Proximest Point and Steepest Descent Algorithm Controlled by a Slowly Vanishing Term. Applications to Hierarchical Minimization

Alexandre Cabot

Rapport de recherche n° 2003-06
Déposé le 18 septembre 2003
PROXIMAL POINT AND STEEPEST DESCENT ALGORITHM CONTROLLED BY A SLOWLY VANISHING TERM. APPLICATIONS TO HIERARCHICAL MINIMIZATION

ALEXANDRE CABOT

Abstract. Given $H$ a Hilbert space, we consider $\Phi_0 : H \rightarrow \mathbb{R} \cup \{+\infty\}$ a closed convex function and $\Phi_1 : H \rightarrow \mathbb{R}$ a finite convex function that are bounded from below. Our goal is to build an algorithm which first minimizes the map $\Phi_0$ and secondly the map $\Phi_1$ over the set $S_0 := \text{argmin} \Phi_0$. For that purpose, we define the following proximal-type algorithm

$$(A_1) \quad -(x_{n+1} - x_n)/\lambda_n \in \partial_{\eta_n} (\Phi_0 + \varepsilon_n \Phi_1)(x_{n+1}),$$

where $(\lambda_n)$ is a positive step sequence; $(\eta_n)$ is a summable error sequence and $(\varepsilon_n)$ is a control sequence tending toward 0; $\partial_{\eta_n}$ denotes the $\eta_n$-approximate subdifferential. When $H$ is finite dimensional and $(\varepsilon_n)$ is a slow control, i.e. $\sum_{n=0}^{\infty} \varepsilon_n = +\infty$, we prove that, under adequate conditions, the sequence $(x_n)$ defined by $(A_1)$ tends toward an element of $S_1 := \text{argmin} S \Phi_1$.

More generally, given finite convex functions $\Phi_2, \ldots, \Phi_N : H \rightarrow \mathbb{R}$, let us define the sets $(S_i)_{i \in \{1, \ldots, N\}}$ by the recursive relation $S_i := \text{argmin} S_{i-1} \Phi_i$. We then introduce the algorithm $(A_N)$, which tends to minimize hierarchically each function $\Phi_i$ on the set $S_{i-1}$, for $i \in \{1, \ldots, N\}$.

Key words. steepest descent system, proximal point method, hierarchical optimization, convex minimization, slow control.

AMS subject classifications. 37N40, 49M30, 65K10.

1. Introduction. Let $H$ be a Hilbert space endowed with scalar product and corresponding norm respectively denoted by $(.,.)$ and $|.|$. Let $\Phi_1 : H \rightarrow \mathbb{R}$ a smooth convex function that we wish to minimize over the convex set $S$. A powerful method consists in following the orbits of a discrete or continuous dynamical system, hopefully converging toward some element of $\text{argmin} S \Phi_1$. It is classical to apply the steepest descent system to the function $\Phi_1 + \delta_S$ ($\delta_S$ is the indicator function of $S$), thus leading to

$$\dot{x}(t) + \nabla \Phi_1 (x(t)) \in -N_S(x(t)), \quad t \geq 0,$$

where $N_S(x(t))$ is the normal cone to $S$ at the point $x(t)$. It can be shown that this differential inclusion falls into the framework of gradient-projection methods (see Brezis [5]). Antipin [1] has initiated another continuous gradient-projection system, where the constraint $S$ is integrated through the projection operator $P_S$. In another direction, Cabot [6] has considered in a recent paper the following continuous dynamical system

$$(SDC) \quad \dot{x}(t) + \nabla \Phi_0(x(t)) + \varepsilon(t) \nabla \Phi_1 (x(t)) = 0, \quad t \geq 0,$$

where $\Phi_0 : H \rightarrow \mathbb{R}$ is a smooth convex function satisfying $\text{argmin} \Phi_0 = S$ and $\varepsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a control parameter tending to 0 when $t \rightarrow +\infty$. In the $(SDC)$ system, the information on the constraint $S$ is contained in the function $\Phi_0$. The main difference with the previous methods lies in the fact that we do not have to handle non-smooth operators, like the normal cone $N_S$ or the projection $P_S$. However, the difficulty comes from the choice of the control parameter $\varepsilon$. If the map $\varepsilon$ tends to 0 too fast, the potential $\Phi_1$ cannot sufficiently influence the trajectory $x(\cdot)$, so
that the minimization of $\Phi_1$ may not occur. We feel the interest of a “slow control” and it is shown in [6] that the adequate condition on $\varepsilon$ is \( \int_0^{+\infty} \varepsilon(t) \, dt = +\infty \). The notion of slow control has already been pointed out by several authors, essentially for continuous dynamical systems. See for example Cominetti [9] and Attouch-Cominetti [3] where the slow control is aimed at stabilizing a continuous gradient-like system toward a peculiar equilibrium. In the same direction, Attouch-Czarnecki [4], Cabot-Czarnecki [8] and Cabot [7] apply the notion of slow control to stabilize second-order in time systems.

Coming back to the (SDC) system and keeping in mind numerical applications, it is natural to deal with a discretized version of (SDC). In this paper, we will be especially interested in the following implicit discretization of (SDC):

\[-(x_{n+1} - x_n) / \lambda_n = \nabla (\Phi_0 + \varepsilon_n \Phi_1)(x_{n+1}), \quad n \in \mathbb{N},\]

where $\lambda_n$ is the step length at iteration $n$ and $\varepsilon_n$ is the value of $\varepsilon(\cdot)$ at time $t_n := \sum_{k=0}^{n-1} \lambda_k$. If the closed convex functions $\Phi_0$ and $\Phi_1$ take their values in $\mathbb{R} \cup \{+\infty\}$ without regularity assumptions, one can easily adapt the previous algorithm as follows:

\[-(x_{n+1} - x_n) / \lambda_n \in \partial (\Phi_0 + \varepsilon_n \Phi_1)(x_{n+1}),\]

where $\partial$ denotes the subdifferential in the sense of convex analysis. Notice that this algorithm falls into the field of proximal point methods proposed in [11, 12] and inspired by [13]. In the previous algorithm, the iterate $x_{n+1}$ is uniquely determined by $x_{n+1} = J_{\lambda_n}^A(x_n)$, where $A_n$ is the maximal monotone operator $\partial(\Phi_0 + \varepsilon_n \Phi_1)$ and $J_{\lambda}^A := (I + \lambda A)^{-1} : H \to H$ is the resolvent of $A$ of parameter $\lambda$ (see for example [5] for further details on maximal monotone operators and their resolvents). In order to deal with numerical applications, it is convenient to authorize at each iteration $n$ an error $\eta_n$ in the evaluation of the subdifferential. More precisely, denoting by $\partial_\eta$ the $\eta$-approximate subdifferential, we are led to the following algorithm:

\[(A_1) \quad -(x_{n+1} - x_n) / \lambda_n \in \partial_\eta (\Phi_0 + \varepsilon_n \Phi_1)(x_{n+1}).\]

The sequence $(\eta_n)$ of errors is assumed to be summable so as to remain close to the exact subdifferential. We show that, under the slow control criterion, i.e. $\sum_{n=0}^{+\infty} \varepsilon_n = +\infty$, then each sequence generated by $(A_1)$ tends to minimize $\Phi_1$ over $\text{argmin} \Phi_0$ in a sense that will be precised throughout the paper.

The next stage consists in building an algorithm which is able to minimize hierarchically several functions over their successive argmin sets. More precisely, consider finite convex functions $\Phi_2, \ldots, \Phi_N$ (with $N \geq 2$) and define their successive argmin sets by $S_0 := \text{argmin} \Phi_0$ and $S_i := \text{argmin} S_{i-1} \Phi_i$, for $i \in \{1, \ldots, N\}$. We introduce in the paper the algorithm $(A_N)$, given by

\[(A_N) \quad -(x_{n+1} - x_n) / \lambda_n \in \partial_\eta (\Phi_0 + \varepsilon_n \Phi_1 + \varepsilon_n^{q_2} \Phi_2 + \ldots + \varepsilon_n^{q_N} \Phi_N)(x_{n+1}),\]

where the choice of the exponents $q_2, \ldots, q_N$ strongly depends on the behaviour of the functions $\Phi_0, \ldots, \Phi_N$. We prove that, under adequate conditions on the sequence $(\varepsilon_n)$, the algorithm $(A_N)$ tends to minimize hierarchically each function $\Phi_i$ on the set $S_{i-1}$, for $i \in \{1, \ldots, N\}$. The introduction of the algorithms $(A_i)_{i \in \{1, \ldots, N\}}$ seems to be a new and promising tool in hierarchical minimization problems.
PROXIMAL POINT AND STEEPEST DESCENT ALGORITHM

The paper is organized as follows. In Section 2, we establish the general features about the algorithm \((A_1)\), based on energy estimates. We also study the case of a fast control \((\varepsilon_n)\) and we then show the weak convergence of the sequence \((x_n)\) toward some element of \(S_0 = \text{argmin } \Phi_0\). Section 3 is devoted to the case of a slow control. In the finite dimensional setting, we prove that the distance of the iterate \(x_n\) to the set \(S_1 = \text{argmin } S_0, \Phi_1\) tends toward 0. We give sufficient conditions ensuring that the sequence \((x_n)\) converges toward some element of \(S_1\). Finally in Section 4, we generalize the previous results by considering the algorithm \((A_N)\), which is shown to minimize hierarchically the respective functions \(\Phi_1, \ldots, \Phi_N\) over the respective sets \(S_0, \ldots, S_N\). For pedagogical reasons, the first paragraph of Section 4 starts with the case \(N = 2\).

2. General results. Case of a fastly vanishing term.

2.1. Algorithm \((A_1)\). General results. Let \(H\) be a Hilbert space endowed with scalar product and corresponding norm respectively denoted by \(\langle \cdot, \cdot \rangle\) and \(|\cdot|\). Let \(\Phi_0 : H \to \mathbb{R} \cup \{+\infty\}\) a closed convex function and \(\Phi_1 : H \to \mathbb{R}\) a finite convex function that are bounded from below. We are also given non-negative sequences \((\varepsilon_n)\), \((\lambda_n)\) and \((\eta_n)\). Denoting by \(\partial \eta\) the \(\eta\)-approximate subdifferential, we consider the following algorithm:

\[
(A_1) \quad - \frac{x_{n+1} - x_n}{\lambda_n} \in \partial \eta_n (\Phi_0 + \varepsilon_n \Phi_1)(x_{n+1}).
\]

There is no uniqueness in the choice of \(x_{n+1}\) and we let the reader check that the iterate \(x_{n+1}\) is characterized by the following inclusion:

\[
x_{n+1} \in \eta_n - \text{argmin } \left( \frac{|x - x_n|^2}{2 \lambda_n} + (\Phi_0 + \varepsilon_n \Phi_1)(x) \right),
\]

where \(\eta_n - \text{argmin}\) denotes the \(\eta\)-approximate set of absolute minimizers.

In the whole paper, we will assume the following hypotheses on the sequences \((\varepsilon_n)\), \((\lambda_n)\), \((\eta_n)\):

\(H_e\) The sequence \((\varepsilon_n)\) is non-increasing and \(\lim_{n \to +\infty} \varepsilon_n = 0\).

\(H_\lambda\) There exist \(\underline{\lambda} > 0\) and \(\overline{\lambda} > 0\) such that \(\underline{\lambda} \leq \lambda_n \leq \overline{\lambda}\), for every \(n \in \mathbb{N}\).

\(H_\eta\) The sequence \((\eta_n)\) is summable, i.e. \(\sum_{n=0}^{+\infty} \eta_n < +\infty\).

The summability condition \((H_\eta)\) means that the authorized error \(\eta_n\) in the computation of the subdifferential is small enough so as to remain close to the exact subdifferential. The following proposition states the main general features about the algorithm \((A_1)\).

Proposition 2.1. Let \(H\) be a Hilbert space endowed with scalar product and corresponding norm respectively denoted by \(\langle \cdot, \cdot \rangle\) and \(|\cdot|\). Let \(\Phi_0 : H \to \mathbb{R} \cup \{+\infty\}\) a closed convex function and \(\Phi_1 : H \to \mathbb{R}\) a finite convex function that are bounded from below. We are given non-negative sequences \((\varepsilon_n)\), \((\lambda_n)\), \((\eta_n)\) satisfying respectively \((H_e)\), \((H_\lambda)\), \((H_\eta)\). Let us consider a sequence \((x_n)\) generated by the algorithm \((A_1)\) and satisfying the initial condition \(x_0 \in \text{dom } \Phi_0\). Then, the following holds:

(i) \(\sum_{n=0}^{+\infty} |x_{n+1} - x_n|^2 < +\infty\) and in particular \(\lim_{n \to +\infty} x_{n+1} - x_n = 0\).

(ii) \(\lim_{n \to +\infty} \Phi_0(x_n) + \varepsilon_n \Phi_1(x_n)\) exists and hence the sequence \((\Phi_0(x_n))\) is bounded from above.

(iii) Assuming that the subsequence \((x_{n_k})\) of \((x_n)\) is bounded, then we have \(\lim_{k \to +\infty} \Phi_0(x_{n_k}) = \inf \Phi_0\). If moreover \(x_{n_k} \rightharpoonup \bar{x}\) weakly in \(H\), then \(\bar{x} \in \text{argmin } \Phi_0\).
Proof. (i) It is clear that the algorithm (A1) is unchanged by replacing \( \Phi_0 \) (resp. \( \Phi_1 \)) by \( \Phi_0 + \alpha \) (resp. \( \Phi_1 + \beta \)), where \( \alpha, \beta \in \mathbb{R} \). Therefore, without any loss of generality, one can assume that \( \inf \Phi_0 = \inf \Phi_1 = 0 \). We define the discrete energy by

\[
E_n = \Phi_0(x_n) + \varepsilon_n \Phi_1(x_n).
\]

Let us compute the quantity \( E_{k+1} - E_k, k \in \mathbb{N} \):

\[
E_{k+1} - E_k = (\Phi_0 + \varepsilon_{k+1} \Phi_1)(x_{k+1}) - (\Phi_0 + \varepsilon_k \Phi_1)(x_k)
= (\Phi_0 + \varepsilon_{k} \Phi_1)(x_{k+1}) - (\Phi_0 + \varepsilon_k \Phi_1)(x_k) - (\varepsilon_k - \varepsilon_{k+1}) \Phi_1(x_{k+1}).
\]

Since \( \varepsilon_{k+1} \leq \varepsilon_k \) and \( -(x_{k+1} - x_k)/\lambda_k \in \partial \eta_k (\Phi_0 + \varepsilon_k \Phi_1)(x_{k+1}) \), the previous equality gives

\[
E_{k+1} - E_k \leq \langle -(x_{k+1} - x_k)/\lambda_k, x_{k+1} - x_k \rangle + \eta_k
= -\left| x_{k+1} - x_k \right|^2/\lambda_k + \eta_k.
\]  \hspace{1cm} \text{(2.1)}

By summation from \( k = 0 \) to \( n - 1 \) of the previous inequalities, one obtains

\[
E_n - E_0 \leq - \sum_{k=0}^{n-1} |x_{k+1} - x_k|^2/\lambda_k + \sum_{k=0}^{n-1} \eta_k.
\]

Since \( \sum_{k=0}^{+\infty} \eta_k < +\infty \) and \( E_n \geq 0 \), we deduce

\[
\sum_{k=0}^{+\infty} |x_{k+1} - x_k|^2/\lambda_k \leq E_0 + \sum_{k=0}^{+\infty} \eta_k < +\infty.
\]

Since \( (\lambda_k) \) is bounded from below by a positive constant, we infer that \( \sum_{k=0}^{+\infty} |x_{k+1} - x_k|^2 < +\infty \), which proves (i).

(ii) From (2.1), we deduce that, for every \( k \geq 0 \),

\[
(E_{k+1} - E_k)_+ \leq \eta_k,
\]

where \([t]_+ \) denotes the positive part of \( t \): \([t]_+ := \max\{t, 0\} \). In view of the following lemma, we conclude that \( \lim_{n \to +\infty} E_n \) exists.

**Lemma 2.2.** Let \((g_n)\) be a sequence of non-negative reals satisfying \( \sum_{k=0}^{+\infty} (g_{k+1} - g_k)_+ < +\infty \). Then \( \lim_{n \to +\infty} g_n \) exists.

The proof of Lemma 2.2 is immediate and left to the reader. The rest of (ii) is immediate.

(iii) Assume that \((x_{n_k})\) is bounded. From the subdifferential inclusion (A1), we have for every \( \xi \in H \),

\[
(\Phi_0 + \varepsilon_{n_k} \Phi_1)(\xi) \geq (\Phi_0 + \varepsilon_{n_k} \Phi_1)(x_{n_k+1}) + \langle \xi - x_{n_k+1}, -(x_{n_k+1} - x_{n_k})/\lambda_{n_k} \rangle - \eta_{n_k}
\]

\[
\geq (\Phi_0)(x_{n_k+1}) + \inf \Phi_1 \varepsilon_{n_k} - |\xi - x_{n_k+1}||x_{n_k+1} - x_{n_k}|/\lambda_{n_k} - \eta_{n_k}.
\]

Let \( k \) go to \( +\infty \) and take the upper limit in the previous inequality. Taking into account the fact that \( \lim_{k \to +\infty} \varepsilon_{n_k} = \lim_{k \to +\infty} \eta_{n_k} = 0 \), that \( \lim_{k \to +\infty} |x_{n_k+1} - x_{n_k}| = 0 \) and that \((x_{n_k})\) is bounded, we obtain for every \( \xi \in H \)

\[
\Phi_0(\xi) \geq \limsup_{k \to +\infty} \Phi_0(x_{n_k+1}),
\]
i.e. \( \inf \Phi_0 \geq \limsup_{k \to +\infty} \Phi_0(x_{n_k+1}) \). Since the inequality \( \liminf_{k \to +\infty} \Phi_0(x_{n_k+1}) \geq \inf \Phi_0 \) always holds, we conclude that \( \lim_{k \to +\infty} \Phi_0(x_{n_k}) = \inf \Phi_0 \).

Assume now that \( x_{n_k} \rightharpoonup \bar{x} \) weakly in \( H \). Since the convex function \( \Phi_0 \) is lower semicontinuous for the weak topology, we have

\[
\Phi_0(\bar{x}) \leq \liminf_{k \to +\infty} \Phi_0(x_{n_k}) = \inf \Phi_0,
\]

and therefore \( \bar{x} \in \text{argmin} \Phi_0. \) \( \square \)

2.2. Fast control. First convergence results. When \( \varepsilon_n = 0 \) for every \( n \geq 0 \), the \((A_1)\) algorithm reduces to the standard proximal point method applied to \( \Phi_0 \). The sequence \((x_n)\) generated by \((A_1)\) is then known to weakly converge toward a minimum of \( \Phi_0 \); the arguments of the proof rely on the Opial lemma (see Opial [14]). This result can be generalized when the sequence \((\varepsilon_n)\) tends to zero fast enough. The key condition is \( \sum_{n=0}^{+\infty} \varepsilon_n < +\infty \) and any sequence \((\varepsilon_n)\) satisfying such a criterion will be referred to as a fast control (or sequence). Let us now state:

**Proposition 2.3.** Under the hypotheses of Proposition 2.1, assume moreover that \( \text{argmin} \Phi_0 \neq \emptyset \) and \( \sum_{n=0}^{+\infty} \varepsilon_n < +\infty \). Let \((x_n)\) be a sequence defined by the algorithm \((A_1)\). Then \((x_n)\) satisfies the following properties:

(i) \( \sum_{n=0}^{+\infty} (\Phi_0(x_n) - \min \Phi_0) < +\infty \).

(ii) There exists \( x_\infty \in \text{argmin} \Phi_0 \) such that \( w\lim_{n \to +\infty} x_n = x_\infty \).

**Proof.** (i) Given any \( z \in \text{argmin} \Phi_0 \), let us define the sequence \((g_n)\) by

\[
g_n = \frac{1}{2} |x_n - z|^2.
\]

Decomposing \( x_k - z = x_k - x_{k+1} + x_{k+1} - z \), we obtain for every \( k \in \mathbb{N} \)

\[
g_k = \frac{1}{2} |x_k - x_{k+1}|^2 + g_{k+1} + \langle x_k - x_{k+1}, x_{k+1} - z \rangle,
\]

and hence

\[
g_{k+1} = g_k - \frac{1}{2} |x_{k+1} - x_k|^2 + \langle x_{k+1} - x_k, x_{k+1} - z \rangle \\
\leq g_k - \frac{1}{2} \langle (x_{k+1} - x_k)/\lambda_k, x_{k+1} - z \rangle.
\]

Since \( -(x_{k+1} - x_k)/\lambda_k \in \partial_{\varepsilon_k}(\Phi_0 + \varepsilon_k \Phi_1)(x_{k+1}) \), the previous equality gives

\[
g_{k+1} \leq g_k - \frac{1}{2} \langle (\Phi_0(x_{k+1}) + \varepsilon_k \Phi_1(x_{k+1}) - \Phi_0(z) - \varepsilon_k \Phi_1(z) - \eta_k \rangle.
\]

As a consequence, we have

\[
g_{k+1} - g_k + \lambda (\Phi_0(x_{k+1}) - \min \Phi_0) \leq \overline{\lambda} \varepsilon_k (\Phi_1(z) - \inf \Phi_1) + \overline{\lambda} \eta_k.
\]

A summation from \( k = 0 \) to \( n - 1 \) of these inequalities yields

\[
g_n - g_0 + \lambda \sum_{k=0}^{n-1} (\Phi_0(x_{k+1}) - \min \Phi_0) \leq \overline{\lambda} (\Phi_1(z) - \inf \Phi_1) \sum_{k=0}^{n-1} \varepsilon_k + \overline{\lambda} \sum_{k=0}^{n-1} \eta_k.
\]

Since \( \sum_{k=0}^{+\infty} \varepsilon_k < +\infty \), \( \sum_{k=0}^{+\infty} \eta_k < +\infty \) and \( g_n \geq 0 \), we deduce

\[
\sum_{k=0}^{+\infty} (\Phi_0(x_{k+1}) - \min \Phi_0) < +\infty.
\]
(ii) The main ingredient to prove the weak convergence of the iterates $x_n$ toward $\bar{x}$ is the Opial lemma [14].

**Lemma 2.4.** Let $H$ be a Hilbert space and $(x_n)$ a sequence such that there exists a nonempty set $S \subset H$ verifying

(a) For every $z \in S$, $\lim_{n \to +\infty} |x_n - z|$ exists.

(b) If $x_{n_k} \to \bar{x}$ weakly in $H$ for a subsequence $n_k \to +\infty$ then $\bar{x} \in S$.

Then, there exists $x_\infty \in S$ such that $x_n \to x_\infty$ weakly in $H$ as $n \to +\infty$.

Let us apply the Opial lemma with $S = \text{argmin} \Phi_0$. Taking the positive part of (2.2), we obtain

$$(g_{k+1} - g_k)_+ \leq \bar{\lambda} \varepsilon_k \left( \Phi_1(z) - \inf \Phi_1 \right) + \bar{\lambda} \eta_k.$$ 

In view of Lemma 2.2, we conclude that $\lim_{n \to +\infty} g_n$ exists and hence $\lim_{n \to +\infty} |x_n - z|$ exists.

Let us now assume that $x_{n_k} \to \bar{x}$ weakly in $H$ for a subsequence $n_k \to +\infty$. From Proposition 2.1 (iii), we deduce that $\bar{x} \in S = \text{argmin} \Phi_0$. As a conclusion, the Opial lemma applies and provides the existence of $x_\infty \in \text{argmin} \Phi_0$ such that $x_n \to x_\infty$. □

It is interesting to notice the analogy between the discrete case and the continuous one. In the continuous case, it has been proved in [6] that the trajectories of (SDC) weakly converge toward a minimum of $\Phi_0$ under the condition $\int_0^{+\infty} \varepsilon(t) \, dt < +\infty$. This result is exactly the continuous version of Proposition 2.3. Conversely, it is shown in [6] that the assumption $\int_0^{+\infty} \varepsilon(t) \, dt = +\infty$ allows to rescale conveniently the (SDC) system, then giving rise to the minimization of $\Phi_1$ over the set $\text{argmin} \Phi_0$. The same phenomenon occurs in the discrete case: this is the subject of the next section.

3. **Slow control: minimization of $\Phi_1$ over $\text{argmin} \Phi_0$.**

3.1. **Convergence of the distance to the set $\text{argmin}_S \Phi_1$.** When $(\varepsilon_n)$ is a fast control, Proposition 2.3 shows the weak convergence of the algorithm, but the limit does not depend explicitly on $\Phi_1$. The potential $\Phi_0$ plays no crucial role because the sequence $(\varepsilon_n)$ vanishes too fastly, whence the idea of introducing a slow control satisfying $\sum_{n=0}^{+\infty} \varepsilon_n = +\infty$. The next theorem states that, under the condition $\sum_{n=0}^{+\infty} \varepsilon_n = +\infty$ the algorithm $(A_1)$ tends to minimize $\Phi_1$ over the set $\text{argmin} \Phi_0$.

**Theorem 3.1.** In addition to the hypotheses of Proposition 2.1, assume moreover that $H$ is finite dimensional, that $S := \text{argmin} \Phi_0 \neq \emptyset$ and $S_1 := \text{argmin}_S \Phi_1 \neq \emptyset$. The sequence $(\varepsilon_n)$ is assumed to satisfy $\sum_{n=0}^{+\infty} \varepsilon_n = +\infty$. If the sequence $(x_n)$ defined by the algorithm $(A_1)$ is bounded, then

(i) $\lim_{n \to +\infty} d(x_n, S_1) = 0$,

(ii) $\lim_{n \to +\infty} (\Phi_0(x_n), \Phi_1(x_n)) = (\min \Phi_0, \min_S \Phi_1),$

where $d(., S_1)$ stands for the distance function to the set $S_1$.

**Proof.** (i) The proof relies on the study of the sequence $(h_n)$ defined by

$$h_n = \frac{1}{2} d(x_n, S_1)^2.$$ 

Denoting by $P_{S_1}$ the projection operator onto the convex set $S_1$, we have

$$h_n = \frac{1}{2} [x_k - P_{S_1}(x_k)]^2$$

$$= \frac{1}{2} [x_k - P_{S_1}(x_k) - (x_{k+1} - P_{S_1}(x_{k+1}))^2 + \frac{1}{2} [x_{k+1} - P_{S_1}(x_{k+1})]^2$$

$$+ (x_k - P_{S_1}(x_k) - (x_{k+1} - P_{S_1}(x_{k+1})), x_{k+1} - P_{S_1}(x_{k+1}))$$

$$\geq h_{k+1} + \langle x_k - x_{k+1}, x_{k+1} - P_{S_1}(x_{k+1}) \rangle + \langle P_{S_1}(x_{k+1}) - P_{S_1}(x_k), x_{k+1} - P_{S_1}(x_{k+1}) \rangle.$$
Since \( P_{S_1}(x_k) \in S_1 \), we classically have
\[
\langle P_{S_1}(x_{k+1}) - P_{S_1}(x_k), x_{k+1} - P_{S_1}(x_{k+1}) \rangle \geq 0
\]
and finally
\[
h_{k+1} - h_k \leq \langle x_{k+1} - x_k, x_{k+1} - x_k \rangle. \tag{3.1}
\]
From the fact that \(-\langle x_{k+1} - x_k \rangle/\lambda_k \in \partial_{\Phi_k}(\Phi_0 + \varepsilon_k \Phi_1)(x_{k+1})\), we infer
\[
\langle -(x_{k+1} - x_k) / \lambda_k, x_{k+1} - P_{S_1}(x_{k+1}) \rangle \geq \Phi_0(x_{k+1}) - \min \Phi_0 + \varepsilon_k (\Phi_1(x_{k+1}) - \min \Phi_1) - \eta_k.
\]
By combining (3.1) and (3.2), we are led to
\[
h_{k+1} - h_k + \lambda_k (\Phi_0(x_{k+1}) - \min \Phi_0) + \lambda_k \varepsilon_k (\Phi_1(x_{k+1}) - \min \Phi_1) \leq \lambda \eta_k
\]
and therefore
\[
h_{k+1} - h_k + \lambda_k \varepsilon_k (\Phi_1(x_{k+1}) - \min \Phi_1) \leq \lambda \eta_k. \tag{3.3}
\]
We now distinguish the following cases:
(a) \( \forall n \in \mathbb{N}, \quad \exists n_0 \geq n_0, \quad \Phi_1(x_n) > \min_S \Phi_1 \).
(b) \( \forall n_0 \in \mathbb{N}, \quad \exists n \geq n_0, \quad \Phi_1(x_n) \leq \min_S \Phi_1 \).

Case (a). We assume that there exists \( n_0 \in \mathbb{N} \) such that, for every \( n \geq n_0 \), \( \Phi_1(x_n) > \min_S \Phi_1 \). From (3.3), we deduce that, for every \( k \geq n_0 \), \( h_{k+1} - h_k \leq \lambda \eta_k \). Therefore, we have \( h_{k+1} - h_k \leq \lambda \eta_k \), which in view of Lemma 2.2 implies that
\[
\lim_{n \to +\infty} h_n \quad \text{exists.} \tag{3.4}
\]
The end of the proof consists in showing that this limit equals zero. Let us add inequalities (3.3) from \( k = n_0 \) to \( n - 1 \); we find
\[
h_n - h_{n_0} + \lambda \sum_{k=n_0}^{n-1} \varepsilon_k (\Phi_1(x_{k+1}) - \min \Phi_1) \leq \lambda \sum_{k=n_0}^{n-1} \eta_k.
\]
Since \( h_n \geq 0 \) and \( \sum_{k=n_0}^{+\infty} \eta_k < +\infty \), we deduce that
\[
\sum_{k=n_0}^{+\infty} \varepsilon_k (\Phi_1(x_{k+1}) - \min \Phi_1) < +\infty. \tag{3.5}
\]
From (3.5), it is immediate that \( \liminf_{n \to +\infty} \Phi_1(x_n) = \min_S \Phi_1 \) (indeed, the assumption \( \liminf_{n \to +\infty} \Phi_1(x_n) > \min_S \Phi_1 \) would lead to a contradiction with the fact that \( \sum_{k=0}^{+\infty} \varepsilon_k = +\infty \)). Consider a subsequence \( (x_n) \), still denoted \( (x_n) \) such that \( \lim_{n \to +\infty} \Phi_1(x_n) = \min_S \Phi_1 \). Since the sequence \( (x_n) \) is bounded, we can extract a converging subsequence \( (x_{n_k}) \) of \( (x_n) \): there exists \( \overline{x} \in H \) such that \( \lim_{k \to +\infty} x_{n_k} = \overline{x} \). In view of Proposition 2.1 (iii), we have \( \overline{x} \in \arg\min \Phi_0 = S \). The continuity of the finite convex function \( \Phi_1 \) implies that
\[
\Phi_1(\overline{x}) = \lim_{k \to +\infty} \Phi_1(x_{n_k}) = \lim_{n \to +\infty} \Phi_1(x_n) = \min_S \Phi_1,
\]
and hence \( \overline{x} \in \arg\min_S \Phi_1 = S_1 \). On the other hand, we have
\[
\lim_{k \to +\infty} h_{n_k} = \lim_{k \to +\infty} \frac{1}{2} d(x_{n_k}, S_1)^2 = \frac{1}{2} d(\overline{x}, S_1)^2 = 0.
\]
which means that 0 is a limit point of \((h_n)\). Since the sequence \((h_n)\) is convergent, we conclude that \(\lim_{n \to +\infty} h_n = 0\).

**Case (b).** We assume that for every \(n_0 \in \mathbb{N}\) there exists \(n \geq n_0\) such that \(\Phi_1(x_n) \leq \min_S \Phi_1\). Let us consider the sequence \((\tau_n)\) defined by:

\[
\tau_n = \max\{k \in \mathbb{N}, \quad k \leq n \quad \text{and} \quad \Phi_1(x_k) \leq \min_S \Phi_1\}.
\]

From the above assumption, it is immediate that \((\tau_n)\) is defined for \(n\) large enough and that \(\lim_{n \to +\infty} \tau_n = +\infty\). Suppose now that \(\tau_n \leq n - 1\). From the definition of \(\tau_n\) and formula (3.3), we have

\[
\forall k \in \{\tau_n, n - 1\}, \quad h_{k+1} - h_k \leq \lambda \eta_k.
\]

By adding these \((n - \tau_n)\) inequalities, we obtain

\[
h_n - h_{\tau_n} \leq \lambda \sum_{k=\tau_n}^{n-1} \eta_k
\]

and therefore

\[
h_n \leq h_{\tau_n} + \lambda \sum_{k=\tau_n}^{+\infty} \eta_k.
\]  \(\text{(3.6)}\)

Notice that, if \(\tau_n = n\), then we have \(h_{\tau_n} = h_n\) so that relation (3.6) is always true. If we are able to prove that \(\lim_{n \to +\infty} h_{\tau_n} = 0\), then inequality (3.6) combined with the fact that \(\lim_{n \to +\infty} \sum_{k=\tau_n}^{+\infty} \eta_k = 0\), will immediately imply that \(\lim_{n \to +\infty} h_n = 0\).

Let us now show that \(\lim_{n \to +\infty} h_{\tau_n} = 0\). Consider a convergent subsequence of the bounded sequence \((x_{\tau_n})\), still denoted by \((x_{\tau_n})\); there exists \(\bar{x} \in H\) such that \(\lim_{n \to +\infty} x_{\tau_n} = \bar{x}\). The set \([\Phi_1 \leq \min_S \Phi_1]\) is closed as a sublevel set of the continuous function \(\Phi_1\). From the definition of \(\tau_n\), we have \(x_{\tau_n} \in [\Phi_1 \leq \min_S \Phi_1]\) for every \(n \in \mathbb{N}\) thus implying \(\bar{x} \in [\Phi_1 \leq \min_S \Phi_1]\). On the other hand, from Proposition 2.1 (iii), we have \(\bar{x} \in S = \operatorname{argmin} \Phi_0\). We immediately deduce that \(\bar{x} \in S_1 = \operatorname{argmin} \Phi_1\), which means that every limit point of the sequence \((x_{\tau_n})\) belongs to \(S_1\). It is then clear that 0 is the unique limit point of the bounded sequence \((d(x_{\tau_n}, S_1))\). We easily conclude that \(\lim_{n \to +\infty} h_{\tau_n} = \lim_{n \to +\infty} \frac{1}{\lambda} d(x_{\tau_n}, S_1)^2 = 0\).

\(\text{(ii)}\) The fact that \(\lim_{n \to +\infty} \Phi_0(x_n) = \min \Phi_0\) is a direct consequence of Proposition 2.1 (iii). Let us now prove that \(\lim_{n \to +\infty} \Phi_1(x_n) = \min_S \Phi_1\). The sequence \((\Phi_1(x_n))\) is the image of the bounded sequence \((x_n)\) by the continuous function \(\Phi_1\), hence \((\Phi_1(x_n))\) is bounded. Let \((\Phi_1(x_{n_k}))\) be a converging subsequence of \((\Phi_1(x_n))\). Since \((x_{n_k})\) is bounded, there is a subsequence of \((x_{n_k})\), still denoted by \((x_{n_k})\) which converges to \(\bar{x} \in H\). From (i), we have \(\lim_{k \to +\infty} d(x_{n_k}, S_1) = 0\) and hence \(\bar{x} \in S_1\). From the continuity of \(\Phi_1\), we deduce that \(\lim_{k \to +\infty} \Phi_1(x_{n_k}) = \min_S \Phi_1\). Since \(\min_S \Phi_1\) is the limit of every converging subsequence of \((\Phi_1(x_n))\), we conclude that \(\lim_{n \to +\infty} \Phi_1(x_n) = \min_S \Phi_1\). \(\square\)

**Remark 3.1.** Take \(\Phi_1 = \|\cdot\|^2/2\) in Theorem 3.1: at each iteration \(n\), the algorithm \((A_1)\) consists in a proximal point method applied to the \(\varepsilon_n\)-Tikhonov regularization of \(\Phi_0\) (see [15] for details on Tikhonov regularization). In such conditions,
Theorem 3.1 shows that the sequence \((x_n)\) generated by \((A_1)\) converges toward the element of minimal norm of \(S = \text{argmin} \Phi_0\). This result presents interesting similarities with the corresponding results of Attouch-Cominetti [3] and Attouch-Czarnecki [4] in the continuous case.

In Theorem 3.1, the sequence \((x_n)\) is supposed to be a priori bounded. We now give a sufficient condition implying the boundedness of \((x_n)\).

**Proposition 3.2.** Under the assumptions of Theorem 3.1, let \((x_n)\) be a sequence generated by the algorithm \((A_1)\). If the following condition holds: for every \(M > 0\),

\[(C) \quad \text{The set} \{x \in H, \quad \Phi_0(x) \leq M\} \cap \{x \in H, \quad \Phi_1(x) \leq \min_S \Phi_1\} \quad \text{is bounded},\]

then the sequence \((x_n)\) is bounded and hence, in view of Theorem 3.1, \(\lim_{n \to +\infty} d(x_n, S_1) = 0\).

Moreover, condition \((C)\) is satisfied in each of the following cases

(a) \(\Phi_0\) is coercive, i.e. \(\lim_{|x| \to +\infty} \Phi_0(x) = +\infty\).

(b) \(\Phi_1\) is coercive, i.e. \(\lim_{|x| \to +\infty} \Phi_1(x) = +\infty\).

**Proof.** We use the same notations as in the proof of Theorem 3.1. From Proposition 2.1(ii), there exists \(M_0 \in \mathbb{R}\) such that, for every \(n \in \mathbb{N}\), \(\Phi_0(x_n) \leq M_0\), i.e.

\[\{x_n, n \in \mathbb{N}\} \subset [\Phi_0 \leq M_0]. \tag{3.7}\]

The set \([\Phi_0 \leq M_0] \cap [\Phi_1 \leq \min_S \Phi_1]\) is bounded in view of condition \((C)\) and we denote by \(\rho\) its radius. The set \(S_1\) is clearly bounded as a subset of \([\Phi_0 \leq M_0] \cap [\Phi_1 \leq \min_S \Phi_1]\). We now distinguish the same cases \((a)\) and \((b)\) as in the proof of Theorem 3.1.

**Case (a).** From (3.4), \(\lim_{n \to +\infty} d(x_n, S_1)\) exists, which implies that the sequence \((d(x_n, S_1))\) is bounded and since \(S_1\) is bounded, we conclude that the sequence \((x_n)\) is bounded too.

**Case (b).** Recall that, from inequality (3.6), we have \(h_n \leq h_{n_k} + \lambda \sum_{k=0}^{+\infty} \eta_k\). From the definition of \(\tau_n\), the point \(x_{\tau_n}\) belongs to \([\Phi_1 \leq \min_S \Phi_1]\) and in view of (3.7), we also have \(x_{\tau_n} \in [\Phi_0 \leq M_0]\), so that

\[x_{\tau_n} \in [\Phi_0 \leq M_0] \cap [\Phi_1 \leq \min_S \Phi_1].\]

This, combined with the fact that \(S_1 \subset [\Phi_0 \leq M_0] \cap [\Phi_1 \leq \min_S \Phi_1]\) implies that \(d(x_{\tau_n}, S_1) \leq \rho\), i.e. \(h_n \leq \rho^2/2\). Finally, by taking into account inequality (3.6), we conclude that, for every \(n \in \mathbb{N}\),

\[h_n \leq \rho^2/2 + \lambda \sum_{k=0}^{+\infty} \eta_k,\]

which proves that the sequence \((x_n)\) is bounded. \(\Box\)

### 3.2. Convergence of the sequence \((x_n)\).

From Theorem 3.1, the distance of the sequence \((x_n)\) (generated by \((A_1)\)) to the set \(S_1\) tends to 0. In particular, if \(S_1\) is reduced to a singleton, then the sequence \((x_n)\) converges. However, in the slow case we do not have any general result of convergence for \((x_n)\) like Proposition 2.3 in the fast case. The convergence of the sequence \((x_n)\) can be obtained by strengthening the assumptions on \(\Phi_0, \Phi_1\) and \((\varepsilon_n)\) as shows the following proposition.

**Proposition 3.3.** In addition to the hypotheses of Theorem 3.1, assume moreover that there exist \(\alpha \geq 0\) and \(p > 1\) such that, for \(\varepsilon\) small enough

\[(D_1) \quad \Phi_0 - \min_S \Phi_0 + \varepsilon (\Phi_1 - \min_S \Phi_1) \geq -\alpha \varepsilon^p.\]
The sequence \((\varepsilon_n)\) is supposed to satisfy
\[
\sum_{n=0}^{+\infty} \varepsilon_n = +\infty \quad \text{and} \quad \alpha \sum_{n=0}^{+\infty} \varepsilon^p_n < +\infty,
\]
with the convention \(0 \times \infty = 0\). Then, the sequence \((x_n)\) generated by the algorithm \((A_1)\) converges to some \(\bar{x} \in S_1\).

**Proof.** Given any \(z \in S_1\), let us define the sequence \((g_n)\) by
\[
g_n = \frac{1}{2} |x_n - z|^2.
\]
The same computation as in the proof of Proposition 2.3 shows that
\[
g_{k+1} - g_k + \lambda_k \left( \Phi_0(x_{k+1}) - \min S \Phi_0 + \varepsilon_k (\Phi_1(x_{k+1}) - \min S \Phi_1) \right) \leq \lambda \eta_k.
\]
This last inequality combined with condition \((D_1)\) implies
\[
g_{k+1} - g_k \leq \lambda (\eta_k + \alpha \varepsilon^p_k).
\]
Taking the positive part of the previous inequality and using the fact that \(\sum_{k=0}^{+\infty} \eta_k < +\infty\) and \(\alpha \sum_{k=0}^{+\infty} \varepsilon^p_k < +\infty\), we obtain in view of Lemma 2.2 that \(\lim_{n \to +\infty} g_n\) exists and hence
\[
\lim_{n \to +\infty} |x_n - z| \quad \text{exists for any } z \in S_1. \quad (3.8)
\]
Hence, in particular the sequence \((x_n)\) is bounded and we can extract a converging subsequence: there exist \(\bar{x} \in H\) and \((x_{n_k})\) such that \(\lim_{k \to +\infty} x_{n_k} = \bar{x}\). Since \((x_n)\) is bounded, Theorem 3.1 applies and we have \(\bar{x} \in S_1\). Taking \(z = \bar{x}\) in \((3.8)\), we deduce that \(\lim_{n \to +\infty} |x_n - \bar{x}|\) exists, which combined with \(\lim_{k \to +\infty} |x_{n_k} - \bar{x}| = 0\), finally yields \(\lim_{n \to +\infty} |x_n - \bar{x}| = 0\). \(\square\)

**Remark 3.2.** When criterion \((D_1)\) is realized with \(\alpha = 0\), the required condition on \((\varepsilon_n)\) reduces to \(\sum_{n=0}^{+\infty} \varepsilon_n = +\infty\).

**Remark 3.3.** Let us now assume that \(\alpha > 0\) and denote by \(l^r\) \((r \geq 1)\) the set of non negative sequences \((a_n)\) satisfying \(\sum_{n=0}^{+\infty} a^r_n < +\infty\). The assumptions of Proposition 3.3 relative to \((\varepsilon_n)\) can be rewritten as \((\varepsilon_n) \in l^r \setminus l^1\). This means that \((\varepsilon_n)\) is a slow control which is however sufficiently fast so as to ensure \((\varepsilon_n) \in l^r\). The question of the convergence is open in the case of a “very slow” control. Let us now study the meaning of condition \((D_1)\).

**Proposition 3.4.** Let \(H\) be a finite dimensional Hilbert space, \(\Phi_0 : H \to \mathbb{R} \cup \{+\infty\}\) a closed convex function and \(\Phi_1 : H \to \mathbb{R}\) a finite convex function. Assume that \(\Phi_0\) and \(\Phi_1\) are bounded from below and that the sets \(S := \arg\min \Phi_0\) and \(S_1 := \arg\min_S \Phi_1\) are non-empty. Then, condition \((D_1)\) is satisfied in each of the following cases:

(a) \(\arg\min \Phi_0 \cap \arg\min \Phi_1 \neq \emptyset\).

(b) There exist \(a > 0\), \(b > 0\) and \(r \geq 1\) such that

(i) \(\Phi_0 - \min \Phi_0 \geq a d(S)^r\),

(ii) \(\Phi_1 - \min_S \Phi_1 \geq -bd(S \Phi_1 \geq \min_S \Phi_1)\).
Proof. (a) Notice that the assumption \( \arg \min \Phi_0 \cap \arg \min \Phi_1 \neq \emptyset \) implies that \( \arg \min_S \Phi_1 = \arg \min \Phi_0 \cap \arg \min \Phi_1 \) and \( \min_S \Phi_1 = \min \Phi_1 \). As a consequence,

\[
\Phi_0 - \min \Phi_0 + \varepsilon \left( \Phi_1 - \min \Phi_1 \right) = \Phi_0 - \min \Phi_0 + \varepsilon \left( \Phi_1 - \min \Phi_1 \right) \geq 0
\]

and condition \((D_1)\) is satisfied with \( \alpha = 0 \).

(b) Since the set \( S \) is included in the set \([\Phi_1 \geq \min_S \Phi_1]\), we have \( d(.) , S) \geq d(\cdot, [\Phi_1 \geq \min_S \Phi_1]) \) which combined with the assumption on \( \Phi_1 \), implies

\[
\Phi_1 - \min_S \Phi_1 \geq -\beta d(\cdot, S).
\]

Taking into account assumption (i), we deduce from the previous inequality that, for every \( x \in H \),

\[
\Phi_0(x) - \min \Phi_0 + \varepsilon \left( \Phi_1(x) - \min_S \Phi_1 \right) \geq a d(x, S)^r - \beta d(x, S).
\]

First assume that \( r = 1 \). It is then immediate that we have, for \( \varepsilon \) small enough

\[
\Phi_0(x) - \min \Phi_0 + \varepsilon \left( \Phi_1(x) - \min_S \Phi_1 \right) \geq 0,
\]

and condition \((D_1)\) is satisfied with \( \alpha = 0 \). Now assume that \( r > 1 \). An elementary computation then shows that

\[
ad(x, S)^r - \beta d(x, S) \geq -b \left( \frac{\beta}{ra} \right)^{\frac{r-1}{r}} \left( \frac{r-1}{r} \right) \varepsilon^{\frac{r}{r-1}},
\]

so that condition \((D_1)\) is satisfied with \( \alpha = b \left( \frac{\beta}{ra} \right)^{\frac{r-1}{r}} \left( \frac{r-1}{r} \right) \) and \( p = \frac{r}{r-1} \). \( \square \)

Remark 3.4. The Lipschitz-type condition (ii) on \( \Phi_1 \) is verified in many situations. Indeed, since \( H \) is finite dimensional, the finite convex function \( \Phi_1 \) is continuous on \( H \) and even Lipschitz continuous on the bounded subsets of \( H \). Assuming for example that the set \([\Phi_1 \leq \min_S \Phi_1]\) is bounded, one can easily prove that condition (ii) is satisfied.


4.1. Algorithm \((A_2)\): minimization of \( \Phi_2 \) over \( \arg \min_S \Phi_1 \). Given three convex functions \( \Phi_0, \Phi_1 \) and \( \Phi_2 \), it is natural in view of Theorem 3.1 to try to minimize \( \Phi_2 \) over the set \( \arg \min_S \Phi_1 \). For that purpose, we define the algorithm \((A_2)\):

\[
(A_2) \quad - \frac{x_{n+1} - x_n}{\lambda_n} \in \partial q_n (\Phi_0 + \varepsilon_n \Phi_1 + \varepsilon_n^p \Phi_2) (x_{n+1}),
\]

where the choice of the exponent \( q \) has to be precised. Denoting by \( p \) the exponent arising in condition \((D_1)\), the next theorem indicates to take \( q \) in the interval \([1,p]\) and a sequence \((\varepsilon_n)\) satisfying \((\varepsilon_n) \in L^p \setminus L^p \).

Theorem 4.1. Let \( H \) be a finite dimensional Hilbert space, \( \Phi_0 : H \to \mathbb{R} \cup \{+\infty\} \) a closed convex function and \( \Phi_1, \Phi_2 : H \to \mathbb{R} \) finite convex functions. Assume that the functions \( \Phi_0, \Phi_1, \Phi_2 \) are bounded from below and that the sets \( S := \arg \min \Phi_0, S_1 := \arg \min_S \Phi_1, S_2 := \arg \min_S \Phi_2 \) are non-empty. Suppose moreover that there exist \( \alpha \geq 0 \) and \( p > 1 \) such that, for \( \varepsilon \) small enough

\[
(D_1) \quad \Phi_0 - \min \Phi_0 + \varepsilon (\Phi_1 - \min_S \Phi_1) \geq -\alpha \varepsilon^p.
\]
Given some \( q \in [1,p], \) we consider a sequence \( (\varepsilon_n) \) satisfying \((\mathcal{H}_q)\) and

\[
\sum_{n=0}^{+\infty} \varepsilon_n^q = +\infty \quad \text{and} \quad \alpha \sum_{n=0}^{+\infty} \varepsilon_n^p < +\infty,
\]

with the convention \( 0 \times \infty = 0. \) We are also given non-negative sequences \((\lambda_n), (\eta_n)\) verifying respectively \((\mathcal{H}_\lambda)\) and \((\mathcal{H}_\eta)\). We then define a sequence \((x_n)\) generated by the algorithm \((A_2)\) and satisfying the initial condition \( x_0 \in \text{dom} \Phi_0 \). If \((x_n)\) is bounded, then:

(i) \( \lim_{n \to +\infty} d(x_n, S_2) = 0 \),

(ii) \( \lim_{n \to +\infty} (\Phi_0(x_n), \Phi_1(x_n), \Phi_2(x_n)) = (\min \Phi_0, \min_S \Phi_1, \min_{S_1} \Phi_2) \).

Proof. (i) We decompose the proof into three steps: first we prove that \( \lim_{n \to +\infty} d(x_n, S) = 0 \), then that \( \lim_{n \to +\infty} d(x_n, S_1) = 0 \) and finally that \( \lim_{n \to +\infty} d(x_n, S_2) = 0 \).

Step 1. The results of Proposition 2.1 for \((A_1)\) can be immediately extended to \((A_2)\) by replacing the discrete energy \( E_n = \Phi_0(x_n) + \varepsilon_n \Phi_1(x_n) \) by \( E_n^{(2)} = \Phi_0(x_n) + \varepsilon_n \Phi_1(x_n) + \varepsilon_n^q \Phi_2(x_n) \) in the proof of Proposition 2.1. It is then easy to establish that \( \lim_{n \to +\infty} \Phi_0(x_n) = \min \Phi_0 \) (cf. Proposition 2.1 (iii)). Since \( H \) is finite dimensional and \((x_n)\) is bounded, we immediately deduce that \( \lim_{n \to +\infty} d(x_n, S) = 0 \).

Step 2. Setting \( h_n = \frac{1}{2} d(x_n, S_1)^2 \), the same computation as in the proof of Theorem 3.1 (i) shows that

\[
h_{k+1} - h_k + \lambda_k \varepsilon_k (\Phi_1(x_{k+1}) - \min_S \Phi_2 - M \varepsilon_k^{q-1}) \leq \lambda \eta_k.
\]

Since the sequence \((x_n)\) is bounded, the sequences \((\Phi_2(x_{n+1}))\) and \((\Phi_2(P_{S_1}(x_{n+1})))\) are also bounded, so that there exists \( M \in \mathbb{R}_+ \) such that, for every \( k \geq 0 \),

\[
\Phi_2(x_{k+1}) - \Phi_2(P_{S_1}(x_{k+1})) \geq - M.
\]

This inequality combined with the previous one yields

\[
h_{k+1} - h_k + \lambda_k \varepsilon_k (\Phi_1(x_{k+1}) - \min_S \Phi_1 - M \varepsilon_k^{q-1}) \leq \lambda \eta_k.
\]

The rest of the proof consists in distinguishing the following cases:

(\( a_1 \)) \( \exists n_0 \in \mathbb{N} \quad \forall n \geq n_0, \quad \Phi_1(x_{n+1}) \geq \min_S \Phi_1 + M \varepsilon_k^{q-1}. \)

(\( b_1 \)) \( \forall n_0 \in \mathbb{N} \quad \exists n \geq n_0, \quad \Phi_1(x_{n+1}) \leq \min_S \Phi_1 + M \varepsilon_k^{q-1}. \)

The arguments of the proof of Theorem 3.1 (i) still apply here in so far as the quantity \( \varepsilon_k^{q-1} \) tends to 0 when \( n \to +\infty \). As a consequence, we obtain that \( \lim_{n \to +\infty} d(x_n, S_1) = 0 \). The details are left to the reader.

Step 3. Let us define the sequence \((h_n^{(2)})\) by

\[
h_n^{(2)} = \frac{1}{2} d(x_n, S_2)^2.
\]

The same computation as in the proof of Theorem 3.1 (i) leads to

\[
h_{k+1}^{(2)} - h_k^{(2)} + \lambda_k (\Phi_0(x_{k+1}) - \min \Phi_0) + \lambda_k \varepsilon_k (\Phi_1(x_{k+1}) - \min_S \Phi_1) + \lambda_k \varepsilon_k^q (\Phi_2(x_{k+1}) - \min_{S_1} \Phi_2) \leq \lambda \eta_k.
\]

This last inequality combined with condition \((D_1)\) implies

\[
h_{k+1}^{(2)} - h_k^{(2)} + \lambda_k \varepsilon_k^q (\Phi_2(x_{k+1}) - \min_{S_1} \Phi_2) \leq \lambda (\eta_k + \alpha \varepsilon_k^p).
\]
We are led to distinguish the following cases:

(a) \( \exists n_0 \in \mathbb{N}, \quad \forall n \geq n_0, \quad \Phi_2(x_n) > \min_{S_1} \Phi_2. \)

(b) \( \forall n_0 \in \mathbb{N}, \quad \exists n \geq n_0, \quad \Phi_2(x_n) \leq \min_{S_1} \Phi_2. \)

It suffices now to reproduce the arguments of Theorem 3.1 (i) by taking into account the fact that \( \alpha \sum_{k=0}^{\infty} \varepsilon_k^q < +\infty \) and \( \lim_{n \to +\infty} d(x_n, S_1) = 0 \) (see Step 2). We let the reader check that we have \( \lim_{n \to +\infty} d(x_n, S_2) = 0 \) in both cases (a2) and (b2), which concludes the proof of (i). \( \square \)

(ii) is an immediate consequence of (i). For further details, we refer the reader to the proof of Theorem 3.1 (ii). \( \square \)

REMARK 4.1. When criterion \((D_1)\) is realized with \( \alpha = 0 \), the exponent \( q \) can take any value in the interval \([1, +\infty[\) and the condition on \((\varepsilon_n)\) reduces to \((\varepsilon_n) \not\in I^p\).

Like for the algorithm \((A_1)\), the convergence of the sequence \((x_n)\) generated by \((A_2)\) can be obtained by strengthening the assumptions on \( \Phi_0, \Phi_1 \) and \( \Phi_2 \). In the following proposition, we introduce the adequate condition \((D_2)\) which is more stringent than \((D_1)\).

PROPOSITION 4.2. In addition to the hypotheses of Theorem 4.1, assume moreover that there exist \( \alpha > 0, p > 1 \) and \( q \in ]1, p[\) such that, for \( \varepsilon \) small enough

\[
\begin{align*}
(D_2) & \quad \begin{cases}
\Phi_0 - \min \Phi_0 + \varepsilon (\Phi_1 - \min_S \Phi_1) \geq -\alpha \varepsilon^p \\
\Phi_0 - \min \Phi_0 + \varepsilon (\Phi_1 - \min_S \Phi_1) + \varepsilon^q (\Phi_2 - \min_S \Phi_2) \geq -\alpha \varepsilon^p.
\end{cases}
\end{align*}
\]

Then, the sequence \((x_n)\) generated by the algorithm \((A_2)\) converges to some \( \bar{x} \in S_2. \)

Proof. Given any \( z \in S_2, \) let us define the sequence \((g_n)\) by \( g_n = \frac{1}{2} |x_n - z|^2. \) An elementary computation shows that

\[
g_{k+1} - g_k + \lambda_k \left( \Phi_0(x_{k+1}) - \min \Phi_0 + \varepsilon_k (\Phi_1(x_{k+1}) - \min_S \Phi_1) + \varepsilon_k^q (\Phi_2(x_{k+1}) - \min_S \Phi_2) \right) \leq \alpha k. \eta_k.
\]

This last inequality combined with the second inequality of condition \((D_2)\) implies

\[
g_{k+1} - g_k \leq \alpha k + \alpha \varepsilon_k^p.
\]

The sequel of the proof is similar to the one of Proposition 3.3 and the reader is referred to it. \( \square \)

Let us now exhibit sufficient conditions under which assumption \((D_2)\) is satisfied.

PROPOSITION 4.3. Let \( H \) be a finite dimensional Hilbert space, \( \Phi_0 : H \to \mathbb{R} \cup \{+\infty\} \) a closed convex function and \( \Phi_1, \Phi_2 : H \to \mathbb{R} \) finite convex functions. Assume that the functions \( \Phi_0, \Phi_1, \Phi_2 \) are bounded from below and that the sets \( S := \text{argmin} \Phi_0, S_1 := \text{argmin} \Phi_1, S_2 := \text{argmin} \Phi_2 \) are non-empty. Then, condition \((D_2)\) is satisfied in each of the following cases:

(a) \( \text{argmin} \Phi_0 \cap \text{argmin} \Phi_1 \cap \text{argmin} \Phi_2 \neq \emptyset. \)

(b) There exist \( a_0, a_1, b_1, b_2, c \in \mathbb{R}^*_{\geq} \) and \( n_0, r_1 \in ]1, +\infty[ \) such that

\[
\begin{align*}
&\begin{cases}
(i) \quad \Phi_0 - \min \Phi_0 \geq a_0 d(., S)^{r_0}, \\
(ii) \quad \Phi_1 - \min_S \Phi_1 \geq a_1 d(., [\Phi_1 \leq \min_S \Phi_1])^{r_1} - b_1 d(., [\Phi_1 \geq \min_S \Phi_1]), \\
(iii) \quad \Phi_2 - \min_S \Phi_2 \geq -b_2 d(., [\Phi_2 \geq \min_S \Phi_2]), \\
(iv) \quad d(., S_1) \leq c (d(., [\Phi_1 \leq \min_S \Phi_1]) + d(., S)).
\end{cases}
\end{align*}
\]
Proof. (a) Notice that the assumption \( \arg \min \Phi_0 \cap \arg \min \Phi_1 \cap \arg \min \Phi_2 \neq \emptyset \) implies that \( \min_S \Phi_1 = \min \Phi_1 \) and \( \min_{S_1} \Phi_2 = \min \Phi_2 \). As a consequence,

\[
\Phi_0 - \min \Phi_0 + \varepsilon (\Phi_1 - \min \Phi_1) + \varepsilon^q (\Phi_2 - \min \Phi_2) \geq 0
\]

and condition \((D_2)\) is satisfied with \( \alpha = 0 \).

(b) Let us first minorize the quantity

\[
A_\varepsilon(x) := \Phi_0(x) - \min \Phi_0 + \varepsilon (\Phi_1(x) - \min \Phi_1).
\]

Since \( S \subset [\Phi_1 \geq \min_S \Phi_1] \), we have \( d(\cdot, S) \geq d(\cdot, [\Phi_1 \geq \min_S \Phi_1]) \) which combined with (ii) implies

\[
\Phi_1(x) - \min \Phi_1 \geq a_1 d(x, [\Phi_1 \leq \min \Phi_1])^{r_1} - b_1 d(x, S). \tag{4.2}
\]

In view of (i), we deduce that \( A_\varepsilon(x) \geq a_0 d(x, S)^{r_0} - b_1 \varepsilon d(x, S) \). The same computation as in the proof of Proposition 3.4 (b) shows that there exist \( m_1 \geq 0 \) and \( p_0 > 1 \) such that, for \( \varepsilon \) small enough,

\[
a_0 d(x, S)^{r_0} - b_1 \varepsilon d(x, S) \geq -m_1 \varepsilon^{p_0}.
\]

More precisely, if \( r_0 > 1 \) one can take \( p_0 = r_0 / (r_0 - 1) \) and if \( r_0 = 1 \), the previous inequality is satisfied with \( m_1 = 0 \). This leads to the following minorization of \( A_\varepsilon \):

\[
A_\varepsilon(x) \geq -m_1 \varepsilon^{p_0}. \tag{4.3}
\]

Let us now find a lower bound for the quantity

\[
B_\varepsilon(x) := \Phi_0(x) - \min \Phi_0 + \varepsilon (\Phi_1(x) - \min \Phi_1) + \varepsilon^q (\Phi_2(x) - \min \Phi_2).
\]

The inclusion \( S_1 \subset [\Phi_2 \geq \min_{S_1} \Phi_2] \) implies \( d(\cdot, S_1) \geq d(\cdot, [\Phi_2 \geq \min_{S_1} \Phi_2]) \), and hence in view of (iii) and (iv),

\[
\Phi_2(x) - \min_{S_1} \Phi_2 \geq -b_2 c (d(x, [\Phi_1 \leq \min_S \Phi_1]) + d(x, S)) \tag{4.4}
\]

Combining (i), (4.2) and (4.4), we obtain

\[
B_\varepsilon(x) \geq a_0 d(x, S)^{r_0} - b_1 \varepsilon d(x, S) - b_2 c \varepsilon^q d(x, S)
\]

\[
+ a_1 \varepsilon d(x, [\Phi_1 \leq \min_S \Phi_1])^{r_1} - b_2 c \varepsilon^q d(x, [\Phi_1 \leq \min_S \Phi_1]). \tag{4.5}
\]

Since the quantity \( \varepsilon^q \) is negligible with respect to \( \varepsilon \) when \( \varepsilon \to 0 \), the same arguments as above show that there exists \( m_2 \geq 0 \) such that

\[
a_0 d(x, S)^{r_0} - b_1 \varepsilon d(x, S) - b_2 c \varepsilon^q d(x, S) \geq -m_2 \varepsilon^{p_0}. \tag{4.6}
\]

(If \( r_0 = 1 \) then inequality (4.6) is satisfied with \( m_2 = 0 \).) In the same way, we let the reader check that there exist \( m_3 \geq 0 \) and \( p_1 > 1 \) such that

\[
a_1 \varepsilon d(x, [\Phi_1 \leq \min_S \Phi_1])^{r_1} - b_2 c \varepsilon^q d(x, [\Phi_1 \leq \min_S \Phi_1]) \geq -m_3 \varepsilon^{p_1}. \tag{4.7}
\]
(If \( r_1 > 1 \) then one can take \( p_1 = (q - 1) \left( \frac{r_1}{r_1 - 1} \right) + 1 \) and if \( r_1 = 1 \) then inequality (4.7) is satisfied with \( m_3 = 0 \). In view of (4.5), (4.6) and (4.7), we deduce that
\[
B_\varepsilon(x) \geq -m_2 \varepsilon^{r_2} - m_3 \varepsilon^{r_3}.
\] (4.8)

Let us now conclude by distinguishing the following cases:

- If \( r_0 > 1 \) and \( r_1 > 1 \), choose the value \( q \) so that
\[
p_1 = (q - 1) \left( \frac{r_1}{r_1 - 1} \right) + 1 \geq p_0,
\]
or equivalently, \( q \in [(p_0 - 1) \left( \frac{r_1}{r_1 - 1} \right) + 1, p_0] \). In view of (4.3) and (4.8), condition (\( D_2 \)) is then clearly realized by taking \( p := p_0 \).

- If \( r_0 > 1 \) and \( r_1 = 1 \), condition (\( D_2 \)) holds as soon as \( q \in [1, p_0] \) and \( p := p_0 \).

- If \( r_0 = 1 \) and \( r_1 > 1 \), inequalities (4.3) and (4.8) applied with \( m_1 = m_2 = 0 \) show that condition (\( D_2 \)) is satisfied for any \( q > 1 \) and \( p \) given by \( p := p_1 > q \).

- If \( r_0 = r_1 = 1 \), we have \( m_1 = m_2 = m_3 = 0 \) and condition (\( D_2 \)) trivially holds for \( \alpha = 0 \) and for any couple \((q, p)\) satisfying \( q \in [1, p] \).

Let us now comment items (ii), (iii) and (iv) of Proposition 4.3 (b) by means of the following remarks.

**Remark 4.2.** Assumption (ii) of Proposition 4.3 (b) is a generic condition on \( \Phi_1 \), which is satisfied in many cases. Indeed, first assume that \( \min_S \Phi_1 = \min \Phi_1 \). It is classical to suppose that the convex function \( \Phi_1 \) verifies
\[
\Phi_1 - \min \Phi_1 \geq a_1 d([\Phi_1 \leq \min \Phi_1])^{r_1},
\]
which is exactly (ii) with \( b_1 = 0 \). Conversely assume that \( \min_S \Phi_1 > \min \Phi_1 \) and that \( \Phi_1 \) is differentiable. In such conditions, we have \( |\nabla \Phi_1(x)| \neq 0 \) for every \( x \in \Phi_1 = \min_S \Phi_1 \). Let us denote by \( m \) (resp. \( M \)) the quantity
\[
m := \inf_{x \in \Phi_1 = \min_S \Phi_1} |\nabla \Phi_1(x)| \quad \text{(resp. } M := \sup_{x \in \Phi_1 = \min_S \Phi_1} |\nabla \Phi_1(x)|).\]

Assuming that \( m > 0 \) and \( M < +\infty \), we let the reader check that
\[
\Phi_1 - \min \Phi_1 \geq m d([\Phi_1 \leq \min \Phi_1]) - M d([\Phi_1 \geq \min \Phi_1]),
\]
which is exactly (ii) with \( r_1 = 1 \), \( a_1 = m \) and \( b_1 = M \).

**Remark 4.3.** Assuming that \( \Phi_2 \) is differentiable, let us denote by \( M' \) the quantity
\[
M' := \sup_{x \in [\Phi_2 = \min_S \Phi_2]} |\nabla \Phi_2(x)|.
\]

Supposing moreover that \( M' < +\infty \), it is easy to check that
\[
\Phi_2 - \min \Phi_2 \geq -M' d([\Phi_2 \geq \min \Phi_2]),
\]
which corresponds to assumption (iii) of Proposition 4.3 (b), with \( b_2 = M' \). For other comments relative to assumption (iii), we refer the reader to Remark 3.4.

**Remark 4.4.** Assumption (iv) of Proposition 4.3 (b) is closely connected to the fact that the set \( S_1 \) equals the intersection \( \{ \Phi_1 \leq \min_S \Phi_1 \} \cap S \). We let the reader check that assumption (iv) can be replaced by the following less stringent one: there exist \( r_2, r_3 \in [0, 1] \) such that
\[
(iv - bis) \quad d(., S_1) \leq c \left( d(., [\Phi_1 \leq \min_S \Phi_1])^{r_2} + d(., S)^{r_3} \right).
\]
This last assumption allows a large class of intersecting sets \( S \) and \( [\Phi_1 \leq \min_S \Phi_1] \).
4.2. Algorithm ($A_N$) and hierarchical minimization. Let $\Phi_0 : H \to \mathbb{R} \cup \{+\infty\}$ a closed convex function and $\Phi_1, ..., \Phi_N : H \to \mathbb{R}$ finite convex functions ($N \geq 1$). Assume that $\Phi_0, \Phi_1, ..., \Phi_N$ are bounded from below and define the sets $(S_i)_{i=1}^N$ as follows: $S_{-1} := H$ and for $i \in \{0, ..., N\}$, $S_i := \text{argmin}_{S_{i-1}} \Phi_i$. A challenging task consists in minimizing each function $\Phi_i$ on the set $S_{i-1}$, for $i \in \{0, ..., N\}$. This question of hierarchical minimization has been addressed by many authors (see for example Attouch [2], Cominetti-Courdurier [10]).

According to the previous results of the paper, the algorithm ($A_i$) (resp. ($A_2$)) allows to generate a sequence $(x_n)$ which minimizes the function $\Phi_1$ (resp. $\Phi_2$) over the set $S$ (resp. $S_1$). It is then natural to define the algorithm ($A_N$) by:

$$(A_N) \quad \frac{x_{n+1} - x_n}{\lambda_n} \in \partial_{\varepsilon_n} (\Phi_0 + \varepsilon_1 \Phi_1 + \varepsilon_2^2 \Phi_2 + ... + \varepsilon_n^N \Phi_N)(x_{n+1}),$$

where the exponents $q_2, ..., q_N$ have to be carefully chosen. In the following theorem, the functions $\Phi_0, ..., \Phi_{N-1}$ are assumed to satisfy condition ($D_{N-1}$) which is a generalization at order $N - 1$ of ($D_1$). The choice of the exponents $q_2, ..., q_N$ is then intimately connected with the property ($D_{N-1}$) satisfied by $\Phi_0, ..., \Phi_{N-1}$.

**Theorem 4.4.** Let $H$ be a finite dimensional Hilbert space, $\Phi_0 : H \to \mathbb{R} \cup \{+\infty\}$ a closed convex function and $\Phi_1, ..., \Phi_N : H \to \mathbb{R}$ finite convex functions. Assume that $\Phi_0, \Phi_1, ..., \Phi_N$ are bounded from below and that the sets $(S_i)_{i=1}^N$ defined as above are non-empty. Suppose moreover that there exist $\alpha > 0$, $p > 1$ and $(q_i)_{2 \leq i \leq N-1}$ such that $1 < q_2 < ... < q_{N-1} < p$ and for $\varepsilon$ small enough

$$(D_{N-1}) \quad \forall i \in \{1, ..., N-1\}, \quad \sum_{j=0}^{i} \varepsilon_j^q \left( \Phi_j - \min_{S_{j-1}} \Phi_j \right) \geq -\alpha \varepsilon^p,$$

with the conventions $q_0 = 0$, $q_1 = 1$ and $S_{-1} = H$. Given some $q_N \in [q_{N-1}, p]$, let us consider a sequence $(\varepsilon_n)$ satisfying ($H_0$) and

$$\sum_{n=0}^{+\infty} \varepsilon_n^q = +\infty \quad \text{and} \quad \alpha \sum_{n=0}^{+\infty} \varepsilon_n^p < +\infty,$$

with the convention $0 \times \infty = 0$. We are also given non-negative sequences $(\lambda_n)$, $(\eta_n)$ verifying respectively ($H_0$) and ($H_0$). We then define a sequence $(x_n)$ generated by the algorithm ($A_N$) and satisfying the initial condition $x_0 \in \text{dom} \ \Phi_0$. If $(x_n)$ is bounded, then we have:

(i) $\lim_{n \to +\infty} d(x_n, S_N) = 0$,
(ii) $\forall i \in \{0, ..., N\}$, $\lim_{n \to +\infty} \Phi_i(x_n) = \min_{S_{i-1}} \Phi_i$.

**Proof.** (i) Let us argue by recurrence and let us denote by ($R_i$) the following recurrence assumption:

$$(R_i) \quad \lim_{n \to +\infty} d(x_n, S_i) = 0.$$

Let us first prove that ($R_0$) is satisfied. The results of Proposition 2.1 for ($A_i$) can be immediately extended to ($A_N$) by replacing the discrete energy $E_n = \Phi_0(x_n) + \varepsilon_n \Phi_1(x_n)$ by $E_n^{(N)} = \Phi_0(x_n) + \varepsilon_n \Phi_1(x_n) + \varepsilon_n^2 \Phi_2(x_n) + ... + \varepsilon_n^N \Phi_N(x_n)$ in the proof of Proposition 2.1. It is then easy to establish that $\lim_{n \to +\infty} \Phi_0(x_n) = \min \Phi_0$ (cf. Proposition 2.1 (iii)). Since $H$ is finite dimensional and $(x_n)$ is bounded, we immediately deduce that $\lim_{n \to +\infty} d(x_n, S) = 0$, i.e. ($R_0$) is satisfied.
Let us now prove the implication \((\mathcal{R}_{i-1}) \implies (\mathcal{R}_i)\), for every \(i \in \{1,...,N\}\). For that purpose, we define the sequence \(h_n^{(i)}\) by
\[
h_n^{(i)} = \frac{1}{2} d(x_n, S_i)^2.
\]
We let the reader check that
\[
h_{k+1}^{(i)} - h_k^{(i)} + \lambda_k \sum_{j=0}^{N} \varepsilon_k^{(j)} (\Phi_j(x_{k+1}) - \Phi_j(P_{S_i}(x_{k+1}))) \leq \lambda \eta_k.
\]
Notice that, since \(P_{S_i}(x_{k+1}) \in S_i\), we have
\[
\forall j \in \{0,...,i\}, \quad \Phi_j(P_{S_i}(x_{k+1})) = \min_{S_{j-1}} \Phi_j,
\]
so that the previous inequality can be rewritten as:
\[
h_{k+1}^{(i)} - h_k^{(i)} + \lambda_k \sum_{j=0}^{i} \varepsilon_k^{(j)} (\Phi_j(x_{k+1}) - \min_{S_{j-1}} \Phi_j) + \lambda_k \sum_{j=i+1}^{N} \varepsilon_k^{(j)} (\Phi_j(x_{k+1}) - \Phi_j(P_{S_i}(x_{k+1}))) \leq \lambda \eta_k.
\]
Since the sequence \((x_n)\) is bounded, the sequences \((\Phi_j(x_{n+1}))\) and \((\Phi_j(P_{S_i}(x_{n+1})))\) are also bounded for \(j \in \{i+1,...,N\}\), so that there exists \(M \in \mathbb{R}_+\) such that for \(k\) large enough,
\[
\sum_{j=i+1}^{N} \varepsilon_k^{(j)} (\Phi_j(x_{k+1}) - \Phi_j(P_{S_i}(x_{k+1}))) \geq -M \varepsilon_k^{(i+1)}.
\]
On the other hand, from the \((i-1)\)th inequality of \((D_{N-1})\), we have
\[
\sum_{j=0}^{i-1} \varepsilon_k^{(j)} (\Phi_j - \min_{S_{j-1}} \Phi_j) \geq -\alpha \varepsilon_k.
\]
By combining (4.9), (4.10) and (4.11), we deduce
\[
h_{k+1}^{(i)} - h_k^{(i)} + \lambda_k \varepsilon_k^{(i)} (\Phi_i(x_{k+1}) - \min_{S_{i-1}} \Phi_i - M \varepsilon_k^{(i+1)-\alpha}) \leq \lambda \eta_k + \alpha \varepsilon_k.
\]
The rest of the proof consists in distinguishing the following cases:
(a) \(\forall n \in \mathbb{N}, \quad \Phi_i(x_{n+1}) > \min_{S_{i-1}} \Phi_i + M \varepsilon_n^{(i+1)-\alpha}\).
(b) \(\forall n \in \mathbb{N}, \quad \Phi_i(x_{n+1}) \leq \min_{S_{i-1}} \Phi_i + M \varepsilon_n^{(i+1)-\alpha}\).
The arguments of the proof of Theorem 3.1 (i) still apply here in so far as the quantity \(\varepsilon_k^{(i+1)-\alpha}\) tends to 0 when \(n \to +\infty\). As a consequence, we obtain that \(\lim_{n \to +\infty} h_n^{(i)} = \lim_{n \to +\infty} \frac{1}{2} d(x_n, S_i)^2 = 0\) and \((\mathcal{R}_i)\) is satisfied. The details are left to the reader.

(ii) is an immediate consequence of (i). For further details, we refer the reader to the proof of Theorem 3.1 (ii). \(\square\)

Remark 4.5. When criterion \((D_{N-1})\) is realized with \(\alpha = 0\), the exponent \(q_N\) can take any value in the interval \([q_{N-1}, +\infty]\) and the condition on \(\varepsilon_n\) reduces to \((\varepsilon_n) \not\in l^{\infty}N\).
Like for the algorithms \((A_1)\) and \((A_2)\), the convergence of the sequence \((x_n)\) generated by \((A_N)\) can be obtained by strengthening the assumptions on \(\Phi_0, \ldots, \Phi_N\). In the following proposition, we introduce the adequate condition \((D_N)\) which is more stringent than \((D_{N-1})\).

**Proposition 4.5.** In addition to the hypotheses and conventions of Theorem 4.4, assume that there exist \(\alpha \geq 0\), \(p > 1\) and \((q_i)\) such that \(1 < q_1 < \cdots < q_N < p\) and for \(\varepsilon\) small enough:

\[
(D_N) \quad \forall i \in \{1, \ldots, N\}, \quad \sum_{j=0}^{i} \varepsilon_{ij} (\Phi_j - \min_{S_{j-1}} \Phi_j) \geq -\alpha \varepsilon^p.
\]

Then, the sequence \((x_n)\) generated by the algorithm \((A_N)\) converges to some \(\bar{x} \in S_N\).

**Proof.** Given any \(z \in S_N\), let us define the sequence \((g_n)\) by \(g_n = \frac{1}{2} |x_n - z|^2\). An elementary computation shows that

\[
g_{k+1} - g_k + \lambda_k \sum_{j=0}^{N} \varepsilon_{kj}^2 (\Phi_j - \min_{S_{j-1}} \Phi_j) \leq \lambda \eta_k.
\]

This last inequality combined with the \(N^{th}\) inequality of condition \((D_N)\) implies

\[
g_{k+1} - g_k \leq \lambda \eta_k + \alpha \varepsilon_{k}^p.
\]

The sequel of the proof is similar to the one of Proposition 3.3 and the reader is referred to it. \(\square\)

**References**


