Serdica
Mathematical Journal
Сердика
Математическо списание

Serdica Math. J. **26** (2000), 177-206

Serdica
Mathematical Journal

Institute of Mathematics
Bulgarian Academy of Sciences

PERTURBED PROXIMAL POINT ALGORITHM WITH NONQUADRATIC KERNEL

M. Brohe, P. Tossings

Communicated by A. L. Dontchev

Abstract. Let H be a real Hilbert space and T be a maximal monotone operator on H.

A well-known algorithm, developed by R. T. Rockafellar [16], for solving the problem

(P) "To find $\overline{x} \in H$ such that $0 \in T\overline{x}$"

is the proximal point algorithm.

Several generalizations have been considered by several authors: introduction of a perturbation, introduction of a variable metric in the perturbed algorithm, introduction of a pseudo-metric in place of the classical regularization, ...

We summarize some of these extensions by taking simultaneously into account a pseudo-metric as regularization and a perturbation in an inexact version of the algorithm.

2000 *Mathematics Subject Classification:* 47H04, 47H05, 47H09, 47H10.

Key words: Proximal point algorithm, Bregman functions, generalized resolvent operator, variational convergence.

1. Introduction and preliminaries. Let H be a real Hilbert space with inner product $\langle .,. \rangle$ and associated norm $\|.\|$.

Maximal monotone operators on H have been extensively studied because of their role in convex analysis. In this context, a fundamental problem consists in

(P) "To find $\overline{x} \in H$ such that $0 \in T\overline{x}$."

A well-known approach for solving problem (P) is to use the *proximal point algorithm* developed by R. T. Rockafellar [16]. This algorithm generates, from any starting point $y_0 \in H$, a sequence (y_n) in H, by the scheme

$$y_n = J^T_{\lambda_n} y_{n-1} + e_n, \quad \forall\, n \in \mathbb{N}^*,$$

where (λ_n) is a sequence of positive real numbers bounded away from zero, (e_n) a sequence in H taking into account a possible inexact computation and $J^T_\lambda = (I+\lambda T)^{-1}$ the resolvent operator associated with T with parameter $\lambda > 0$ [1]. This is a single-valued nonexpansive mapping defined everywhere on H:

$$\|J^T_\lambda x - J^T_\lambda y\| \le \|x - y\|, \quad \forall\, x, y \in H,$$

and such that

$$J^T_\lambda \overline{x} = \overline{x} \quad \Leftrightarrow \quad 0 \in T\overline{x}.$$

In a first time, B. Lemaire [10] studied the perturbation of this algorithm for $T = \partial f$, the subdifferential of a proper closed convex function on H.

Inspired by H. Attouch and R. J. B. Wets' work [2], P. Tossings [17] introduced the *variational metric* between two maximal monotone operators T^1 and T^2 with parameters $\lambda > 0$ and $\rho \ge 0$:

$$\delta_{\lambda,\rho}(T^1, T^2) = \sup_{\|x\| \le \rho} \left\| J^{T^1}_\lambda x - J^{T^2}_\lambda x \right\|,$$

and an associated notion of convergence.

Thanks to this notion of convergence, P. Tossings [18] studied a perturbed version of the proximal point algorithm, which generates, from any starting point $x_0 \in H$, a sequence (x_n) in H, by the recursive rule

$$x_n = J^{T^n}_{\lambda_n} x_{n-1} + e_n, \quad \forall\, n \in \mathbb{N}^*,$$

the maximal monotone operators T^n approaching T in a certain sense tied to the variational metric.

Another classical algorithm for solving initial problem (P) is the A. Renaud and G. Cohen's Auxiliary Problem Principle [14]. In the symmetric case,

[1] We will sometimes call it the "classical" resolvent operator.

their algorithm generates a sequence (x_n) by the iterative scheme

$$x_n = (\nabla h + \lambda_n T)^{-1} \nabla h(x_{n-1}), \quad \forall\, n \in \mathbb{N}^*, \tag{1}$$

where h denotes a real-valued strongly convex function, assumed to be Gateaux differentiable. Using the nonlinear change of coordinates $\nabla h(x_n) = u_n$ $(n \in \mathbb{N})$, the scheme (1) may be rewritten as

$$u_n = \Big(I + \lambda_n T(\nabla h)^{-1}\Big)^{-1} u_{n-1}, \quad \forall\, n \in \mathbb{N}^*. \tag{2}$$

At this point, scheme (2) is nothing but the proximal point algorithm applied to the operator $T(\nabla h)^{-1}$. However, the composition of monotone operators fails to be monotone, in general. In consequence, the results obtained in the context of the proximal point algorithm are not directly applicable.

When T is equal to ∂f, the subdifferential of a lower semicontinuous convex function f, then iteration (1) can be rewritten in the following equivalent form, studied by S. Kabbadj [9],

$$x_n = \operatorname*{Argmin}_{x \in H} \left\{ f(x) + \frac{1}{\lambda_n} D_h(x, x_{n-1}) \right\}, \quad \forall\, n \in \mathbb{N}^*, \tag{3}$$

where

$$D_h(x,y) = h(x) - h(y) - \langle \nabla h(y), x - y\rangle, \quad \forall\, x, y \in H. \tag{4}$$

This equality appears as a generalized proximal rule, based upon a Moreau-Yosida regularization, since

$$D_h(x,y) = \frac{1}{2}\|x-y\|^2 \quad \text{if} \quad h(x) = \frac{1}{2}\|x\|^2, \qquad \forall\, x, y \in H.$$

One calls h a *Bregman function* if it is defined on a nonempty open convex subset S of H and if it has certain additional properties. In his original paper [3], Bregman gave a set of axioms describing D-functions and offered (4) as one mean of constructing them. The term "Bregman function" was coined by Y. Censor and A. Lent [5].

The aim of this paper is to introduce, for a strongly convex function h defined on a nonempty open convex subset S of H, an error term and a perturbation, in the symmetric case of A. Renaud and G. Cohen's rule, i.e.

$$x_n \in S, \quad \Big\| x_n - (\nabla h + \lambda T^n)^{-1} \nabla h(x_{n-1}) \Big\| \le \varepsilon_n, \quad \forall\, n \in \mathbb{N}^*,$$

where (ε_n) is a sequence of positive real numbers, introduced to take into account a possible inexact computation.

This rule is equivalent to

$$x_n \in S, \quad x_n = (\nabla h + \lambda T^n)^{-1} \nabla h(x_{n-1}) + e_n, \quad \forall\, n \in \mathbb{N}^*, \tag{5}$$

where (e_n) is a sequence of H satisfying

$$\|e_n\| \le \varepsilon_n, \quad \forall\, n \in \mathbb{N}^*.$$

In the next section, the most important properties of function $D_h(x,y)$ are presented as well as a definition of the generalized resolvent operator and his properties. In Section 3, we establish, as a preliminary result, the weak convergence of the sequence (x_n) generated by the *nonperturbed* rule

$$x_n \in S, \quad \left\| x_n - (\nabla h + \lambda T)^{-1} \nabla h(x_{n-1}) \right\| \le \varepsilon_n, \quad \forall\, n \in \mathbb{N}^*,$$

where (ε_n) is a sequence of positive real numbers. As in the nonperturbed case, this rule takes the equivalent form

$$\text{(PRh)} \qquad x_n = (\nabla h + \lambda T)^{-1} \nabla h(x_{n-1}) + e_n, \quad \forall\, n \in \mathbb{N}^*.$$

Finally, Section 4 presents a convergence theory for iterates of the form (5), analogous to the R. T. Rockafellar's one for the nonperturbed version.

In Sections 3 and 4, assumptions on the error terms are formulated on the sequence (e_n). Obviously, they can be expressed, more restrictively, in terms of the sequence (ε_n).

At the end of this paper, Theorem 16 appears as a particular case of Theorem 19. However, the technic for proving Theorem 19 consists precisely in generating an auxiliary sequence satisfying assumptions ensuring the convergence in the nonperturbed case. Therefore, it appears necessary to establish Theorem 16 in a first time.

The convergence of a sequence (x_n) generated by the rule (PRh) has been studied recently by J. Eckstein [8]. He reduces hypothesis on the auxiliary function h and assumes that h is "only" strictly convex (and satisfies other subordinated assumptions) on a nonempty open convex subset S. Nevertheless, he imposes more conditions on the error sequence (e_n) than us. Our formulation presents, in addition of the perturbation, the great advantage to assume, on the error sequence (e_n), exactly the same hypothesis than in the classical proximal point algorithm.

2. Pseudo-metric and resolvent operator with nonquadratic kernel. In this section, h will denote a real-valued strongly convex function on $\overline{S}$, with constant $\alpha > 0$,

$$h(\theta x + (1-\theta)y) \le \theta h(x) + (1-\theta)h(y) - \frac{\alpha}{2}\theta(1-\theta)\|x-y\|^2\,,\ \forall\, x\,, y \in \overline{S},\ \theta \in [0,1],$$

assumed to be Gateaux differentiable on S, a nonempty open convex subset of H.

2.1.Pseudo-metric associated with h. Under certain conditions, Proposition 2.1 below, proved in S. Kabbadj [9, Proposition 1.3.6] shows the equivalence between the pseudo-metric D_h defined by (4) on $\overline{S} \times S$ and the norm associated with the inner product.

Proposition 1.

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

Furthermore, if ∇h is Lipschitz continuous with constant $M > 0$, then

$$D_h(x, y) \leq \frac{M}{2}\|x - y\|^2, \quad \forall\,(x, y) \in \overline{S} \times S.$$

Lemma 2.

$$D_h(x, b) - D_h(x, a) + D_h(b, a) = \langle \nabla h(a) - \nabla h(b), x - b\rangle, \quad \forall\, x \in \overline{S},\ \forall a, b \in S.$$

Definition 3. *A mapping P is said to be* firmly nonexpansive for D_h *if its restriction to S takes its values in S and if*

$$\begin{aligned} D_h(Px_1, Px_2) + D_h(Px_2, Px_1) \leq{}& D_h(Px_1, x_2) - D_h(Px_1, x_1) \\ &+ D_h(Px_2, x_1) - D_h(Px_2, x_2), \quad \forall\, x_1, x_2 \in D(P) \cap S. \end{aligned}$$

Remark 4.

- The function D_h defined on $H \times H$, coincides with the half of the square of the distance associated with the norm when h is the half of the square of the norm.
- In this case, a firmly nonexpansive mapping for D_h is a 1-firmly nonexpansive mapping for the norm.

Proposition 5. *Let f be a real-valued convex function on H, Gateaux differentiable, whose gradient ∇f is Lipschitz continuous with constant $M > 0$.*

For all $b \in \left]0, \frac{2}{M}\right[$, $\left(I - b\nabla f\right)$ is c-firmly nonexpansive with $c = \frac{2}{bM} - 1$.

Moreover, if f is strongly convex with constant $\alpha > 0$, there is $\beta \in]0, 1[$ such that

$$\|(I - b\nabla f)(x) - (I - b\nabla f)(y)\| \leq \beta\|x - y\|, \quad \forall\, x, y \in H.$$

Proof. The first part of the result is established in B. Lemaire [12, Lemma 3.2], and gives us

$$\|(I - b\,\nabla f)(x) - (I - b\,\nabla f)(y)\|^2 + \Big(\frac{2}{bM} - 1\Big)\|b\,\nabla f(x) - b\,\nabla f(y)\|^2 \\ \leq \|x - y\|^2, \quad \forall\, x, y \in H.$$

For the second part, it suffices to note that the strong monotonicity of ∇f and the Cauchy-Schwarz' inequality lead to

$$\|(I - b\nabla f)x - (I - b\nabla f)y\|^2 \leq \|x - y\|^2 + \Big(1 - \frac{2}{bM}\Big)b^2\alpha^2\|x - y\|^2,$$

that is to say the conclusion for

$$\beta = \sqrt{1 + \Big(1 - \frac{2}{bM}\Big)b^2\alpha^2}. \qquad \square$$

Proposition 6. *Let T be a maximal monotone operator on H, S a nonempty open convex subset of H and h a real-valued convex function on $\overline{S}$, Gateaux differentiable on S, with hemicontinuous gradient. If either*

(i) $S = H$ *or*

(ii) $S \cap D(T) \neq \emptyset$ *and* $R\,(\nabla h) = H$ *or*

(iii) $D(T) \subset S$,

then the operator $\nabla h + T$ is maximal monotone on $S \cap D(T)$.

Proof. Let A be a monotone operator enclosing ∇h. Since S is open and convex, we prove, as H. Brezis [4, Proposition 2.4] that A and ∇h coincide on S. The conclusion follows immediately for the first hypothesis.

For the second ones, let $u \in Ax$, $x \notin S$. The convex separation theorem ensures the existence of $\gamma \in H$ satisfying $\langle \gamma, x - y\rangle < 0$, for all $y \in S$. Since $R\ (\nabla h) = H$, there is $\overline{y} \in S$ such that $\langle u - \nabla h(\overline{y}), x - \overline{y}\rangle < 0$. This is not possible since A is monotone and, so, $A = \nabla h$. We conclude by using H. Brezis [4, Corollary 2.7].

Finally, for the third hypothesis, assume that A is a maximal monotone operator. H. Brezis [4, Corollary 2.7] involves that $A + T$ is maximal monotone too and so is $\nabla h + T$ over $D(T)$. $\square$

Proposition 7. *Let T be a maximal monotone operator on H, S a nonempty open convex subset of H and h a real-valued function on $\overline{S}$, Gateaux differentiable on S, such that the sum $\nabla h + T$ is maximal monotone. If either*

(i) *h is a strictly convex function and S is bounded or*

(ii) *h is a strongly convex function,*

then the operator $(\nabla h + T)^{-1}$ is defined over all H. Moreover, the operator $(\nabla h + T)^{-1}\nabla h$ is a firmly nonexpansive mapping for D_h.

Proof. For (i), $\nabla h + T$ is a maximal monotone operator, with bounded domain. The conclusion follows from H. Brezis [4, Corollary 2.2] and the strict monotonicity of $\nabla h + T$. For (ii), see H. Brezis [4].

For the firm nonexpansivity of the above-mentioned operator, use the monotonicity of the operator T and the Lemma 2. □

Proposition 8. *Let P be a firmly nonexpansive mapping for D_h. If ∇h is Lipschitz continuous with constant $M > 0$ then*

$$\|Px_1 - Px_2\| \leq \frac{M}{\alpha}\|x_1 - x_2\|, \quad \forall\, x_1, x_2 \in D(P) \cap S.$$

Proof. Three successive applications of Lemma 2 give us, for $x_1, x_2 \in D(P) \cap S$,

$$\begin{aligned} D_h(Px_1, x_2) - D_h(Px_1, x_1) &= -D_h(x_2, x_1) + \langle \nabla h(x_1) - \nabla h(x_2), Px_1 - x_2\rangle, \\ D_h(Px_2, x_1) - D_h(Px_2, x_2) &= -D_h(x_1, x_2) + \langle \nabla h(x_2) - \nabla h(x_1), Px_2 - x_1\rangle \end{aligned}$$

and

$$D_h(x_2, x_1) + D_h(x_1, x_2) = \langle \nabla h(x_2) - \nabla h(x_1), x_2 - x_1\rangle.$$

So, we can rewrite the firm nonexpansivity of the mapping P for D_h as

$$D_h(Px_1, Px_2) + D_h(Px_2, Px_1) \leq \langle \nabla h(x_1) - \nabla h(x_2), Px_1 - Px_2\rangle.$$

The inequality

$$2\langle a, b\rangle \leq \gamma\|a\|^2 + \frac{1}{\gamma}\|b\|^2, \quad \forall\, a, b \in H,\ \forall\, \gamma \in\,]0, +\infty[, \tag{6}$$

written with $\gamma = \alpha^{-1}$, joined to Proposition 1, implies

$$D_h(Px_1, Px_2) + D_h(Px_2, Px_1) \leq \frac{M^2}{\alpha}\|x_1 - x_2\|^2, \quad \forall\, x_1, x_2 \in D(P) \cap S.$$

The conclusion arises by applying Proposition 1 again. □

2.2. Resolvent operator with nonquadratic kernel.

Definition 9. *Let T be a maximal monotone operator on H and λ a positive number. If one condition of Proposition 6 is satisfied, we can define the* resolvent operator with nonquadratic kernel *associated with T, with parameters h, λ or, more simply,* generalized resolvent operator *associated with T, with* *parameters* h*,* λ*, by*

$$J_\lambda^{h,T} = (\nabla h + \lambda T)^{-1}\nabla h,$$

and

$$A_\lambda^{h,T}x = \frac{\nabla h(x) - \nabla h(J_\lambda^{h,T}x)}{\lambda}$$

the corresponding Yosida approximation.

Consequently, the immediate following properties state.

Proposition 10.

(i) $J_\lambda^{h,T}$ *and* $A_\lambda^{h,T}$ *are single-valued operators that satisfy*

$$\overline{x} \in S \cap T^{-1}(0) \quad \Leftrightarrow \quad J_\lambda^{h,T}\overline{x} = \overline{x}, \ \forall \lambda > 0 \quad \Leftrightarrow \quad A_\lambda^{h,T}\overline{x} = 0, \ \forall \lambda > 0$$

and

$$A_\lambda^{h,T}x \in T\Big(J_\lambda^{h,T}x\Big), \quad \forall x \in S, \ \forall \lambda > 0.$$

(ii)

$$\nabla h(x) - \nabla h(J_\lambda^{h,T}x) \in \left(I + \nabla h\, T^{-1}\frac{1}{\lambda}\right)^{-1} \nabla h(x), \quad \forall x \in S, \quad \forall \lambda > 0.$$

(iii)

$$J_\lambda^{h,T}x = J_\mu^T \left[\frac{\mu}{\lambda}\nabla h(x) + J_\lambda^{h,T}x - \frac{\mu}{\lambda}\nabla h\Big(J_\lambda^{h,T}x\Big)\right], \quad \forall x \in S, \ \forall \lambda, \mu > 0.$$

2.3. Variational metric associated with h

Convention. *From now on,* S *denotes a nonempty open convex subset of* H *and* h *a real valued strongly convex function, with positive constant* α *on* $\overline{S}$*, Gateaux differentiable on* S*, with Lispchitz and weakly continuous gradient* ∇h*, with positive constant* $M > 0$*.*

Moreover, T *and* T^n $(n \in \mathbb{N}^*)$ *denote maximal monotone operators on* H*, satisfying* $T^{-1}(0) \neq \emptyset$*,* $D(T) \subset S$ *and one of the following conditions:*

(i) $S = H$*,*

(ii) $S \cap D(T^n) \neq \emptyset$ *and* $R(\nabla h) = H$*,*

(iii) $D(T^n) \subset S$*.*

Remark 11.

• In finite dimension, the weak continuity of the gradient ∇h is ensured by the Lipschitz property.

• The hypothesis imposed on h are apparently more restrictive than those imposed by Bregman, for example. Nevertheless, an appropriate choice of the function h and the associated parameters allows us to work over any bounded domain and so to cover the optimal set of any practical problem.

Compared with those realized with the classical proximal point algorithm, our numerical tests show clearly a significant improvement both in the number of iterations and in the precision on the obtained solution.

Moreover, in some cases, optimal point, lying on the boundary of the domain of h, are attained with a surprising facility. This fact recalls us that the given theoretical conditions are sufficient but not necessary and open the door to future researches.

Definition 12. *Let T^1 and T^2 be two maximal monotone operators on H, $\lambda > 0$ and $\rho \geq 0$. The* variational metric between T^1 and T^2, associated with h, with parameters λ, ρ *is the semi-distance*[2]

$$\delta^h_{\lambda,\rho}(T^1, T^2) = \sup_{x \in S, \|x\| \leq \rho} \left\| J^{h,T^1}_\lambda x - J^{h,T^2}_\lambda x \right\|.$$

Proposition 10 (iii), allows us to compare this new metric to the classical one.

Proposition 13. *Let T^1 and T^2 be two maximal monotone operators on H and $\rho \geq 0$. If $\lambda^*, \mu > 0$ satisfy $\dfrac{\mu}{\lambda^*} < \dfrac{2}{M}$, then there is a constant $C > 0$ such that, for all $\lambda \geq \lambda^*$,*

$$\delta^h_{\lambda,\rho}(T^1, T^2) \leq C\, \lambda\, \delta_{\mu,\rho_0}(T^1, T^2),$$

for all

$$\rho_0 \geq \left[\frac{\mu}{\lambda} M + \Big(\frac{\mu}{\lambda} M + 1\Big)\frac{M}{\alpha}\right]\rho + \Big(\frac{\mu}{\lambda} M + 1\Big)\Big(\frac{M}{\alpha}\|x^*\| + \left\|J^{h,T^1}_\lambda x^*\right\|\Big),$$

where x^* is some point of S.

Moreover, if T^1 has at least one zero x^*, then the minimal value imposed on ρ_0 can be replaced by

$$\Big(\frac{\mu}{\lambda} M + \frac{M}{\alpha} + \frac{\mu}{\lambda}\frac{M^2}{\alpha}\Big)\rho + \Big(1 + \frac{\mu}{\lambda} M\Big)\Big(\frac{M}{\alpha} + 1\Big)\|x^*\|.$$

[2]We simply say generalized variational metric.

P r o o f. Proposition 10 (iii), involves

$$\begin{aligned}\delta^h_{\lambda,\rho}(T^1,T^2) \quad \leq \quad & \sup_{x\in S,\|x\|\leq\rho}\Big\| J^{T^1}_\mu\Big[\frac{\mu}{\lambda}\nabla h(x)+J^{h,T^1}_\lambda x-\frac{\mu}{\lambda}\nabla h(J^{h,T^1}_\lambda x)\Big] \\ & \qquad - J^{T^2}_\mu\Big[\frac{\mu}{\lambda}\nabla h(x)+J^{h,T^1}_\lambda x-\frac{\mu}{\lambda}\nabla h(J^{h,T^1}_\lambda x)\Big]\Big\| \\ & + \sup_{x\in S,\|x\|\leq\rho}\Big\| J^{T^2}_\mu\Big[\frac{\mu}{\lambda}\nabla h(x)+J^{h,T^1}_\lambda x-\frac{\mu}{\lambda}\nabla h(J^{h,T^1}_\lambda x)\Big] \\ & \qquad - J^{T^2}_\mu\Big[\frac{\mu}{\lambda}\nabla h(x)+J^{h,T^2}_\lambda x-\frac{\mu}{\lambda}\nabla h(J^{h,T^2}_\lambda x)\Big]\Big\|.\end{aligned}$$

For the first term of the overestimation, we have, by Proposition 8,

$$\begin{aligned}&\Big\|\frac{\mu}{\lambda}\nabla h(x)+J^{h,T^1}_\lambda x-\frac{\mu}{\lambda}\nabla h(J^{h,T^1}_\lambda x)\Big\| \\ &\qquad \leq \Big[\frac{\mu}{\lambda}M+\Big(\frac{\mu}{\lambda}M+1\Big)\frac{M}{\alpha}\Big]\|x\|+\Big(\frac{\mu}{\lambda}M+1\Big)\Big(\frac{M}{\alpha}\|x^*\|+\Big\|J^{h,T^1}_\lambda x^*\Big\|\Big),\end{aligned}$$

so

$$\begin{aligned}&\sup_{x\in S,\|x\|\leq\rho}\Big\| J^{T^1}_\mu\Big[\frac{\mu}{\lambda}\nabla h(x)+J^{h,T^1}_\lambda x-\frac{\mu}{\lambda}\nabla h(J^{h,T^1}_\lambda x)\Big] \\ &\qquad\qquad -J^{T^2}_\mu\Big[\frac{\mu}{\lambda}\nabla h(x)+J^{h,T^1}_\lambda x-\frac{\mu}{\lambda}\nabla h(J^{h,T^1}_\lambda x)\Big]\Big\| \\ &\qquad \leq \delta_{\mu,\rho_0}(T^1,T^2),\end{aligned}\tag{7}$$

for all ρ_0 such that

$$\rho_0\geq\Big[\frac{\mu}{\lambda}M+\Big(\frac{\mu}{\lambda}M+1\Big)\frac{M}{\alpha}\Big]\rho+\Big(\frac{\mu}{\lambda}M+1\Big)\Big(\frac{M}{\alpha}\|x^*\|+\Big\|J^{h,T^1}_\lambda x^*\Big\|\Big).$$

For the second term, the resolvent operator's nonexpansivity and Proposition 5 imply

$$\begin{aligned}&\Big\|J^{T^2}_\mu\Big[\frac{\mu}{\lambda}\nabla h(x)+J^{h,T^1}_\lambda x-\frac{\mu}{\lambda}\nabla h(J^{h,T^1}_\lambda x)\Big]-J^{T^2}_\mu\Big[\frac{\mu}{\lambda}\nabla h(x)+J^{h,T^2}_\lambda x-\frac{\mu}{\lambda}\nabla h(J^{h,T^2}_\lambda x)\Big]\Big\| \\ &\qquad \leq \beta\Big\|J^{h,T^1}_\lambda x-J^{h,T^2}_\lambda x\Big\|\end{aligned}\tag{8}$$

for

$$\beta = \sqrt{1 + \Big(1 - \frac{2\lambda}{M\mu}\Big)\frac{\mu^2}{\lambda^2}\alpha^2}.$$

Consequently, inequalities (7) and (8) give

$$\delta^h_{\lambda,\rho}(T^1, T^2) \le \frac{1}{1-\beta}\delta_{\mu,\rho_0}(T^1, T^2),$$

for all ρ_0 such that

$$\rho_0 \ge \Big[\frac{\mu}{\lambda}M + \Big(\frac{\mu}{\lambda}M + 1\Big)\frac{M}{\alpha}\Big]\,\rho + \Big(\frac{\mu}{\lambda}M + 1\Big)\Big(\frac{M}{\alpha}\|x^*\| + \Big\|J^{h,T^1}_\lambda x^*\Big\|\Big),$$

where the definition of β allows us to write

$$\frac{1}{1-\beta} = \frac{1+\beta}{1 - \left(1 + \Big(2 - \dfrac{2\lambda}{M\mu}\Big)\dfrac{\mu^2}{\lambda^2}\alpha^2\right)} \le C\,\lambda$$

for

$$C = \frac{2}{\mu\alpha^2\Big(\dfrac{2}{M} - \dfrac{\mu}{\lambda^*}\Big)}.$$

For the particular case, since $x^* \in (T^1)^{-1}(0) \cap S$ if and only if $J^{h,T^1}_\lambda x^* = x^*$, we get, from Proposition 8,

$$\Big\|x^* - J^{h,T^1}_\lambda x\Big\| \le \frac{M}{\alpha}\|x^* - x\|$$

and, finally,

$$\Big\|J^{h,T^1}_\lambda x\Big\| \le \frac{M}{\alpha}\|x\| + \left(\frac{M}{\alpha} + 1\right)\|x^*\|,$$

for all $x \in S$.

The conclusion follows immediately. □

Proposition 14. *Let T^n $(n \in \mathbb{N}^*)$ and T be maximal monotone operators on H. If $0 < \underline{\lambda} \le \lambda_n$, $\forall\, n \in \mathbb{N}^*$ and*

$$\lim_{n\to+\infty} \delta^h_{\lambda_n,\rho}(T^n, T) = 0, \quad \forall\, \rho \ge 0,$$

then

$$\text{(P)} \qquad \lim_{n\to+\infty} \Big\|A^{h,T^n}_{\lambda_n} x - A^{h,T}_{\lambda_n} x\Big\| = 0, \quad \forall\, x \in S.$$

In particular

$$\lim_{n\to+\infty} \Big\|A^{h,T^n}_{\lambda_n} \overline{x}\Big\| = 0, \quad \forall\, \overline{x} \in H \text{ such that } 0 \in T\overline{x}.$$

Proof. The assumption made on $\delta^h_{\lambda_n,\rho}$ and the Lipschitz continuity of ∇h lead us to

$$\lim_{n\to+\infty} \left\| \nabla h(J^{h,T^n}_{\lambda_n} x) - \nabla h(J^{h,T}_{\lambda_n} x) \right\| = 0, \quad \forall\, x \in S,$$

and so to the first part of the result, thanks to hypothesis $0 < \underline{\lambda} \leq \lambda_n$ and Definition 9.

The particular case results from Proposition 10 (i). $\square$

3. Proximal point algorithm with nonquadratic kernel. For proving weak convergence of the sequence (x_n) generated by the rule (PRh), we recall the following technical lemma proved by B. Polyak [13].

Lemma 15. *Let (z_n), (ε_n) and (C_n), $n \in \mathbb{N}^*$, be three sequences of nonnegative numbers such that*

$$\sum_{n=1}^{+\infty} \varepsilon_n < +\infty \quad \text{and} \quad \sum_{n=1}^{+\infty} C_n < +\infty.$$

If there is a range $M \in \mathbb{N}^*$ from which

$$z_n \leq (1 + C_n) z_{n-1} + \varepsilon_n, \tag{9}$$

then (z_n) is convergent.

Theorem 16. *Assume that problem* (P) *has at least one solution and*

(i) $0 < \underline{\lambda} \leq \lambda_n$, $\forall\, n \in \mathbb{N}^*$,

(ii) $\sum_{n=1}^{+\infty} \|e_n\| < +\infty$.

Then the sequence (x_n) generated by the rule (PRh) *weakly converges to some solution of* (P) *and satisfies*

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

Proof. Let $\overline{x} \in S$ be a zero of T. We know that

$$\overline{x} = J^{h,T}_{\lambda_n} \overline{x}, \quad \forall\, n \in \mathbb{N}^*.$$

By another way, we can write

$$x_n = u_n + e_n, \quad \forall\, n \in \mathbb{N}^*,$$

where

$$u_n = J^{h,T}_{\lambda_n} x_{n-1}, \quad \forall\, n \in \mathbb{N}^*.$$

From this, the proof will be separated into four parts.

$\boxed{1}$ *The sequence (x_n) is bounded.*

The firm nonexpansivity of $J^{h,T}_{\lambda_n}$ for D_h, Lemma 2 and the positivity of D_h imply

$$D_h(\overline{x}, x_n) \leq D_h(\overline{x}, x_{n-1}) + \langle \nabla h(u_n) - \nabla h(x_n), \overline{x} - x_n \rangle, \quad \forall\, n \in \mathbb{N}^*. \tag{10}$$

The inequality (6), written with

$$\gamma = \frac{\alpha \|e_n\|}{1 + \|e_n\|},^{3}$$

and Proposition 1 lead us to

$$D_h(\overline{x}, x_n) \leq (1 + \|e_n\|) D_h(\overline{x}, x_{n-1}) + \frac{M^2}{\alpha} \|e_n\|, \quad \forall\, n \geq \mathbb{N}, \tag{11}$$

where $N \in \mathbb{N}^*$ is chosen to have

$$\|e_n\| \leq \sqrt{2} - 1, \quad \forall\, n \geq N.$$

Lemma 15 ensures therefore the convergence of $(D_h(\overline{x}, x_n))$ and so the boundedness of (x_n).

Subsequently, we will denote, for $\overline{x}$ solution of (P),

$$l(\overline{x}) = \lim_{n \to +\infty} D_h(\overline{x}, x_n).$$

$\boxed{2}$ *The sequence (x_n) satisfies*

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

Lemma 2, the positivity of D_h and formula (10) involve

$$D_h(\overline{x}, x_n) + D_h(x_n, x_{n-1}) \leq D_h(\overline{x}, x_{n-1})$$
$$+ \langle \nabla h(u_n) - \nabla h(x_n), \overline{x} - x_n \rangle + \langle \nabla h(x_{n-1}) - \nabla h(x_n), u_n - x_n \rangle,$$

where, from assumption (ii) and the Lipschitz continuity of ∇h, the two inner products and the positive sequence $D_h(x_n, x_{n-1})$ converge to zero.

[3]Remark that, if $\|e_n\| = 0$, then the inner product is also equal to zero and inequality (11) stays true.

$\boxed{3}$ *Every weak cluster point of* (x_n) *is a solution of (P) and so is a point of* S.

Let x^* be such a point and (x_{n_k}) a subsequence of (x_n) weakly convergent to x^*. For all $w, z \in H$ such that $z \in T(w)$, the monotonicity of T implies

$$\Big\langle w - x_{n_k} + e_{n_k}, z - \frac{\nabla h(x_{n_k-1}) - \nabla h(u_{n_k})}{\lambda_{n_k}} \Big\rangle \geq 0, \quad \forall\, k \in \mathbb{N}.$$

Assumption (i), joined to $\boxed{2}$ and the Lipschitz continuity of ∇h, provides

$$\frac{\nabla h(x_{n_k-1}) - \nabla h(x_{n_k})}{\lambda_{n_k}} \xrightarrow{s} 0 \quad \text{when} \quad k \to +\infty,$$

and, passing to the limit for $k \to +\infty$,

$$\langle w - x^*, z \rangle \geq 0.$$

The maximality of T leads in turn to $0 \in T(x^*)$.

$\boxed{4}$ *The sequence* (x_n) *is weakly convergent to a solution of* (P).

Assume that (x_n) has two weak cluster points x_1^* and x_2^* and write, according to the convention taken in $\boxed{1}$,

$$l(x_1^*) = \lim_{n\to+\infty} D_h(x_1^*, x_n) \quad \text{and} \quad l(x_2^*) = \lim_{n\to+\infty} D_h(x_2^*, x_n).$$

If (x_{n_l}) is a subsequence of (x_n) weakly convergent to x_2^*, we get, from the definition of D_h,

$$\begin{aligned} l(x_1^*) - l(x_2^*) &= \lim_{l\to+\infty} \Big[h(x_1^*) - h(x_2^*) - \langle \nabla h(x_{n_l}), x_1^* - x_2^* \rangle\Big] \\ &= D_h(x_1^*, x_2^*), \end{aligned}$$

the last equality following from the weak continuity of ∇h.

Reversing the roles of x_1^* and x_2^*, we obtain

$$D_h(x_1^*, x_2^*) + D_h(x_2^*, x_1^*) = 0$$

and so the conclusion, by using Proposition 1.

Remark 17. Under assumptions (i) and (ii) of Theorem 16, the following assertions are equivalent:

(a) problem (P) has at least one solution;
(b) the sequence (x_n) generated by the rule (PRh) is bounded.

Remark 18. In the exact algorithm, we can weaken assumptions on ∇h, supposing it Lipschitz continuous on bounded sets only.

4. Perturbed proximal point algorithm with nonquadratic kernel. Now, we can establish convergence theorems for the sequence generated by the scheme

$$x_n \in S, \quad \left\| x_n - J_{\lambda_n}^{h,T^n} x_{n-1} \right\| \le \varepsilon_n, \quad \forall\, n \in \mathbb{N}^*,$$

where (ε_n) is a sequence of positive real numbers.

This rule takes the equivalent form

$$\text{(PPRh)} \qquad x_n \in S, \quad x_n = J_{\lambda_n}^{h,T^n} x_{n-1} + e_n, \quad \forall\, n \in \mathbb{N}^*.$$

The proofs of these theorems being rather technical, we give, in this section, their main ideas and, in the appendix, the complete one's.

Theorem 19. *Assume that problem* (P) *has at least one solution and*

(i) $0 < \dfrac{M}{2}\underline{\lambda} < \lambda^* \le \lambda_n$, $\forall\, n \in \mathbb{N}^*$,

(ii) $\sum\limits_{n=1}^{+\infty} \|e_n\| < +\infty$,

(iii) $\sum\limits_{n=1}^{+\infty} \lambda_n \delta_{\underline{\lambda},\rho}(T^n, T) < +\infty$, $\forall\, \rho \ge 0$.

Then the sequence (x_n) *generated by the rule* (PPRh) *weakly converges to some solution of* (P) *and satisfies*

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

P r o o f. In a first time, we prove that $(D_h(\overline{x}, x_n))$ is a convergent sequence and so that the sequence (x_n) is bounded.

Then, we define the auxiliary sequence

$$\begin{cases} \tilde{x}_0 & = & x_0 \\ \tilde{x}_n & = & J_{\lambda_n}^{h,T} x_{n-1} \end{cases}$$

and show that this sequence is generated by a rule (PRh) and satisfies assumptions of Theorem 16. $\square$

Remark 20. Once more, under assumptions of Theorem 19, the two following assertions are equivalent:

(a) problem (P) has at least one solution;
(b) the sequence (x_n) generated by the rule (PPRh) is bounded.

Lemma 21. *Under the assumptions of Theorem* 19, *the sequence* (x_n)

generated by the rule (PPRh) *satisfies*

$$\lim_{n\to+\infty} \left\| A^{h,T^n}_{\lambda_n} x_{n-1} \right\| = 0.$$

P r o o f. The definitions of $A^{h,T^n}_{\lambda_n}$ and (x_n) imply

$$\begin{aligned} A^{h,T^n}_{\lambda_n} x_{n-1} &= \frac{\nabla h(x_{n-1}) - \nabla h(J^{h,T^n}_{\lambda_n} x_{n-1})}{\lambda_n} \\ &= \frac{\nabla h(x_{n-1}) - \nabla h(x_n - e_n)}{\lambda_n}, \quad \forall\, n \in \mathbb{N}^*, \end{aligned}$$

and, since ∇h is Lipschitz continuous and $\lambda_n \geq \frac{M}{2}\underline{\lambda}$, $\forall\, n \in \mathbb{N}^*$,

$$\left\| A^{h,T^n}_{\lambda_n} x_{n-1} \right\| \leq \frac{2}{\underline{\lambda}}(\|x_{n-1} - x_n\| + \|e_n\|).$$

We can conclude immediately, thanks to theorem 19 and assumption (ii). □

Theorem 22. *Assume that*

(i) $0 < \frac{M}{2}\underline{\lambda} < \lambda^* \leq \lambda_n$, $\forall\, n \in \mathbb{N}^*$,

(ii) *the sequence* (x_n) *generated by the rule* (PPRh) *is bounded,*

(iii) $\|e_n\| \leq \theta_n \|x_n - x_{n-1}\|$, $\forall\, n \in \mathbb{N}^*$, *with* $\sum_{n=1}^{+\infty} \theta_n < +\infty$,

(iv) $\sum_{n=1}^{+\infty} \lambda_n \delta_{\underline{\lambda},\rho}(T^n, T) < +\infty$, $\forall\, \rho \geq 0$,

(v) *the operators* $(T^n)^{-1}$ *are uniformly locally Lipschitz continuous at* 0, *i.e. there are two constants* $a \geq 0$ *and* $\tau > 0$ *such that, for all* $n \in \mathbb{N}^*$,

$$\|w_1\|, \|w_2\| \leq \tau \Rightarrow \|z_1 - z_2\| \leq a\|w_1 - w_2\|,$$

$$\forall\, z_1 \in (T^n)^{-1}(w_1),\ \forall\, z_2 \in (T^n)^{-1}(w_2).$$

Then problem (P) *has a unique solution* $\overline{x}$ *and the sequence* (x_n) *generated by the rule* (PPRh) *strongly converges to* $\overline{x}$. *More precisely, there are two constants* $\eta \in]0, 1[$, $C > 0$ *and a range* $N \in \mathbb{N}^*$ *from which*

$$D_h(\overline{x}, x_n) + \frac{\alpha\underline{\lambda}^2}{8a^2}\|x_n - \overline{x}\|^2$$

$$\leq \eta \left[D_h(\overline{x}, x_{n-1}) + \frac{\alpha \underline{\lambda}^2}{8a^2} \|x_{n-1} - \overline{x}\|^2 \right] + C\delta^h_{\lambda_n, \|\overline{x}\|}(T^n, T). \tag{12}$$

P r o o f. The uniqueness of the solution of problem (P) results from Theorem 19 and P. Tossings [18, proof of Theorem 3.7]. The idea of the proof consists in showing that the sequence

$$D_h(\overline{x}, x_n) + \frac{\alpha \underline{\lambda}^2}{8a^2} \|x_n - \overline{x}\|^2$$

is decreasing by using, notably, the firm nonexpansivity of the mapping $J^{h,T^n}_{\lambda_n}$ for D_h, Lemma 2 and Proposition 1. □

Remark 23. The Lipschitz condition imposed in the previous theorem is satisfied if the operators $(T^n)^{-1}$ $(n \in \mathbb{N}^*)$ are uniformly globally Lipschitz continuous, what holds when the operators (T^n) $(n \in \mathbb{N}^*)$ are uniformly strongly monotone.

Remark 24. Property (12) could be rewritten more weakly (see Proposition 13): there are a range $N \in \mathbb{N}^*$ and a real $\rho^* \geq 0$ such that

$$\|x_n - \overline{x}\| \leq \eta \|x_{n-1} - \overline{x}\| + C\lambda_n \delta_{\underline{\lambda}, \rho^*}(T^n, T), \quad \forall\, n \geq N,$$

where $0 \leq \eta < 1$ and $C > 0$.

This formulation has the advantage to present the well-known variational metrics $\delta_{\underline{\lambda}, \rho^*}(T^n, T)$ with a fixed parameter $\underline{\lambda}$ in place of λ_n.

The following theorem gives us, in the nonperturbed *exact* context, a finite convergence result similar to R.T. Rockafellar's one [16].

Theorem 25. *Assume that*

(i) $0 < \dfrac{M}{2} \underline{\lambda} < \lambda^* \leq \lambda_n$, $\forall\, n \in \mathbb{N}^*$,

(ii) $\displaystyle\sum_{n=1}^{+\infty} \|e_n\| < +\infty$,

(iii) $\displaystyle\sum_{n=1}^{+\infty} \lambda_n \delta_{\underline{\lambda}, \rho}(T^n, T) < +\infty$, $\forall\, \rho \geq 0$,

(iv) *there is* $\overline{x} \in H$ *such that* $0 \in \operatorname{int} T\overline{x}$.

Then problem (P) *has a unique solution* $\overline{x}$, *the sequence* (x_n) *generated by the* *rule* (PPRh) *is bounded and there is a range* $N \in \mathbb{N}^*$ *from which*

$$\|x_n - \overline{x}\| \leq \delta^h_{\lambda_n,\rho}(T^n, T) + \|e_n\|,$$

where $\rho = \sup_{n\in\mathbb{N}} \|x_n\|$.

P r o o f. The assumptions of theorem 19 being satisfied, the sequence (x_n) is bounded and such that

$$\lim_{n\to+\infty} \|x_n - x_{n-1}\| = 0. \tag{13}$$

Set

$$\rho = \sup_{n\in\mathbb{N}} \|x_n\|$$

and define $(\tilde{x}_n)$ by

$$\begin{cases} \tilde{x}_0 &= x_0, \\ \tilde{x}_n &= J^{h,T}_{\lambda_n} x_{n-1}, \quad \forall\, n \in \mathbb{N}^*. \end{cases}$$

We get

$$\|x_n - \tilde{x}_n\| \leq \delta^h_{\lambda_n,\rho}(T^n, T) + \|e_n\|, \quad \forall\, n \in \mathbb{N}^*. \tag{14}$$

Since ∇h is Lipschitz continuous, (13) and (14) lead us to

$$\lim_{n\to+\infty} \left\| \frac{\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n)}{\lambda_n} \right\| = 0.$$

Moreover, the definition of $(\tilde{x}_n)$ implies

$$\frac{\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n)}{\lambda_n} \in T\tilde{x}_n, \quad \forall\, n \in \mathbb{N}^*,$$

and assumption (iv) ensures the existence of a neighbourhood of the origin where T^{-1} takes the unique value $\overline{x}$ (see R.T. Rockafellar [16, Theorem 3]).

So the conclusion arises. □

Remark 26. In the optimization context, an estimation of the variational metric for the most useful penalty functions, i.e. classical or exact exterior one or yet exponential one, is well known. In this context, assumption (iii) may be expressed in terms of conditions on the penalty parameters.

Remark 27. Like in Theorem 22, the thesis of Theorem 25 can be rewritten more weakly, with only the metrics $\delta_{\underline{\lambda},\rho}(T^n, T)$ $(\rho \geq 0)$, i.e. (x_n) is bounded and there are a range $N \in \mathbb{N}^*$, a real ρ^* and a constant $C > 0$ such that

$$\|x_n - \overline{x}\| \leq C\lambda_n \delta_{\underline{\lambda},\rho^*}(T^n, T) + \|e_n\|, \quad \forall\, n \geq N.$$

Like in R. T. Rockafellar's [16] and P. Alexandre's [1], we achieve with a super-linear convergence result.

Theorem 28. *Assume that*

(i) $1 \le \lambda_n$, $\forall\, n \in \mathbb{N}^*$, *with* $\lim\limits_{n\to+\infty} \lambda_n = +\infty$,

(ii) *the sequence* (x_n) *generated by the rule* (PPRh) *is bounded,*

(iii) $\|e_n\| \le \theta_n \|x_n - x_{n-1}\|$, $\forall\, n \in \mathbb{N}^*$, *with* $\sum\limits_{n=1}^{+\infty} \theta_n < +\infty$,

(iv) $\sum\limits_{n=1}^{+\infty} \lambda_n \delta_{\underline{\lambda},\rho}(T^n, T) < +\infty$, $\forall\, \rho \ge 0$,

(v) *the operators* $(T^n)^{-1}$ *are uniformly differentiable at the origin, i.e. there are a point* $\overline{x}$ *of* H, *a real* $\tau > 0$ *and a sequence of linear applications* (A_n) *such that*

$$\begin{cases} (T^n)^{-1}(0) = \{\overline{x}\}, \quad \forall\, n \in \mathbb{N}^*, \\[2ex] \|w\| \le \tau \Rightarrow \Big[(T^n)^{-1}(w) - \overline{x} - A_n w\Big] \subset o(\|w\|)B, \quad \forall\, n \in \mathbb{N}^*, \\[2ex] \sup\limits_{n\in\mathbb{N}^*} \|A_n\| < +\infty. \end{cases}$$

Then $\overline{x}$ *is the unique solution of problem* (P) *and the sequence* (x_n) *generated by the rule* (PPRh) *strongly converges to this solution.*

Furthermore, there is a sequence (η_n) *converging to zero and a range* $N \in \mathbb{N}^*$ *such that*

$$\|x_n - \overline{x}\| \le \eta_n \|x_{n-1} - \overline{x}\|, \quad \forall\, n \ge N.$$

P r o o f. We first remark that assumptions of Theorem 19 are satisfied and so we get

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

We deduce then the existence of a sequence (β_n) going to zero such that

$$\|\nabla h(\tilde{x}_n) - \nabla h(\overline{x})\| \le \beta_n \|x_{n-1} - x_n\|, \quad \forall\, n \ge N_0,$$

where the auxiliary sequence $(\tilde{x}_n)$ is defined by

$$\tilde{x}_n = J_{\lambda_n}^{h,T^n} x_{n-1}, \quad \forall\, n \in \mathbb{N}^*.$$

The conclusion follows for

$$\eta_n = \frac{M\theta_n + \beta_n}{\alpha - (M\theta_n + \beta_n)}. \qquad \square$$

A. Appendix.

Proof of Theorem 19. Let $\overline{x}$ be a solution of (P) and choose $n \in \mathbb{N}^*$.

On the one hand, two applications of Lemma 2 and the positivity of D_h lead to

$$\begin{aligned} D_h(\overline{x}, x_n) + D_h(x_n, \overline{x}) \quad \leq \quad & D_h(J_{\lambda_n}^{h,T^n}\overline{x}, J_{\lambda_n}^{h,T^n}x_{n-1}) + D_h(J_{\lambda_n}^{h,T^n}x_{n-1}, J_{\lambda_n}^{h,T^n}\overline{x}) \\ & + \Big\langle \nabla h(J_{\lambda_n}^{h,T^n}x_{n-1}) - \nabla h(\overline{x}), J_{\lambda_n}^{h,T^n}\overline{x} - \overline{x} \Big\rangle \\ & \qquad + \Big\langle \nabla h(J_{\lambda_n}^{h,T^n}x_{n-1}) - \nabla h(x_n), \overline{x} - x_n \Big\rangle \\ & + \Big\langle \nabla h(J_{\lambda_n}^{h,T^n}\overline{x}) - \nabla h(\overline{x}), J_{\lambda_n}^{h,T^n}x_{n-1} - \overline{x} \Big\rangle \\ & \qquad + \Big\langle \nabla h(\overline{x}) - \nabla h(x_n), J_{\lambda_n}^{h,T^n}x_{n-1} - x_n \Big\rangle. \end{aligned}$$

On the other hand, the firm nonexpansivity of $J_{\lambda_n}^{h,T^n}$ for D_h and Lemma 2, applied a last time to

$$D_h(J_{\lambda_n}^{h,T^n}x_{n-1}, \overline{x}) - D_h(x_n, \overline{x}) \quad \text{and} \quad D_h(J_{\lambda_n}^{h,T^n}\overline{x}, x_{n-1}) - D_h(\overline{x}, x_{n-1})$$

give

$$\begin{aligned} D_h(\overline{x}, x_n) \quad \leq \quad & D_h(\overline{x}, x_{n-1}) \\ & + \Big\langle \overline{x} - x_n, \nabla h(J_{\lambda_n}^{h,T^n}x_{n-1}) - \nabla h(x_n) \Big\rangle + \Big\langle \nabla h(\overline{x}) - \nabla h(x_n), J_{\lambda_n}^{h,T^n}x_{n-1} - x_n \Big\rangle \\ & + \Big\langle \nabla h(J_{\lambda_n}^{h,T^n}x_{n-1}) - \nabla h(x_n), J_{\lambda_n}^{h,T^n}\overline{x} - \overline{x} \Big\rangle + \Big\langle \nabla h(x_n) - \nabla h(\overline{x}), J_{\lambda_n}^{h,T^n}\overline{x} - \overline{x} \Big\rangle \\ & + \Big\langle J_{\lambda_n}^{h,T^n}x_{n-1} - x_n, \nabla h(J_{\lambda_n}^{h,T^n}\overline{x}) - \nabla h(\overline{x}) \Big\rangle + \Big\langle x_n - \overline{x}, \nabla h(J_{\lambda_n}^{h,T^n}\overline{x}) - \nabla h(\overline{x}) \Big\rangle \qquad (15) \\ & + \Big\langle \nabla h(\overline{x}) - \nabla h(x_n), x_n - J_{\lambda_n}^{h,T^n}x_{n-1} \Big\rangle \\ & + \Big\langle \nabla h(x_n) - \nabla h(J_{\lambda_n}^{h,T^n}x_{n-1}), x_n - J_{\lambda_n}^{h,T^n}x_{n-1} \Big\rangle \end{aligned}$$

$$+\Big\langle \nabla h(x_{n-1}) - \nabla h(\overline{x}), \overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\Big\rangle + \Big\langle \nabla h(\overline{x}) - \nabla h(J_{\lambda_n}^{h,T^n}\overline{x}), \overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\Big\rangle.$$

Applying now the inequality (6) to the ten inner products of (15) and taking for γ the following values

$$\gamma_n^1 = \frac{\alpha}{6}\frac{\|e_n\|}{1+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|}, \quad \gamma_n^2 = \frac{1}{M^2}\gamma_n^1,$$

$$\gamma_n^3 = \frac{1}{M^2}\frac{\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|}{\|e_n\|}, \qquad \gamma_n^4 = \frac{\alpha}{4M^2}\frac{\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|}{1+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|},$$

$$\gamma_n^5 = M^2\gamma_n^3, \qquad \gamma_n^6 = M^2\gamma_n^4,$$

$$\gamma_n^7 = \gamma_n^2, \qquad \gamma_n^8 = \|e_n\|,$$

$$\gamma_n^9 = \frac{\alpha}{2M^2}(\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|), \quad \gamma_n^{10} = M^2\|e_n\|\gamma_n^3,$$

we get, thanks to the Lipschitz continuity of ∇h and Proposition 1,

$$\begin{aligned}\left[1-\frac{1}{\alpha}\Big[3\gamma_n^1+2\gamma_n^6\Big]\right] D_h(\overline{x},x_n) \quad &\le \quad \left[1+\frac{M^2\gamma_n^9}{\alpha}\right] D_h(\overline{x},x_{n-1}) \\ &+\frac{1}{2}\left[\frac{3}{\gamma_n^2}+2\gamma_n^5+M^2\gamma_n^8+\frac{1}{\gamma_n^8}\right]\|e_n\|^2 \\ &+\frac{1}{2}\left[\frac{1}{\gamma_n^3}+\frac{2}{\gamma_n^4}+\frac{M^2}{\gamma_n^5}+\frac{1}{\gamma_n^9}+M^2\gamma_n^{10}+\frac{1}{\gamma_n^{10}}\right]\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|^2.\end{aligned} \tag{16}$$

The coefficients of $D_h(\overline{x},x_n)$ and $D_h(\overline{x},x_{n-1})$ are respectively equal to

$$\frac{2+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|}{2(1+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|)} \quad \text{and} \quad \frac{1}{2}\Big[2+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|\Big],$$

and so, formula (16) can be rewritten as

$$\begin{aligned}D_h(\overline{x},x_n) \quad &\le \quad \Big[1+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|\Big] D_h(\overline{x},x_{n-1}) \\ &+\frac{1+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|}{2+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|}\left[\frac{3}{\gamma_n^2}+2\gamma_n^5+M^2\gamma_n^8+\frac{1}{\gamma_n^8}\right]\|e_n\|^2 \\ &+\frac{1+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|}{2+\|e_n\|+\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\|}\left[\frac{1}{\gamma_n^3}+\frac{2}{\gamma_n^4}+\frac{M^2}{\gamma_n^5}\right.\end{aligned} \tag{17}$$

$$+\frac{1}{\gamma_n^9} + M^2\gamma_n^{10} + \frac{1}{\gamma_n^{10}}\bigg] \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|^2.$$

Now, there is a range $N \in \mathbb{N}^*$ from which

$$\|e_n\| \leq 1, \quad \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\| \leq 1 \quad \text{ and } \quad \frac{1 + \|e_n\| + \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|}{2 + \|e_n\| + \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|} < 1$$

and two constants C_1 and $C_2 > 0$, depending only on α and M, such that

$$\frac{1 + \|e_n\| + \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|}{2 + \|e_n\| + \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|}\left[\frac{3}{\gamma_n^2} + 2\gamma_n^5 + M^2\gamma_n^8 + \frac{1}{\gamma_n^8}\right]\|e_n\|^2 < C_1\|e_n\|,$$

and

$$\frac{1 + \|e_n\| + \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|}{2 + \|e_n\| + \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|}\bigg[\frac{1}{\gamma_n^3} + \frac{2}{\gamma_n^4} + \frac{M^2}{\gamma_n^5} + \frac{1}{\gamma_n^9}$$
$$+M^2\gamma_n^{10} + \frac{1}{\gamma_n^{10}}\bigg] \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|^2 < C_2\|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|,$$

for all $n \geq N$.

That leads finally, by using Proposition 13, to rewrite formula (17) in the form

$$D_h(\overline{x}, x_n) \leq \left[1 + \|e_n\| + \|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\|\right] D_h(\overline{x}, x_{n-1}) + C_1\|e_n\|$$
$$+C_2'\lambda_n\delta_{\underline{\lambda},\rho_0}(T^n, T), \quad \forall\, n \geq N,$$

where $C_2' > 0$ and

$$\rho_0 \geq \left[5 + 6\frac{M}{\alpha}\right]\|\overline{x}\|.$$

So, taking into account assumptions and Lemma 15, the sequence $(D_h(\overline{x}, x_n))$ appears to be convergent.

Therefore, the sequence (x_n) is bounded and we can write

$$\rho = \sup_{n\in\mathbb{N}} \|x_n\|.$$

Let us define the auxiliary sequence

$$\begin{cases} \tilde{x}_0 &= x_0, \\ \tilde{x}_n &= J_{\lambda_n}^{h,T} x_{n-1}, \quad \forall\, n \in \mathbb{N}^*. \end{cases}$$

We have, on the one hand,

$$\|x_n - \tilde{x}_n\| \leq \delta_{\lambda_n,\rho}^h(T^n, T) + \|e_n\|, \quad \forall\, n \in \mathbb{N}, \tag{18}$$

and, on the other hand,

$$\tilde{x}_n = J_{\lambda_n}^{h,T}\tilde{x}_{n-1} + \tilde{e}_n, \quad \forall\, n \in \mathbb{N}^*,$$

where, from Proposition 8,

$$\|\tilde{e}_n\| = \|J_{\lambda_n}^{h,T}x_{n-1} - J_{\lambda_n}^{h,T}\tilde{x}_{n-1}\| \le \frac{M}{\alpha}\|x_{n-1} - \tilde{x}_{n-1}\|, \quad \forall\, n \in \mathbb{N}^*. \tag{19}$$

Inequalities (18) and (19) and Proposition 13 imply that the sequence $(\tilde{x}_n)$ is generated by a rule (PRh) and satisfies assumptions of Theorem 16: it weakly converges to a solution of (P) with

$$\lim_{n\to+\infty} \|\tilde{x}_n - \tilde{x}_{n-1}\| = 0.$$

The conclusion follows immediately thanks to assumptions (ii), (iii) and inequality (18). $\square$

P r o o f o f T h e o r e m 22. The uniqueness of the solution of problem (P) results from Theorem 19 and P. Tossings [18, proof of Theorem 3.7].

Let $\overline{x}$ be this unique solution. Proposition 14 and Lemma 21 ensure the existence of a range $N_1 \in \mathbb{N}^*$ from which

$$\left\|A_{\lambda_n}^{h,T^n}\overline{x}\right\| < \tau \quad \text{and} \quad \left\|A_{\lambda_n}^{h,T^n}x_{n-1}\right\| < \tau.$$

Since

$$A_{\lambda_n}^{h,T^n}\overline{x} \in T^n\Big(J_{\lambda_n}^{h,T^n}\overline{x}\Big) \quad \text{and} \quad A_{\lambda_n}^{h,T^n}x_{n-1} \in T^n\Big(J_{\lambda_n}^{h,T^n}x_{n-1}\Big),$$

assumption (v) and Definition 9 involve

$$\left\|J_{\lambda_n}^{h,T^n}\overline{x} - J_{\lambda_n}^{h,T^n}x_{n-1}\right\|^2 \le \frac{2a^2M^2}{\lambda_n^2}\left(\left\|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\right\|^2 + \left\|x_{n-1} - J_{\lambda_n}^{h,T^n}x_{n-1}\right\|^2\right),$$

for all $n \ge N_1$, and finally, using assumption (i),

$$\frac{\alpha}{2}\frac{\underline{\lambda}^2}{8a^2}\left\|J_{\lambda_n}^{h,T^n}\overline{x} - J_{\lambda_n}^{h,T^n}x_{n-1}\right\|^2 \le \frac{\alpha}{2}\left\|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\right\|^2 + \frac{\alpha}{2}\left\|x_{n-1} - J_{\lambda_n}^{h,T^n}x_{n-1}\right\|^2. \tag{20}$$

for all $n \ge N_1$.

Moreover, Proposition 1 and the firm nonexpansivity of $J_{\lambda_n}^{h,T^n}$ for D_h lead us to

$$\begin{aligned} &D_h(J_{\lambda_n}^{h,T^n}x_{n-1}, J_{\lambda_n}^{h,T^n}\overline{x}) + D_h(J_{\lambda_n}^{h,T^n}\overline{x}, J_{\lambda_n}^{h,T^n}x_{n-1}) \le D_h(J_{\lambda_n}^{h,T^n}x_{n-1}, \overline{x}) \\ &+D_h(J_{\lambda_n}^{h,T^n}\overline{x}, x_{n-1}) - \frac{\alpha}{2}\left\|J_{\lambda_n}^{h,T^n}x_{n-1} - x_{n-1}\right\|^2 - \frac{\alpha}{2}\left\|J_{\lambda_n}^{h,T^n}\overline{x} - \overline{x}\right\|^2. \end{aligned} \tag{21}$$

Working as in the proof of Theorem 19 for obtaining inequality (15), we get

$$\begin{aligned}
D_h(\overline{x}, x_n) \quad \leq \quad & D_h(\overline{x}, x_{n-1}) \\
& + \left\langle \overline{x} - x_n, \nabla h(J^{h,T^n}_{\lambda_n} x_{n-1}) - \nabla h(x_n) \right\rangle + \left\langle \nabla h(\overline{x}) - \nabla h(x_n), J^{h,T^n}_{\lambda_n} x_{n-1} - x_n \right\rangle \\
& + \left\langle \nabla h(J^{h,T^n}_{\lambda_n} x_{n-1}) - \nabla h(x_n), J^{h,T^n}_{\lambda_n} \overline{x} - \overline{x} \right\rangle + \left\langle \nabla h(x_n) - \nabla h(\overline{x}), J^{h,T^n}_{\lambda_n} \overline{x} - \overline{x} \right\rangle \\
& + \left\langle J^{h,T^n}_{\lambda_n} x_{n-1} - x_n, \nabla h(J^{h,T^n}_{\lambda_n} \overline{x}) - \nabla h(\overline{x}) \right\rangle + \left\langle x_n - \overline{x}, \nabla h(J^{h,T^n}_{\lambda_n} \overline{x}) - \nabla h(\overline{x}) \right\rangle \\
& + \left\langle \nabla h(\overline{x}) - \nabla h(x_n), x_n - J^{h,T^n}_{\lambda_n} x_{n-1} \right\rangle \\
& + \left\langle \nabla h(x_n) - \nabla h(J^{h,T^n}_{\lambda_n} x_{n-1}), x_n - J^{h,T^n}_{\lambda_n} x_{n-1} \right\rangle \\
& + \left\langle \nabla h(x_{n-1}) - \nabla h(\overline{x}), \overline{x} - J^{h,T^n}_{\lambda_n} \overline{x} \right\rangle + \left\langle \nabla h(\overline{x}) - \nabla h(J^{h,T^n}_{\lambda_n} \overline{x}), \overline{x} - J^{h,T^n}_{\lambda_n} \overline{x} \right\rangle \\
(22) & - \frac{\alpha}{2} \left\| J^{h,T^n}_{\lambda_n} x_{n-1} - x_{n-1} \right\|^2 - \frac{\alpha}{2} \left\| J^{h,T^n}_{\lambda_n} \overline{x} - \overline{x} \right\|^2 .
\end{aligned}$$

Using the well-known formula

$$\|x - b\|^2 - \|x - a\|^2 + \|b - a\|^2 = 2\langle a - b, x - b \rangle, \quad \forall\, x, a, b \in H,$$

inequalities (21) and (22) involve

$$\begin{aligned}
\Phi_{\overline{x}}(x_n) \quad \leq \quad & \Phi_{\overline{x}}(x_{n-1}) - \frac{\alpha \underline{\lambda}^2}{8a^2} \|x_{n-1} - \overline{x}\|^2 \\
& + \left\langle \overline{x} - x_n, \nabla h(J^{h,T^n}_{\lambda_n} x_{n-1}) - \nabla h(x_n) \right\rangle + \left\langle \nabla h(\overline{x}) - \nabla h(x_n), J^{h,T^n}_{\lambda_n} x_{n-1} - x_n \right\rangle \\
& + \left\langle \nabla h(J^{h,T^n}_{\lambda_n} x_{n-1}) - \nabla h(x_n), J^{h,T^n}_{\lambda_n} \overline{x} - \overline{x} \right\rangle + \left\langle \nabla h(x_n) - \nabla h(\overline{x}), J^{h,T^n}_{\lambda_n} \overline{x} - \overline{x} \right\rangle \\
& + \left\langle \nabla h(J^{h,T^n}_{\lambda_n} \overline{x}) - \nabla h(\overline{x}), J^{h,T^n}_{\lambda_n} x_{n-1} - x_n \right\rangle + \left\langle x_n - \overline{x}, \nabla h(J^{h,T^n}_{\lambda_n} \overline{x}) - \nabla h(\overline{x}) \right\rangle \\
& + \left\langle \nabla h(x_{n-1}) - \nabla h(\overline{x}), \overline{x} - J^{h,T^n}_{\lambda_n} \overline{x} \right\rangle + \left\langle \nabla h(\overline{x}) - \nabla h(J^{h,T^n}_{\lambda_n} \overline{x}), \overline{x} - J^{h,T^n}_{\lambda_n} \overline{x} \right\rangle \\
& + \left\langle \nabla h(\overline{x}) - \nabla h(x_n), x_n - J^{h,T^n}_{\lambda_n} x_{n-1} \right\rangle
\end{aligned}$$

$$+\left\langle \nabla h(x_n) - \nabla h(J_{\lambda_n}^{h,T^n} x_{n-1}), x_n - J_{\lambda_n}^{h,T^n} x_{n-1} \right\rangle$$

$$+\frac{\alpha\underline{\lambda}^2}{4a^2}\left\langle \overline{x} - x_n, J_{\lambda_n}^{h,T^n} x_{n-1} - x_n \right\rangle + \frac{\alpha\underline{\lambda}^2}{4a^2}\left\langle \overline{x} - x_n, \overline{x} - J_{\lambda_n}^{h,T^n} \overline{x} \right\rangle$$

$$(23)\ +\frac{\alpha\underline{\lambda}^2}{4a^2}\left\langle J_{\lambda_n}^{h,T^n} x_{n-1} - x_n, J_{\lambda_n}^{h,T^n} \overline{x} - \overline{x} \right\rangle,$$

where the auxiliary function $\Phi_{\overline{x}}$ is defined by

$$\Phi_{\overline{x}}(x) = D_h(\overline{x}, x) + \frac{\alpha\underline{\lambda}^2}{8a^2}\|x - \overline{x}\|^2, \quad \forall\, x \in S.$$

Applying the inequality (6) to the thirteen inner products of (23) and taking for γ the following values

$$\gamma_n^1 = \frac{\theta_n}{2}, \qquad \gamma_n^2 = \frac{1}{M^2}\gamma_n^1,$$

$$\gamma_n^3 = \frac{1}{M^2}\frac{\left\|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\right\|}{\theta_n}, \qquad \gamma_n^4 = \frac{2\theta_n}{3}\gamma_n^3,$$

$$\gamma_n^5 = \frac{1}{M^2}\frac{\theta_n}{\left\|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\right\|}, \qquad \gamma_n^6 = M^2\gamma_n^4,$$

$$\gamma_n^7 = 2\theta_n\gamma_n^3, \qquad \gamma_n^8 = \left\|\overline{x} - J_{\lambda_n}^{h,T^n}\overline{x}\right\|,$$

$$\gamma_n^9 = \frac{1}{M^2}\gamma_n^1, \qquad \gamma_n^{10} = \theta_n,$$

$$\gamma_n^{11} = \frac{4a^2}{\alpha\underline{\lambda}^2}\gamma_n^1, \qquad \gamma_n^{12} = \frac{4a^2M^2}{\alpha\underline{\lambda}^2}\gamma_n^4,$$

$$\gamma_n^{13} = \gamma_n^3,$$

we get, thanks to the Lipschitz continuity of ∇h, Proposition 1 and relations between γ_n,

$$\Phi_{\overline{x}}(x_n) \;\le\; \Phi_{\overline{x}}(x_{n-1}) - \frac{\alpha\underline{\lambda}^2}{8a^2}\|x_{n-1} - \overline{x}\|^2$$

$$(24)\quad +\frac{M^2\gamma_n^7}{2}\|x_{n-1}-\overline{x}\|^2+\left[2\gamma_n^1+\theta_n\gamma_n^3\right]\|x_n-\overline{x}\|^2$$

$$+\left[\frac{3M^2}{2\gamma_n^1}+\frac{\alpha^2\underline{\lambda}^4}{32a^4\gamma_n^1}+\frac{M^2\gamma_n^3}{2}+\frac{\alpha\underline{\lambda}^2\gamma_n^3}{8a^2}+\frac{1}{2\gamma_n^5}+\frac{M^2\gamma_n^{10}}{2}+\frac{1}{2\gamma_n^{10}}\right]\|e_n\|^2$$

$$+\left[\frac{1}{2\gamma_n^3}+\frac{\alpha\underline{\lambda}^2}{8a^2\gamma_n^3}+\frac{7M^2}{4\theta_n\gamma_n^3}+\frac{3\alpha^2\underline{\lambda}^4}{64a^4M^2\theta_n\gamma_n^3}\right.$$

$$\left.+\frac{M^2\gamma_n^5}{2}+\frac{M^2\gamma_n^8}{2}+\frac{1}{2\gamma_n^8}\right]\left\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\right\|^2.$$

Now, there is a range $N_2\geq N_1$ from which

$$0\leq\theta_n\leq 1\quad\text{ and }\quad 0\leq\left\|J_{\lambda_n}^{h,T^n}\overline{x}-\overline{x}\right\|\leq 1$$

and so, the explicit values of constants γ_n and assumption (iii) ensure the existence of three constants C_1, C_2 and $C_3>0$ such that

$$\Phi_{\overline{x}}(x_n)\ \leq\ \Phi_{\overline{x}}(x_{n-1})-\frac{\alpha\underline{\lambda}^2}{8a^2}\|x_{n-1}-\overline{x}\|^2+\left[C_2\theta_n+\left\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\right\|\right]\|x_n-\overline{x}\|^2$$

$$(25)+\left[C_1\theta_n+\left\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\right\|\right]\|x_{n-1}-\overline{x}\|^2+C_3\left\|\overline{x}-J_{\lambda_n}^{h,T^n}\overline{x}\right\|,\quad\forall\, n\geq N_2.$$

Moreover, assumptions (ii) and (iv) ensure the existence of a range $N_3\geq N_2$ such that

$$0\leq\theta_n<\frac{\alpha}{8C_2}\quad\text{ et }\quad 0\leq\left\|J_{\lambda_n}^{h,T^n}\overline{x}-\overline{x}\right\|<\frac{\alpha}{8},$$

and so, Proposition 1 and the definition of function $\Phi_{\overline{x}}$ lead us to

$$\Phi_{\overline{x}}(x_n)\ \leq\ \left[1+C_\theta\theta_n+C_J\left\|J_{\lambda_n}^{h,T^n}\overline{x}-\overline{x}\right\|\right]\Phi_{\overline{x}}(x_{n-1})$$

$$-\frac{\alpha\underline{\lambda}^2}{8a^2}\|x_{n-1}-\overline{x}\|^2+2C_3\left\|J_{\lambda_n}^{h,T^n}\overline{x}-\overline{x}\right\|,\quad\forall\, n\geq N_3,$$

where C_θ and C_J are two positive constants.

On the one hand, Proposition 1 and the definition of function $\Phi_{\overline{x}}$ imply

$$(26)\qquad \Phi_{\overline{x}}(x_{n-1})\leq\left[\frac{M}{2}+\frac{\alpha\underline{\lambda}^2}{8a^2}\right]\|\overline{x}-x_{n-1}\|^2.$$

On the other hand, choosing finally a range $N_4\geq N_3$ from which

$$0\leq\theta_n<\left[\frac{M}{2}+\frac{\alpha\underline{\lambda}^2}{8a^2}\right]^{-1}\frac{1}{C_\theta}\frac{\alpha\underline{\lambda}^2}{32a^2}$$

and

$$0 \leq \left\| J_{\lambda_n}^{h,T^n} \overline{x} - \overline{x} \right\| < \left[\frac{M}{2} + \frac{\alpha \underline{\lambda}^2}{8a^2} \right]^{-1} \frac{1}{C_J} \frac{\alpha \underline{\lambda}^2}{32a^2},$$

it follows that

$$\Phi_{\overline{x}}(x_n) \leq \Phi_{\overline{x}}(x_{n-1}) - \frac{\alpha \underline{\lambda}^2}{8a^2} \|x_{n-1} - \overline{x}\|^2 + 2C_3 \left\| J_{\lambda_n}^{h,T^n} \overline{x} - \overline{x} \right\|$$

and so the announced result using inequality (26), assumption (iv) and Proposition 13.

This establishes the announced result. □

P r o o f o f t h e o r e m 28. First, remark that assumption (v) and the Lipschitz continuity of ∇h ensure the existence of a constant $C > 0$ such that

$$\left\| \nabla h\Big((T^n)^{-1}(w) \Big) - \nabla h(\overline{x}) - A_n w \right\| \leq C \|w\|, \quad \forall\, w \in H, \|w\| \leq \tau.$$

Then, note that assumption (iv) involving the graph-convergence of (T^n) to T, the hypothesis $(T^n)^{-1}(0) = \{\overline{x}\}$, $\forall\, n \in \mathbb{N}^*$, leads to $0 \in T\overline{x}$.

Applying Theorem 19, we get therefore

$$\lim_{n \to +\infty} \|x_n - x_{n-1}\| = 0. \tag{27}$$

Now, define the auxiliary sequence

$$\tilde{x}_n = J_{\lambda_n}^{h,T^n} x_{n-1}, \quad \forall\, n \in \mathbb{N}^*.$$

Proposition 10 (ii) implies

$$\nabla h(\tilde{x}_n) \in \left(\nabla h (T^n)^{-1} \frac{1}{\lambda_n} \right) \Big[\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n) \Big], \quad \forall\, n \in \mathbb{N}^*,$$

and, consequently, the mappings A_n being linear,

$$\begin{aligned} \nabla h(\tilde{x}_n) - \nabla h(\overline{x}) \in & \Big(\nabla h (T^n)^{-1} \frac{1}{\lambda_n} \Big) \Big[\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n) \Big] \\ & - \nabla h(\overline{x}) - A_n \frac{1}{\lambda_n} \Big[\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n) \Big] \\ & + \frac{1}{\lambda_n} A_n \Big[\nabla h(x_{n-1}) - \nabla h(x_n) \Big] + \frac{1}{\lambda_n} A_n \Big[\nabla h(x_n) - \nabla h(\tilde{x}_n) \Big]. \end{aligned}$$

One the one hand, assumption (iii), relation (27) and the Lipschitz continuity of ∇h ensure the existence of a range $N_0 \in \mathbb{N}^*$ from which

$$\theta_n \leq 1, \quad \|x_n - x_{n-1}\| \leq \frac{\tau}{2M}$$

and

$$\|\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n)\| \leq M(1+\theta_n)\|x_n - x_{n-1}\|,$$

giving in turn

$$\left\|\frac{1}{\lambda_n}\Big(\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n)\Big)\right\| \leq \|\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n)\|, \quad \forall\, n \in \mathbb{N}^*.$$

It results from assumption (v) that

$$\begin{aligned}&\Big[\nabla h(T^n)^{-1}\Big[\frac{1}{\lambda_n}\Big(\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n)\Big)\Big] - \nabla h(\overline{x}) - A_n\Big[\frac{1}{\lambda_n}\Big(\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n)\Big)\Big]\Big]\\ &\quad\subset o\left(\frac{1}{\lambda_n}\|\nabla h(x_{n-1}) - \nabla h(\tilde{x}_n)\|\right) B\\ &\quad\subset o\left(\|x_n - x_{n-1}\|\right) B, \quad \forall\, n \geq N_0.\end{aligned}$$

On the other hand, set

$$\nu = \sup_{n\in\mathbb{N}^*} \|A_n\| \qquad \text{and} \qquad \varepsilon_n = \frac{\nu}{\lambda_n} M, \quad \forall\, n \in \mathbb{N}^*.$$

Obviously ε_n goes to zero and

$$\left\|\frac{1}{\lambda_n} A_n\Big[\nabla h(x_{n-1}) - \nabla h(x_n)\Big]\right\| \leq \varepsilon_n \|x_{n-1} - x_n\|, \quad \forall\, n \in \mathbb{N}^*,$$

$$\left\|\frac{1}{\lambda_n} A_n\Big[\nabla h(x_n) - \nabla h(\tilde{x}_n)\Big]\right\| \leq \varepsilon_n \theta_n \|x_{n-1} - x_n\|, \quad \forall\, n \in \mathbb{N}^*,$$

what implies

$$\nabla h(\tilde{x}_n) - \nabla h(\overline{x}) \in o(\|x_{n-1} - x_n\| + 2\varepsilon_n\|x_{n-1} - x_n\|)B.$$

In other words, there is a sequence (β_n) going to zero for which

$$\|\nabla h(\tilde{x}_n) - \nabla h(\overline{x})\| \leq \beta_n \|x_{n-1} - x_n\|, \quad \forall\, n \geq N_0.$$

Finally, since ∇h is strongly monotone, we get

$$\begin{aligned}\alpha\|x_n - \overline{x}\| &\leq \|\nabla h(x_n) - \nabla h(\overline{x})\|\\ &\leq M\theta_n\|x_{n-1} - x_n\| + \beta_n\|x_{n-1} - x_n\|\\ &\leq (M\theta_n + \beta_n)(\|x_{n-1} - \overline{x}\| + \|\overline{x} - x_n\|),\end{aligned}$$

where

$$\lim_{n\to+\infty}(M\theta_n + \beta_n) = 0$$

and the conclusion arises for

$$\eta_n = \frac{M\theta_n + \beta_n}{\alpha - (M\theta_n + \beta_n)}. \qquad \square$$

REFERENCES

[1] P. Alexandre, V. H. Nguyen, P. Tossings. The perturbed Generalized Proximal Point Algorithm. *Math. Model. Numer. Anal.* **32**, *2* (1998), 223–253.

[2] H. Attouch, R. J. B. Wets. Isometries for Legendre-Fenchel Transform. Technical Report, Trans. A. M. S., vol. **296**, *1*, 1986, 33–60.

[3] L. M. Bregman. The Relaxation Method of Finding the Common Point of Convex Sets and its Application to the Solution of Problems in Convex Programming. *Comput. Math. Math. Phys.* **7**, *3* (1967), 200–217.

[4] H. Brezis. Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert. North-Holland, Math. Studies, vol. **5**, Amsterdam, 1973.

[5] Y. Censor, A. Lent. An iterative Row-Action Method for Interval Convex Programming. *J. Optim. Theory Appl.* **34**, *3* (1981), 321–353.

[6] G. Cohen. Auxiliary Problem Principle and Decomposition of Optimization Problems. *J. Optim. Theory Appl.* **32** (1980), 277–305.

[7] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Math. Oper. Res.* **18**, *1* (1993), 202–226.

[8] J. Eckstein. Approximate iterations in Bregman-function-based proximal algorithms. *Math. Program.* **83** (1998), 113–123.

[9] S. Kabbadj. Méthodes proximales entropiques. Thèse, Université de Montpellier II, 1994.

[10] B. Lemaire. Coupling Optimization Methods and Variational Convergence. in: Trends in Mathematical Optimization International Series of Num. Math. (Eds. K. H. Hoffmann, J. B. Hiriart-Urruty, C. Lemarechal, J. Zowe), Birkhäuser Verlag, Basel, vol. **84**, 1988, 163–179.

[11] B. Lemaire. On the convergence of some Iterative Methods for convex Minimization. In: Recent Developments in Optimization (Eds. R. Durier, C. Michelot) Lecture Notes in Econom. and Math. Systems, Springer-Verlag, 1995, 252–268.

[12] B. Lemaire. Stability of the Iteration Method for non expansive Mappings. *Serdica Math. J.* **22**, *3* (1996), 341–358.

[13] B. Polyak. Introduction to Optimization, Translations Series in Mathematics and Engineering, Optimization Software, Inc., Publications Division, New York, 1987.

[14] A. Renaud, G. Cohen. An Extension of the Auxiliary Problem Principle to Nonsymmetric Auxiliary Operators. *ESAIM Control Optim. Calc. Var.* **2** (1997), 281–306.

[15] R. T. Rockafellar. Convex Analysis. Univ. Press, Princeton, New-Jersey, 1970.

[16] R. T. Rockafellar. Monotone Operators and the Proximal Point Algorithm. *SIAM J. Control Optim.* **14**, *5* (1976), 877–898.

[17] P. Tossings. Convergence Variationnelle et Opérateurs Maximaux Monotones d'un Espace de Hilbert réel. In: Bulletin de la Société Royale des Sciences de Liège, 60e année, vol. **2–3**, 1991, 103–132.

[18] P. Tossings. The perturbed Proximal Point Algorithm and some of its Applications. *Appl. Math. Optim.* **29** (1994), 125–159.

Institut de Mathématique (B37)
Grande Traverse, 12
4000 Liège
Belgium
e-mail: Patricia.Tossings@ulg.ac.be
e-mail: M.Brohe@ulg.ac.be

Received July 20, 1999
Revised May 23, 2000