x_n - \bar{x}\| \leq \max\left\{\|x_0 - \bar{x}\|, \frac{1}{1 - \rho}\|h(\bar{x}) - \bar{x}\|\right\}, \quad n \geq 0.$$

In particular, $\{x_n\}_{n=0}^{\infty}$ is bounded.

(2°) We prove that $\|x_{n+1}-x_n\| \to 0$ as $n \to \infty$. Let M be a constant such that

$$M > \max\left\{\sup_{n\geq 0}\|h(x_n)\|,\quad \sup_{k,n\geq 0}\|V_k x_n\|,\quad \sup_{n\geq 0}\|\nabla f(x_n)\|\right\}.$$

We compute

$$\begin{aligned}\|x_{n+1}-x_n\| &= \|\theta_n h(x_n) + (1-\theta_n)V_n x_n - [\theta_{n-1}h(x_{n-1}) + (1-\theta_{n-1})V_{n-1}x_{n-1}]\| \\ &= \|\theta_n[h(x_n)-h(x_{n-1})] + (1-\theta_n)(V_n x_n - V_n x_{n-1}) \\ &\quad + (\theta_n-\theta_{n-1})(h(x_{n-1}) - V_{n-1}x_{n-1}) \\ &\quad + (1-\theta_n)(V_n x_{n-1} - V_{n-1}x_{n-1})\| \\ &\leq (1-(1-\rho)\theta_n)\|x_n - x_{n-1}\| + 2M|\theta_n - \theta_{n-1}| \\ &\quad + \|V_n x_{n-1} - V_{n-1}x_{n-1}\| \end{aligned} \tag{29}$$

and

$$\begin{aligned}\|V_n x_{n-1} - V_{n-1}x_{n-1}\| &= \|\mathrm{Proj}_C(I-\gamma_n\nabla f)x_{n-1} - \mathrm{Proj}_C(I-\gamma_{n-1}\nabla f)x_{n-1}\| \\ &\leq \|(I-\gamma_n\nabla f)x_{n-1} - (I-\gamma_{n-1}\nabla f)x_{n-1}\| \\ &= |\gamma_n - \gamma_{n-1}|\|\nabla f(x_{n-1})\| \leq M|\gamma_n - \gamma_{n-1}|. \end{aligned} \tag{30}$$

Combining (29) and (30), we obtain

$$\|x_{n+1}-x_n\| \leq (1-(1-\rho)\theta_n)\|x_n - x_{n-1}\| + 2M(|\theta_n-\theta_{n-1}| + |\gamma_n - \gamma_{n-1}|). \tag{31}$$

Apply Lemma 3.2 to (31) to conclude that $\|x_{n+1}-x_n\| \to 0$ as $n \to \infty$.

(3°) We prove that $\omega_w(x_n) \subset S$. Let $\hat{x} \in \omega_w(x_n)$ and assume that $x_{n_j} \rightharpoonup \hat{x}$ for some subsequence $\{x_{n_j}\}_{j=1}^{\infty}$ of $\{x_n\}_{n=0}^{\infty}$. We may further assume that $\gamma_{n_j} \to \gamma \in]0, 2/L[$, due to condition (8). Set $V := \mathrm{Proj}_C(I-\gamma\nabla f)$. Notice that V is nonexpansive and *Fix* $T = S$. It turns out that

$$\begin{aligned}\|x_{n_j} - Vx_{n_j}\| &\leq \|x_{n_j} - V_{n_j}x_{n_j}\| + \|V_{n_j}x_{n_j} - Vx_{n_j}\| \\ &\leq \|x_{n_j} - x_{n_j+1}\| + \|x_{n_j+1} - V_{n_j}x_{n_j}\| \\ &\quad + \|\mathrm{Proj}_C(I-\gamma_{n_j}\nabla f)x_{n_j} - \mathrm{Proj}_C(I-\gamma\nabla f)x_{n_j}\| \\ &\leq \|x_{n_j} - x_{n_j+1}\| + \theta_{n_j}\|h(x_{n_j}) - V_{n_j}x_{n_j}\| \\ &\quad + \|(I-\gamma_{n_j}\nabla f)x_{n_j} - (I-\gamma\nabla f)x_{n_j}\| \\ &\leq \|x_{n_j} - x_{n_j+1}\| + 2M(\theta_{n_j} + |\gamma_{n_j} - \gamma|) \to 0 \quad \text{as } j \to \infty. \end{aligned}$$

So Lemma 3.1 guarantees that $\omega_w(x_n) \subset$ *Fix* $V = S$.

(4°) We prove that $x_n \to x^*$ as $n \to \infty$, where x^* is the unique solution of the VI (28). First observe that there is some $\hat{x} \in \omega_w(x_n) \subset S$ such that

$$\limsup_{n\to\infty}\langle h(x^*) - x^*, x_n - x^*\rangle = \langle h(x^*) - x^*, \hat{x} - x^*\rangle \leq 0. \tag{32}$$

We now compute

$$\begin{aligned}
\|x_{n+1}-x^*\|^2 &= \|\theta_n(h(x_n)-x^*)+(1-\theta_n)(V_nx_n-x^*)\|^2\\
&= \|\theta_n(h(x_n)-h(x^*))+(1-\theta_n)(V_nx_n-x^*)+\theta_n(h(x^*)-x^*)\|^2\\
&\le \|\theta_n(h(x_n)-h(x^*))+(1-\theta_n)(V_nx_n-x^*)\|^2\\
&\quad + 2\theta_n\langle h(x^*)-x^*, x_{n+1}-x^*\rangle\\
&\le \theta_n\|h(x_n)-h(x^*)\|^2+(1-\theta_n)\|V_nx_n-x^*\|^2\\
&\quad + 2\theta_n\langle h(x^*)-x^*, x_{n+1}-x^*\rangle\\
&\le \big(1-(1-\rho^2)\theta_n\big)\|x_n-x^*\|^2+2\theta_n\langle h(x^*)-x^*, x_{n+1}-x^*\rangle. \qquad (33)
\end{aligned}$$

Applying Lemma 3.2 to the inequality (33), together with (32), we get $\|x_n-x^*\|\to 0$ as $n\to\infty$. □

Corollary 5.3 *Assume that the minimization problem* (1) *is consistent and that the gradient* ∇f *satisfies the Lipschitz condition* (5). *Let* $u\in C$ *be a fixed point in* C. *Define a sequence* $\{x_n\}_{n=0}^{\infty}$ *by the following hybrid gradient-projection algorithm*:

$$x_{n+1}=\theta_n u+(1-\theta_n)\mathrm{Proj}_C(x_n-\gamma_n\nabla f(x_n)),\quad n=0,1,2,\ldots. \qquad (34)$$

Assume that the sequence $\{\gamma_n\}_{n=0}^{\infty}$ *satisfies the condition* (8) *and, in addition, the conditions* (i)–(iv) *in Theorem* 5.2 *are satisfied. Then the sequence* $\{x_n\}_{n=0}^{\infty}$ *converges in norm to the minimizer* x^* *of* (1) *which is closest to* u *from the solution set* S. *In other words*, $x^*=\mathrm{Proj}_S u$.

Proof Taking $h(x)\equiv u$ in (27) and noting that for this choice of h, the unique solution x^* of the corresponding VI (28) coincides with $\mathrm{Proj}_S u$. □

Our next result shows that additional projections applied to the gradient-projection algorithm can also guarantee strong convergence.

Theorem 5.4 *Assume that the minimization problem* (1) *is consistent and let* S *be its solution set. Assume that the gradient* ∇f *satisfies the Lipschitz condition* (5). *Let* $\{x_n\}_{n=0}^{\infty}$ *be generated by the following algorithm*:

$$\begin{cases}
x_0\in C \quad \text{chosen arbitrarily},\\
y_n=\mathrm{Proj}_C(x_n-\gamma_n\nabla f(x_n)),\\
C_n=\{z\in C:\|y_n-z\|\le\|x_n-z\|\},\\
Q_n=\{z\in C:\langle x_n-z, x_0-x_n\rangle\ge 0\},\\
x_{n+1}=\mathrm{Proj}_{C_n\cap Q_n}x_0.
\end{cases} \qquad (35)$$

Assume that the sequence $\{\gamma_n\}_{n=0}^{\infty}$ *satisfies the condition* (8). *Then* $\{x_n\}_{n=0}^{\infty}$ *converges strongly to* $\mathrm{Proj}_S x_0$.

To prove this theorem, we include a lemma below.

Lemma 5.1 [18] *Let K be a nonempty closed and convex subset of H. Let $\{x_n\}_{n=0}^{\infty}$ be a sequence in H and $u \in H$. Let $q = \mathrm{Proj}_K u$. If $\{x_n\}_{n=0}^{\infty}$ satisfies the conditions*

(i) $\omega_w(x_n) \subset K$,
(ii) $\|x_n - u\| \le \|u - q\|$ *for all* n,

then $x_n \to q$ as $n \to \infty$.

Proof of Theorem 5.4 First observe that C_n is convex. Set

$$V_n := \mathrm{Proj}_C(I - \gamma_n \nabla f).$$

Next we show that $S \subset C_n \cap Q_n$ for all $n \ge 0$. Indeed, for all $p \in S$, since $V_n p = p$, we immediately get $\|y_n - p\| = \|V_n x_n - V_n p\| \le \|x_n - p\|$, showing that $p \in C_n$ for all $n \ge 0$.

To see that $S \subset Q_n$, we use induction. That $S \subset Q_0$ is evident since $Q_0 = C$. Assuming $S \subset Q_n$, we see that $x_{n+1} = \mathrm{Proj}_{C_n \cap Q_n} x_0$ is well defined; moreover, we have

$$\langle x_{n+1} - z, x_0 - x_{n+1} \rangle \ge 0 \quad \forall z \in C_n \cap Q_n.$$

As $S \subset C_n \cap Q_n$ by the induction assumption, the last inequality holds, in particular, for all $z \in S$. This, together with the definition of Q_{n+1}, implies that $S \subset Q_{n+1}$. Therefore, $S \subset Q_n$ holds for all $n \ge 0$.

Note that the definition of Q_n actually implies $x_n = \mathrm{Proj}_{Q_n} x_0$. This, together with the fact *Fix* $T \subset Q_n$, further implies

$$\|x_n - x_0\| \le \|p - x_0\| \quad \text{for all } p \in S.$$

In particular, $\{x_n\}_{n=0}^{\infty}$ is bounded and

$$\|x_n - x_0\| \le \|q - x_0\|, \quad \text{where } q = \mathrm{Proj}_S x_0. \tag{36}$$

The fact $x_{n+1} \in Q_n$ asserts that $\langle x_{n+1} - x_n, x_n - x_0 \rangle \ge 0$. Therefore,

$$\begin{aligned} \|x_{n+1} - x_n\|^2 &= \|(x_{n+1} - x_0) - (x_n - x_0)\|^2 \\ &= \|x_{n+1} - x_0\|^2 - \|x_n - x_0\|^2 - 2\langle x_{n+1} - x_n, x_n - x_0 \rangle \\ &\le \|x_{n+1} - x_0\|^2 - \|x_n - x_0\|^2. \end{aligned}$$

It turns out that

$$\|x_{n+1} - x_n\| \to 0. \tag{37}$$

By the fact $x_{n+1} \in C_n$ we get

$$\|x_{n+1} - y_n\| \le \|x_{n+1} - x_n\| \to 0 \quad \text{and} \quad \|x_n - y_n\| \to 0.$$

Since $V_n = (1 - \beta_n)I + \beta_n T_n$, with $\beta_n = (2 + \gamma_n L)/4 \ge 1/2$, we find

$$\|x_n - T_n x_n\| \le 2\|x_n - V_n x_n\| = 2\|x_n - y_n\| \to 0. \tag{38}$$

Now repeating the proof of Theorem 3.1, we conclude that $\omega_w(x_n) \subset S$. This, together with (36) and Lemma 5.1, implies that $x_n \to q$ as $n \to \infty$. □

6 Regularization and Minimum-Norm Solution

We now look at the minimum-norm solution of the minimization problem (1), assuming again $S := \arg\min_C f \neq \emptyset$. Let $x^\dagger$ be the element in S with minimum-norm; that is, $x^\dagger$ satisfies the properties: $x^\dagger \in S$ and $\|x^\dagger\| = \min\{\|x\| : x \in S\}$. In other words, $x^\dagger$ is the projection of the origin onto S. The objective of this section is to design an iterative algorithm that can be used to find $x^\dagger$. We observe that Corollary 5.3 and Theorem 5.4 actually provide the minimum-norm solution $x^\dagger$ if $0 \in C$. However, in this section, we use the idea of regularization to attach the general case (indeed, the regularization approach seems more popular in finding the minimum-norm solution).

Consider the regularized minimization problem

$$\min_{x\in C} f_\alpha(x) := f(x) + \frac{\alpha}{2}\|x\|^2. \tag{39}$$

Here $\alpha > 0$ is the regularization parameter, and again f is convex with L-Lipschitz continuous gradient ∇f.

Since now the gradient ∇f_α is α-strongly monotone and $(L+\alpha)$-Lipschitzian, (39) has a unique solution which is denoted as $x_\alpha \in C$ and which can be obtained via the Banach Contraction Principle. Indeed, if we choose γ such that $0 < \gamma < 2\alpha/(L+\alpha)^2$, then x_α is the unique fixed point of the mapping

$$V_\alpha := \mathrm{Proj}_C(I - \gamma\nabla f_\alpha) = \mathrm{Proj}_C(I - \gamma(\nabla f + \alpha I)).$$

Note that V_α is a contraction on C. As a matter of fact, it is easy to find that

$$\|V_\alpha x - V_\alpha y\| \le \sqrt{1 - \gamma\big(2\alpha - \gamma(L+\alpha)^2\big)}\|x - y\| \le \left(1 - \frac{1}{2}\alpha\gamma\right)\|x - y\| \tag{40}$$

provided $0 < \gamma \le \alpha/(L+\alpha)^2$.

Lemma 6.1 *Assume* $0 < \gamma \le \alpha/(L+\alpha)^2$ *and let* x_α *be the solution of* (39). *Then the strong* $\lim_{\alpha\to 0} x_\alpha = x^\dagger$.

Proof For any $\hat{x} \in S = \mathrm{argmin}_C f$, we have

$$\begin{aligned} f(\hat{x}) + \frac{\alpha}{2}\|x_\alpha\|^2 &\le f(x_\alpha) + \frac{\alpha}{2}\|x_\alpha\|^2 = f_\alpha(x_\alpha) \\ &\le f_\alpha(\hat{x}) = f(\hat{x}) + \frac{\alpha}{2}\|\hat{x}\|^2. \end{aligned}$$

It follows that

$$\|x_\alpha\| \le \|\hat{x}\| \quad (\forall \alpha > 0,\ \forall \hat{x} \in S). \tag{41}$$

Assume $\alpha_j \to 0$ and $x_{\alpha_j} \rightharpoonup \tilde{x}$. Then the weak lower semicontinuity of f implies that, for $x \in C$,

$$\begin{aligned} f(\tilde{x}) &\le \liminf_{j\to\infty} f(x_{\alpha_j}) \le \liminf_{j\to\infty} f_{\alpha_j}(x_{\alpha_j}) \\ &\le \liminf_{j\to\infty} f_{\alpha_j}(x) = \liminf_{j\to\infty}\left[f(x) + \frac{\alpha_j}{2}\|x\|^2\right] = f(x). \end{aligned}$$

It turns out that $\tilde{x} \in S$. This, together with (41), is sufficient to ensure that $x_\alpha \to x^\dagger$ which can also be obtained by applying Lemma 5.1 to the case where $K := S$, $u = 0$ and $q = x^\dagger$. □

Lemma 6.1 tells us that we can get the minimum-norm solution $x^\dagger$ through two steps. The first step is to employ Banach's Contraction Principle to get x_α via Picard's successive approximations: $V_\alpha^n x_0 \to x_\alpha$ as $n \to \infty$; the second step is to let $\alpha \to 0$, then the limit of the regularized solutions $\{x_\alpha\}$ is the minimum-norm solution $x^\dagger$. The result below shows that if we appropriately select the regularized parameters α and the stepsize parameter γ in the projection-gradient algorithm, then we can combine the two steps above to get a single iterative algorithm that generates a sequence strongly convergent to $x^\dagger$. Our combined algorithm generates a sequence $\{x_n\}_{n=0}^\infty$ in the following manner:

$$x_{n+1} = \mathrm{Proj}_C(I - \gamma_n \nabla f_{\alpha_n})x_n = \mathrm{Proj}_C(I - \gamma_n(\nabla f + \alpha_n I))x_n, \quad n \geq 0, \tag{42}$$

where the initial guess is $x_0 \in C$ and $\{\alpha_n\}_{n=0}^\infty$ and $\{\gamma_n\}_{n=0}^\infty$ are the parameter sequences satisfying certain conditions. We have the following convergence result.

Theorem 6.1 *Assume that the minimization problem* (1) *is consistent and let S denote its solution set. Assume that the gradient ∇f satisfies the Lipschitz condition* (5). *Let $\{x_n\}_{n=0}^\infty$ be generated by the iterative algorithm* (42). *Assume*

(i) $0 < \gamma_n \leq \alpha_n/(L + \alpha_n)^2$ *for all* n;
(ii) $\alpha_n \to 0$ (*and* $\gamma_n \to 0$) *as* $n \to \infty$;
(iii) $\sum_{n=1}^\infty \alpha_n \gamma_n = \infty$;
(iv) $(|\gamma_n - \gamma_{n-1}| + |\alpha_n \gamma_n - \alpha_{n-1}\gamma_{n-1}|)/(\alpha_n \gamma_n)^2 \to 0$ *as* $n \to \infty$.

Then $x_n \to x^\dagger$ *as* $n \to \infty$.

Proof First we show that $\{x_n\}_{n=0}^\infty$ is bounded. To see this, we take $\hat{x} \in S$ to get

$$\begin{aligned}\|x_{n+1} - \hat{x}\| &= \|\mathrm{Proj}_C(I - \gamma_n \nabla f_{\alpha_n})x_n - \mathrm{Proj}_C(I - \gamma_n \nabla f)\hat{x}\| \\ &\leq \|\mathrm{Proj}_C(I - \gamma_n \nabla f_{\alpha_n})x_n - \mathrm{Proj}_C(I - \gamma_n \nabla f_{\alpha_n})\hat{x}\| \\ &\quad + \|\mathrm{Proj}_C(I - \gamma_n \nabla f_{\alpha_n})\hat{x} - \mathrm{Proj}_C(I - \gamma_n \nabla f)\hat{x}\| \\ &\leq \left(1 - \frac{1}{2}\alpha_n \gamma_n\right)\|x_n - \hat{x}\| + \alpha_n \gamma_n \|\hat{x}\| \leq \max\{\|x_n - \hat{x}\|, 2\|\hat{x}\|\}.\end{aligned}$$

This implies by induction that

$$\|x_n - \hat{x}\| \leq \max\{\|x_0 - \hat{x}\|, 2\|\hat{x}\|\}, \quad n \geq 0.$$

Hence, $\{x_n\}_{n=0}^\infty$ is bounded.

Now let $z_n := z_{\alpha_n}$ be the unique fixed point of the contraction V_{α_n}. By Lemma 6.1, we get $z_n \to x^\dagger$. It remains to prove that $\|x_{n+1} - z_n\| \to 0$ as $n \to \infty$. Using (42), the fact that $z_n = V_{\alpha_n} z_n = \mathrm{Proj}_C(I - \gamma_n \nabla f_{\alpha_n}) z_n$ and the fact that V_{α_n} is a contraction with coefficient $(1 - \alpha_n \gamma_n/2)$, we derive that

$$\begin{aligned}\|x_{n+1} - z_n\| &\le \left(1 - \frac{1}{2}\alpha_n \gamma_n\right)\|x_n - z_n\| \\ &\le \left(1 - \frac{1}{2}\alpha_n \gamma_n\right)\|x_n - z_{n-1}\| + \|z_n - z_{n-1}\|. \end{aligned} \tag{43}$$

On the other hand, we have

$$\begin{aligned}\|z_n - z_{n-1}\| &= \|V_{\alpha_n} z_n - V_{\alpha_{n-1}} z_{n-1}\| \\ &\le \|V_{\alpha_n} z_n - V_{\alpha_n} z_{n-1}\| + \|V_{\alpha_n} z_{n-1} - V_{\alpha_{n-1}} z_{n-1}\| \\ &\le \left(1 - \frac{1}{2}\alpha_n \gamma_n\right)\|z_n - z_{n-1}\| \\ &\quad + \|(I - \gamma_n \nabla f_{\alpha_n}) z_{n-1} - (I - \gamma_{n-1} \nabla f_{\alpha_{n-1}}) z_{n-1}\| \\ &\le \left(1 - \frac{1}{2}\alpha_n \gamma_n\right)\|z_n - z_{n-1}\| + |\gamma_n - \gamma_{n-1}|\|\nabla f(z_{n-1})\| \\ &\quad + |\alpha_n \gamma_n - \alpha_{n-1}\gamma_{n-1}|\|z_{n-1}\| \\ &\le \left(1 - \frac{1}{2}\alpha_n \gamma_n\right)\|z_n - z_{n-1}\| + \frac{M}{2}(|\gamma_n - \gamma_{n-1}| \\ &\quad + |\alpha_n \gamma_n - \alpha_{n-1}\gamma_{n-1}|), \end{aligned} \tag{44}$$

where M is a constant big enough so that $M > 2\max\{\|z_n\|, \|\nabla f(z_n)\|\}$ for all n. It follows from (44) that

$$\|z_n - z_{n-1}\| \le M \frac{|\gamma_n - \gamma_{n-1}| + |\alpha_n \gamma_n - \alpha_{n-1}\gamma_{n-1}|}{\alpha_n \gamma_n}. \tag{45}$$

Substituting (45) into (43) we get

$$\|x_{n+1} - z_n\| \le \left(1 - \frac{1}{2}\alpha_n \gamma_n\right)\|x_n - z_{n-1}\| + M \frac{|\gamma_n - \gamma_{n-1}| + |\alpha_n \gamma_n - \alpha_{n-1}\gamma_{n-1}|}{\alpha_n \gamma_n}. \tag{46}$$

By virtue of the conditions (iii) and (iv), we can apply Lemma 3.2 to the relation (46) to get $\|x_{n+1} - z_n\| \to 0$ as $n \to \infty$. □

Remark 6.1 If we take

$$\alpha_n = \frac{1}{(n+1)^\alpha}, \quad \gamma_n = \frac{1}{(n+1)^\gamma},$$

where α and γ are such that $0 < \alpha < \gamma < 1$ and $2\alpha + \gamma < 1$, then it is not hard to verify that conditions (i)–(iv) of Theorem 6.1 are all satisfied.

7 Conclusions

The gradient-projection algorithm (GPA) for solving constrained optimization problems has extensively been studied in both finite- and infinite-dimensional Hilbert spaces for quite a long time. In this paper, we have, first time in the literature, shown an averaged mapping approach to the GPA. We have also shown that relaxed GPAs (i.e., (18) and (21)) can also be used to solve constrained optimization problems.

Since the GPA fails, in general, to converge in norm in infinite-dimensional Hilbert spaces, we have provided two strongly convergent modifications of it; one of which is of viscosity nature and the other of projection nature.

We have introduced a regularization technique and its related iterative algorithm that has been proved to converge in norm to the minimum-norm solution of the minimization problem (1). The feature of our method is a strategic combination of regularization and contractiveness, together with appropriate selections of the regularization parameter and the stepsize at each iteration.

References

1. Levitin, E.S., Polyak, B.T.: Constrained minimization methods. Zh. Vychisl. Mat. Mat. Fiz. **6**, 787–823 (1966)
2. Calamai, P.H., Moré, J.J.: Projected gradient methods for linearly constrained problems. Math. Program. **39**, 93–116 (1987)
3. Polyak, B.T.: Introduction to optimization. In: Optimization Software, New York (1987)
4. Su, M., Xu, H.K.: Remarks on the gradient-projection algorithm. J. Nonlinear Anal. Optim. **1**, 35–43 (2010)
5. Censor, Y., Elfving, T.: A multiprojection algorithm using Bregman projections in a product space. Numer. Algorithms **8**, 221–239 (1994)
6. Byrne, C.: A unified treatment of some iterative algorithms in signal processing and image reconstruction. Inverse Probl. **20**, 103–120 (2004)
7. Censor, Y., Elfving, T., Kopf, N., Bortfeld, T.: The multiple-sets split feasibility problem and its applications for inverse problems. Inverse Probl. **21**, 2071–2084 (2005)
8. Censor, Y., Bortfeld, T., Martin, B., Trofimov, A.: A unified approach for inversion problems in intensity-modulated radiation therapy. Phys. Med. Biol. **51**, 2353–2365 (2006)
9. Xu, H.K.: A variable Krasnosel'skii–Mann algorithm and the multiple-set split feasibility problem. Inverse Probl. **22**, 2021–2034 (2006)
10. Xu, H.K.: Iterative methods for the split feasibility problem in infinite-dimensional Hilbert spaces. Inverse Probl. **26**, 105018 (2010)
11. Lopez, G., Martin, V., Xu, H.K.: Perturbation techniques for nonexpansive mappings with applications. Nonlinear Anal., Real World Appl. **10**, 2369–2383 (2009)
12. Lopez, G., Martin, V., Xu, H.K.: Iterative algorithms for the multiple-sets split feasibility problem. In: Censor, Y., Jiang, M., Wang, G. (eds.) Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems, pp. 243–279. Medical Physics Publishing, Madison (2009)
13. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optimization **53**, 475–504 (2004)
14. Geobel, K., Kirk, W.A.: Topics in Metric Fixed Point Theory. Cambridge Studies in Advanced Mathematics, vol. 28. Cambridge University Press, Cambridge (1990)
15. Brezis, H.: Operateurs Maximaux Monotones et Semi-Groups de Contractions dans les Espaces de Hilbert. North-Holland, Amsterdam (1973)
16. Bertsekas, D.P., Gafni, E.M.: Projection methods for variational inequalities with applications to the traffic assignment problem. Math. Program. Stud. **17**, 139–159 (1982)
17. Han, D., Lo, H.K.: Solving non-additive traffic assignment problems: a descent method for co-coercive variational inequalities. Eur. J. Oper. Res. **159**, 529–544 (2004)

18. Martinez-Yanes, C., Xu, H.K.: Strong convergence of the CQ method for fixed-point iteration processes. Nonlinear Anal. **64**, 2400–2411 (2006)
19. Xu, H.K.: Iterative algorithms for nonlinear operators. J. Lond. Math. Soc. **66**, 240–256 (2002)
20. Opial, Z.: Weak convergence of the sequence of successive approximations of nonexpansive mappings. Bull. Am. Math. Soc. **73**, 595–597 (1967)
21. Bauschke, H.H., Combettes, P.L.: A weak-to-strong convergence principle for Féjer-monotone methods in Hilbert spaces. Math. Oper. Res. **26**, 248–264 (2001)
22. Browder, F.E.: Convergence theorems for sequences of nonlinear operators in Banach spaces. Math. Z. **100**, 201–225 (1967)
23. Baillon, J.B., Haddad, G.: Quelques proprietes des operateurs angle-bornes et n-cycliquement monotones. Isr. J. Math. **26**, 137–150 (1977)
24. Reich, S.: Weak convergence theorems for nonexpansive mappings in Banach spaces. J. Math. Anal. Appl. **67**, 274–276 (1979)
25. Hundal, H.: An alternating projection that does not converge in norm. Nonlinear Anal. **57**, 35–61 (2004)
26. Bauschke, H.H., Burke, J.V., Deutsch, F.R., Hundal, H.S., Vanderwerff, J.D.: A new proximal point iteration that converges weakly but not in norm. Proc. Am. Math. Soc. **133**, 1829–1835 (2005)
27. Bauschke, H.H., Matouskova, E., Reich, S.: Projections and proximal point methods: Convergence results and counterexamples. Nonlinear Anal. **56**, 715–738 (2004)
28. Matouskova, E., Reich, S.: The Hundal example revisited. J. Nonlinear Convex Anal. **4**, 411–427 (2003)
29. Solodov, M.V., Svaiter, B.F.: Forcing strong convergence of proximal point iterations in a Hilbert space. Math. Program., Ser. A **87**, 189–202 (2000)
30. Marino, G., Xu, H.K.: Convergence of generalized proximal point algorithm. Commun. Pure Appl. Anal. **3**, 791–808 (2004)
31. Xu, H.K.: A regularization method for the proximal point algorithm. J. Glob. Optim. **36**, 115–125 (2006)
32. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
33. Güler, O.: On the convergence of the proximal point algorithm for convex optimization. SIAM J. Control Optim. **29**, 403–419 (1991)
34. Moudafi, A.: Viscosity approximation methods for fixed-points problems. J. Math. Anal. Appl. **241**, 46–55 (2000)
35. Xu, H.K.: Viscosity approximation methods for nonexpansive mappings. J. Math. Anal. Appl. **298**, 279–291 (2004)