Algorithms for the Calculating
the Proximal Point

S. Kabbadj

Department of Mathematics
Faculty of Sciences of Meknes, B.P. 11201, Morocco

This article is distributed under the Creative Commons by-nc-nd Attribution License.
Copyright © 2019 Hikari Ltd.

Abstract

We propose two algorithms for calculating \(x^*\) proximal point of \(x\) defined by:

\[
x^* = \left(\nabla h + \lambda \partial f\right)^{-1}(\nabla h(x))
\]

The main results extend and improve the existing results. Moreover, the supposed conditions in our results are weaker than those of the existing results.

Keywords: convex optimization, proximal point, proximal point algorithm

1. Introduction

Let \(H\) a real Hilbert with the scalar product \(< . , . >\) and norm \(\| . \|\), \(T\) a maximal monotone operator. The problem

\[
(P) : " \text{find } \bar{x} \text{ such as } \bar{x} = (T + I)^{-1}(x)"
\]

has been studied by Aulender [2] when \(T = T_0 + \partial \chi_C\) where \(C\) is a closed convex and \(T_0\) is bounded on \(C\). Bruck [6] treated the case where the domain of \(T\) is open.

Alart [1] has studied \((P)\) when \(T = \partial f\) subdifferential from a convex function \(f\), proper and lower semicontinuous on \(R^d\) and that \(C := \text{dom} f\), of interior noted \(\text{int} (C)\) not empty, verifies the following conditions:
$H_1$: $C$ is a convex compact of $R^d$.

$H_2$: $\exists \varepsilon > 0, \forall x \in Fr(C), \forall u \in \partial \chi_C(x)$ and $u \neq 0$, $x - \alpha \frac{u}{\|u\|} \in \text{int}(C)$, $\forall \alpha \in ]0, \varepsilon].$

In other words, in every point $x$ of the boundary $Fr(C)$ of $C$, the opposite of every normal is an admissible direction with a step of a uniforme displacement.

This study allows to calculate the iterated $x^{k+1}$ from $x^k$ in Proximal Point Algorithm (PM) see [3,4,11,12,13]. In order to perform this calculation in Entropic Proximal Point Algorithm (PMD), see [7,9,14], we consider the following problem:

(Q): "Find $x^*$ such as $x^* = \text{prox}^h_{\lambda f}x := (\nabla h + \lambda \partial f)^{-1}(\nabla h(x))''$.

In this present labor, we propose two algorithms allowing to calculate $x^*$ while deleting $H_2$ and lightening $H_1$.

Throughout this paper, we assume:

(i) $f: R^d \to R \cup \{+\infty\}$ is a convex function, proper and lower semicontinuous.

(ii) $h: S \to R$ is a continuous, strictly convex on $S$ and continuously differentiable on $S$, where $S$ is an convex open subset of $R^d$.

(iii) $C := \text{dom} f \subset S$.

(iv) The problem (Q) admits a unique solution $x^*$ in $S$. (Conditions on $h$ and $f$ are required in [8, 10] to ensure the existence and uniqueness of $x^*$ in $S$).

Our notation is fairly standard, the closure of the set $A$ is denoted by $\overline{A}$ and $\chi_A$ is the characteristic function of $A$. For any convex function $f$, we denote by:

(1) $\text{dom} f = \{x \in R^d; f(x) < +\infty\}$ its effective domain,

(2) $\partial f(.) = \{v, f(y) \geq f(.) + \langle v, y - . \rangle - \varepsilon, \forall y\}$ its $\varepsilon$-subdifferential,
2. Study of a first algorithm

By eliminating the hypothesis $H_2$ and by keeping $H_1$, we propose the following algorithm:

--- Algorithm: $A_1(h)$

1: input: $x_0 \in C$
2: for $n = 0, 1, 2, \ldots$, do
3: pick $v_n \in \partial \varepsilon_n f(x_n)$, with $\varepsilon_n > 0$,
4: $w_n = x_n - \rho_n u_n$ with $\rho_n \geq 0$ and
   $u_n = \frac{\nabla h(x_n) - \nabla h(x)}{\lambda} + v_n$
5: $x_{n+1} = P(w_n)$, $P(w_n)$ is the projection of $w_n$ on $C$
6: end for.

The existence of the sequence $\{x_n\}_n$ generated by $A_1(h)$ is related to that of the sequence $\{v_n\}_n$.

For every $n$, $x_n \in \text{dom } f$ and $\varepsilon_n > 0$, so $\partial \varepsilon_n f(x_n) \neq \emptyset$. Which justifies the existence of the sequence $\{x_n\}_n$.

In the following we define the $D_h(\ldots)$ kernel by:

$D_h(x, y) := h(x) - h(y) - \langle x - y, \nabla h(y) \rangle$.

Lemma 2.1. [8, 10]

(i) $D_h(x, y) = \begin{cases} 0 & \text{if } x = y, \\ > 0 & \text{if } x \neq y \end{cases}$

(ii) $\forall x, y \in S : D_h(x, y) + D_h(y, x) = \langle x - y, \nabla h(x) - \nabla h(y) \rangle$.

(iii) If $h$ is strongly convex on $S$ with parametr $\alpha$ then,

$\forall x \in \mathcal{S}, \forall y \in S, D_h(x, y) \geq \frac{\alpha}{2} \|x - y\|^2$.

Proposition 2.2. Let $x$, $x'$, $y$ and $y' \in \mathbb{R}^d$. Let $\varepsilon, \varepsilon' \geq 0$ such as:

$x \in \partial \varepsilon f(x')$ and $y \in \partial \varepsilon' f(y')$. 

Then: \( < x - y, x' - y' > \geq -(\varepsilon + \varepsilon') \).

Proof. Since \( x \in \partial \varepsilon f(x') \) and \( y \in \partial \varepsilon' f(y') \), we have:

\[
\forall \ y \in \mathbb{R}^d, \ f(y) \geq f(x') + < y - x', x > - \varepsilon. \tag{4}
\]

\[
\forall \ z \in \mathbb{R}^d, \ f(z) \geq f(y') + < z - y', y > - \varepsilon'. \tag{5}
\]

By replacing \( y \) by \( y' \) in (4), \( z \) by \( x' \) in (5) and by adding (4) at (5), we deduce the result. \( \square \)

**Theorem 2.3.** We suppose that:

i. \( C \) is a compact of \( S \),

ii. \( \sum_{n=0}^{\infty} \rho_n = +\infty \) and \( \sum_{n=0}^{\infty} \rho_n (\varepsilon_n + \rho_n) < +\infty \),

iii. \( \{\|v_n\|\}_n \) is bounded.

Then \( \{x_n\}_n \) generated by \( A_1(h) \) converges to the solution \( x^* \) of \( (Q) \).

Proof. We have:

\[
v_n = u_n - \lambda^{-1} (\nabla h(x_n) - \nabla h(x)) \in \partial \varepsilon_n f(x_n)
\]
and

\[
\lambda^{-1} (\nabla h(x) - \nabla h(x^*)) \in \partial f(x^*).
\]

From the proposition 2.2.,

\[
< u_n - \lambda^{-1} (\nabla h(x_n) - \nabla h(x)) - \lambda^{-1} (\nabla h(x) - \nabla h(x^*)) , x_n - x^* > \geq -\varepsilon_n,
\]

From (2), we have:

\[
< u_n , x_n - x^* > \geq -\varepsilon_n + \lambda^{-1} [D_h(x_n, x^*) + D_h(x^*, x_n)].
\]

From (1), we have:

\[
< u_n , x_n - x^* > \geq -\varepsilon_n + \lambda^{-1} D_h(x_n, x^*). \tag{6}
\]

Since \( x^* \in C \), we have:

\[
\|x_{n+1} - x^*\|^2 = \|P(w_n) - P(x^*)\|^2 \leq \|w_n - x^*\|^2.
\]
Let again, 
\[ \|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - 2 < x_n - x^*, \rho_n u_n > + \rho_n^2 \|u_n\|^2. \]  
(7)

From i., the sequence \( \{x_n\}_n \) is bounded. Since \( \{x_n\}_n \subset C \subset \mathcal{S} \) and \( \nabla h \) is continuous on \( \mathcal{S} \), we deduce that \( \{\nabla h(x_n)\}_n \) is bounded. Therefore, the sequence \( \{u_n\}_n \) is bounded. So,

\[ \exists L > 0, \forall n \in \mathbb{N}, \|u_n\|^2 \leq L \]  
(8)

From (6), (7) and (8), we have

\[ 2\rho_n \lambda^{-1} D_h(x_n, x^*) + \|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 + 2\rho_n \varepsilon_n + L^2 \rho_n^2. \]

Let \( L' = \max\{2, L^2\} \), thus we have:

\[ 2\rho_n \lambda^{-1} D_h(x_n, x^*) + \|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 + L' \rho_n (\varepsilon_n + \rho_n). \]  
(9)

Let : \( a_n := \|x_n - x^*\|^2 \) and \( b_n := D_h(x_n, x^*) \). By adding the inequality (9) from 0 to \( k \), we have:

\[ 2\lambda^{-1} \sum_{n=0}^{n=k} \rho_n b_n + a_{k+1} \leq a_0 + L' \sum_{n=0}^{n=k} \rho_n (\varepsilon_n + \rho_n). \]  
(10)

From (10) and ii., we have:

\[ \sum_{n=0}^{\infty} \rho_n b_n < +\infty \]  
(11)

It exists the sub-sequence \( \{b_{j_n}\} \) of \( \{b_n\} \) such as : \( b_{j_n} \to 0 \). Otherwise, it will exist \( \varepsilon > 0, j_0 \) such that

\[ b_j \geq \varepsilon, \forall j \geq j_0 \]  
(12)

From (11) and (12), we have \( \sum_{n=0}^{\infty} \rho_n < +\infty \), which contradicts the hypothesis ii.

\( \{x_{j_n}\} \) is bounded, it exists the sub-sequence \( \{x_{k_n}\} \) of \( \{x_{j_n}\} \) such as : \( x_{k_n} \to u^* \in \mathcal{S} \).

On the other hand,

\[ b_{k_n} = D_h(x_{k_n}, x^*) \Rightarrow \lim b_{k_n} = \lim D_h(x_{k_n}, x^*) \]
\[ \Rightarrow 0 = D_h(u^*, x^*) \]
\[ \Rightarrow u^* = x^* \]
\[ \Rightarrow x_{j_n} \to x^* \]
\[ \Rightarrow a_{j_n} \to 0. \]
From (1) and (9), we have

\[ a_{n+1} \leq a_n + L' \rho_n (\varepsilon_n + \rho_n). \]  

(13)

\[ \sum_{n=0}^{\infty} \rho_n (\varepsilon_n + \rho_n) < +\infty \Rightarrow a_n \to l \in R. \] As \( a_j \to 0 \), this leads to the conclusion that \( a_n \to 0 \), that is \( x_n \to x^* \).

If the sequence \( \{\rho_n\}\) is defined like in Bruck [5] and \( h \) is a function strongly convex, then we can give an estimation the error due to the following proposition:

**Proposition 2.4.** In addition to the hypothesis of the theorem 2.3, let’s suppose:

i. \( h \) is strongly convex on \( S \) with parametr \( 2\alpha \),

ii. \( \forall n \geq 1, \rho_n = \varepsilon_n = \frac{1}{\alpha(n+\sigma)} \),

with \( \sigma = (\frac{\theta_1 K}{d})^2 \) where \( K = \frac{\lambda}{\alpha} \), \( \theta_1^2 = 2 + \sup_{1 \leq i \leq n} \{\|u_i\|^2\} \) and \( d \) designates the diameter of \( C \).

Then \( \forall n \geq 1 \),

\[ \|x_n - x^*\| \leq K d_n \theta_n, \]

where \( d_n := 1/\sqrt{(n + \sigma - 1)} \).

Proof. We have:

\[ 2\rho_n \lambda^{-1} D_h(x_n, x^*) + \|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 + 2\rho_n \varepsilon_n + \rho_n^2 \|u_n\|^2. \]

Since \( h \) is strongly convex on \( S \), from (3) we have,

\[ D_h(x_n, x^*) \geq \alpha \|x_n - x^*\|^2. \]

It follows that:

\[ 2\rho_n \lambda^{-1} \alpha \|x_n - x^*\|^2 + \|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 + 2\rho_n \varepsilon_n + \rho_n^2 \|u_n\|^2. \]

Let again

\[ \|x_{n+1} - x^*\|^2 \leq (1 - 2\rho_n \lambda^{-1} \alpha) \|x_n - x^*\|^2 + \rho_n^2 (2 + \|u_n\|^2). \]
Which leads
\[ \| x_{n+1} - x^* \|^2 \leq (1 - \frac{1}{n+\sigma})^2 \| x_n - x^* \|^2 + K^2 \frac{\theta_n^2}{(n+\sigma)^2} \theta_n^2. \] \quad (14)

Let the recurrence hypothesis:
\[ \| x_n - x^* \| \leq K d_n \theta_n. \]

That is true in the rank \( n=1 \), cause \( K d_1 \theta_1 = d \).

Let’s show that \( \| x_{n+1} - x^* \| \leq K d_{n+1} \theta_{n+1} \).

We have from (14):
\[ \| x_{n+1} - x^* \|^2 \leq (1 - \frac{1}{n+\sigma})^2 d_n^2 \theta_n^2 K^2 + \frac{K^2}{(n+\sigma)^2} \theta_n^2 \]

Let again,
\[ \| x_{n+1} - x^* \|^2 \leq \left[ (1 - \frac{1}{n+\sigma})^2 d_n^2 + \frac{1}{(n+\sigma)^2} \right] K^2 \theta_n^2 \]

Like:
\[ (1 - \frac{1}{n+\sigma})^2 d_n^2 + \frac{1}{(n+\sigma)^2} = \frac{1}{n+\sigma} = d_{n+1}^2 \] \quad (15)

And \( \{ \theta_n^2 \}_n \) is an increasing sequence, it follows that:
\[ \| x_{n+1} - x^* \|^2 \leq d_{n+1}^2 \theta_{n+1}^2 K^2, \]

from which the result.

□

**Remark 2.5.** On the practical plan, this estimation is better than the one obtained by Alart [1], cause it requires the knowledge of upper bound of the sequence \( \{v_n\}_n \).

### 3. Study of a second algorithm

In this paragraph, we give an algorithm whose convergence is realized without assuming the hypotheses \( H_1 \) and \( H_2 \). Let’s consider the algorithm:

```
Algorithm : A_2(h)
```

```
1: input: x_0 \in C
```
2: for \( n = 0, 1, 2, \ldots \), do
3: pick \( v_n \in \partial \varepsilon_n f(x_n) \), with \( \varepsilon_n > 0 \),
4: \( w_n = x_n - \frac{\rho_n}{\|u_n\| + r} u_n \) with \( r > 0 \), \( \rho_n > 0 \) and
\[
\begin{align*}
  u_n &= \frac{\nabla h(x_n) - \nabla h(x)}{\lambda} + v_n
\end{align*}
\]
5: \( x_{n+1} = P(w_n) \), \( P(w_n) \) is the projection of \( w_n \) on \( C \)
6: end for

The existence of the sequence \( \{x_n\}_n \) generated by \( A_2(h) \) is insured like in the algorithm \( A_1(h) \).

**Theorem 3.1.** Let’s suppose that :
- i. \( C \) is a closed of \( S \).
- ii. \( \sum_{n=0}^{\infty} \rho_n = +\infty \) and \( \sum_{n=0}^{\infty} \rho_n (\varepsilon_n + \rho_n) < +\infty \),
- iii. \( \{\|v_n\|\}_n \) is bounded.
Then \( \{x_n\}_n \) generated by \( A_2(h) \) converges to the solution \( x^* \) of \( (Q) \).

Proof. We have :
\[
\|x_{n+1} - x^*\|^2 \leq \|w_n - x^*\|^2.
\]
Which leads :
\[
\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - 2\frac{\rho_n}{\|u_n\| + r} < x_n - x^*, u_n > \rho_n^2 \|u_n\|^2 (\|u_n\| + r)^2
\]
(16)
The relation (6) demonstrated in the section 2. still being verified, we deduce then :
\[
-2\frac{\rho_n}{\|u_n\| + r} < x_n - x^*, u_n > \leq 2\frac{\rho_n \varepsilon_n}{\|u_n\| + r} - 2\frac{\rho_n \lambda^{-1}}{\|u_n\| + r} D_h(x_n, x^*).
\]
(17)
(16) and (17) lead that :
\[
2\frac{\rho_n \lambda^{-1}}{\|u_n\| + r} D_h(x_n, x^*) + \|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 + 2\frac{\rho_n \varepsilon_n}{\|u_n\| + r} + \rho_n^2.
\]
(18)
Let \( M = \max\{\frac{2}{r}, 1\} \), we have then :
\[
2\frac{\rho_n \lambda^{-1}}{\|u_n\| + r} D_h(x_n, x^*) + \|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 + M \rho_n (\rho_n + \varepsilon_n).
\]
(19)
We deduce that :
\[ \|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 + M\rho_n(\rho_n + \varepsilon_n). \]

\[ \sum_{n=0}^{\infty} \rho_n(\varepsilon_n + \rho_n) < +\infty, \quad \text{leqds thqt the sequence } \{x_n\}_n \text{ is bounde}, \quad \text{so the sequence } \{u_n\}_n \text{ is bounded}. \quad \text{Like in the section 2.}, \quad \text{we show that :} \]

\[ \sum_{n=0}^{\infty} \frac{2\lambda^{-1}\rho_n}{\|u_n\|} \cdot D_h(x_n, x^*) < +\infty \]

\{u_n\}_n \text{ is bounded}, \quad \text{so } \sum_{n=0}^{\infty} \rho_n b_n < +\infty, \quad \text{therefore } x_n \longrightarrow x^*, \quad \text{like in the sections 2.} \quad \square \]

By following the same demarche that in the section 2. , we give the following estimation :

**Proposition 3.2.** In addition to the hypotheses of the theorem 3.1. , let’s suppose :

i. \( h \) is strongly convex on \( S \) with parametr \( 2\alpha \).

ii. \( \|x_1 - x^*\| \leq R, \)

ii. \( \forall n \geq 1, \rho_n = \varepsilon_n = \frac{\lambda K}{\alpha(n+\sigma)}, \)

\( K := \sup \{r + \|u_n\|\} \) and \( \sigma = (\frac{L}{K})^2 \) where \( L^2 = (1 + \frac{2}{r}) \frac{\lambda^2 K^2}{\alpha^2}. \)

Then :

\[ \forall n \geq 1, \|x_n - x^*\|^2 \leq d_n^2 L^2. \]

where \( d_n := \frac{1}{\sqrt{(n+\sigma - 1)}} \)

Proof . From (18) we have

\[ \|x_{n+1} - x^*\|^2 \leq (1 - 2\alpha \rho_n \lambda^{-1} K^{-1})\|x_n - x^*\|^2 + \frac{r}{\lambda^2 \rho_n^2 + \rho_n^2}, \]

Let again ,

\[ \|x_{n+1} - x^*\|^2 \leq (1 - \frac{1}{n+\sigma})^2\|x_n - x^*\|^2 + (1 + \frac{2}{r}) \frac{\lambda^2 K^2}{\alpha^2} \frac{1}{(n+\sigma)^2}. \quad (20) \]

Let the hypothesis of the recurrence :

\[ \|x_n - x^*\|^2 \leq d_n^2 L^2. \]

That is true for \( n=1 \), because : \( d_1 L = R \) and \( \|x_1 - x^*\| \leq R. \)

Let’s show that : \( \|x_{n+1} - x^*\|^2 \leq d_{n+1}^2 L^2. \) From (20) ,
\[ \|x_{n+1} - x^*\|^2 \leq (1 - \frac{1}{n+\sigma})^2 d_n^2 L^2 + \frac{L^2}{(n+\sigma)^2} \]

From (15), we have:
\[ \|x_{n+1} - x^*\|^2 \leq L^2((1 - \frac{1}{n+\sigma})^2 d_n^2 + \frac{1}{(n+\sigma)^2}) = L^2 d_{n+1}^2, \]

From where the desired inequality.

\[ \square \]

4. Conclusion

The results of convergence of the algorithms making allowing to solve (P) are established by [1] under restrictive conditions such as \( H_1 \) and \( H_2 \). In our analysis, the considered problem (Q) generalises (P) and the results of convergence of the proposed algorithms are established by deleting \( H_2 \) and by lightning \( H_1 \), more precisely:

(i) For \( \lambda = 1 \) and \( h(.) = \frac{1}{2} \| . \|^2 \), \((P) \Leftrightarrow (Q)\)

(ii) With the compacity of the dom \( f \), the algorithm \( A_1(h) \) converges to \( x^* \) solution of (Q) with an estimation of the error, better than the one given by Bruck [5].

(iii) Without the compacity of the dom \( f \), the algorithm \( A_2(h) \) converges to \( x^* \) which improves the results given by Alart [1].

References


Received: September 7, 2019; Published: October 2, 2019