Serdica

Mathematical Journal

Сердика

Математическо списание

Serdica Math. J. **22** (1996), 331-340

Serdica
Mathematical Journal

Institute of Mathematics
and Informatics, Bulgarian
Academy of Sciences

STABILITY OF THE ITERATION METHOD FOR NON EXPANSIVE MAPPINGS

B. Lemaire

Communicated by A. L. Dontchev

Abstract. The general iteration method for nonexpansive mappings on a Banach space is considered. Under some assumption of fast enough convergence on the sequence of ("almost" nonexpansive) perturbed iteration mappings, if the basic method is $\tau-$convergent for a suitable topology τ weaker than the norm topology, then the perturbed method is also $\tau-$convergent. Application is presented to the gradient-prox method for monotone inclusions in Hilbert spaces.

1. Introduction. Let us consider some problem defined by its data d and its solution set S assumed to be a subset of a given set X. An iterative method for solving this problem, i. e. for finding an element of S, generates a sequence in X by some iteration scheme from a given starting point x in X. We call such an iterative method **basic method** if its iteration mapping from X into X is defined from the data d. With such a basic method can be associated another iterative method that we call **perturbed method**. This perturbed method is defined in the same way than the basic method except that at each iteration k, in the iteration mapping, the exact data d are replaced by perturbed (or approximate) data d_k.

If X is equipped with the topology τ, we say that an iterative method is $\tau-$convergent if, for any starting point, the generated sequence has a $\tau-$limit which is in S.

1991 *Mathematics Subject Classification:* 65K10, 49M07, 90C25

Key words: convex minimization, convergence, iteration method, gradient method, monotone inclusions, prox method, stability

We say that a basic method is $\tau-$stable with respect to the perturbed data $\{d_k\}$ if it is $\tau-$convergent as also the perturbed method with $\{d_k\}$.

In the last seven years, a number of works have been devoted to this kind of stability in the field of convex minimization and monotone inclusions in Hilbert spaces for specific basic methods known to be convergent (prox and gradient methods) ([6, 7, 9, 13, 14, 16]). The step was to prove the convergence of the perturbed method by the same techniques used for the basic method assuming that the perturbed data converge to the exact ones fast enough in a suitable sense allowing to recover the convergence of the basic method as a particular case.

What we would like to present here is a general converse result saying, roughly speaking, that, also under some assumption of fast enough convergence on the perturbed data, if a given basic method is convergent then the perturbed method is also convergent for the same topology.

We present the main results in section 2 and, in section 3, we give an application to the gradient-prox method for monotone inclusions in Hilbert spaces.

2. Main results. Considering some problem with data d and solution set S, a subset of some set X, we call **basic** method any iteration scheme

$$\begin{cases} \xi_n & = & P_n \, \xi_{n-1}, \quad n = 1, 2, \ldots \\ \xi_0 & = & x \in X \end{cases}$$

where the iteration mapping $P_n := P(d, \lambda_n)$, from X into X, is defined from the data d and may depend also on the iteration index n through a sequence of parameters $\{\lambda_n\}$ (for instance, in a descent method for a minimization problem, the sequence of step lengths).

With such a basic method we associate the **perturbed** method, i. e. the iteration scheme

$$\begin{cases} x_k & = & Q_k \, x_{k-1}, \quad k = 1, 2, \ldots \\ x_0 & = & x \in X \end{cases}$$

where the iteration mapping $Q_k := P(d_k, \lambda_k)$, from X into X, is defined in the same way than for the basic method except that, at each iteration k, the exact data d are replaced by perturbed (or approximate) data d_k.

Now on, let us assume that X is equipped with some topology τ.

Definition 2.1. *An iterative method is said to be* τ-**convergent** *if, for any starting point, the generated sequence has a* τ*-limit which is in* S.

Definition 2.2. *A basic method is said to be* τ-**stable** *with respect to the perturbed data* $\{d_k\}$*, if it is* τ*-convergent as also the perturbed method with* $\{d_k\}$.

Definition 2.3. *Let some basic method with iteration mappings P_n be given. For all $k \in \mathbb{N}$, the* **translated** *basic method is defined by the iteration scheme*

$$\begin{cases} \xi_n(k) &= P_{k+n}\,\xi_{n-1}(k), \quad n = 1, 2, \ldots \\ \xi_0(k) &= x \in X. \end{cases}$$

Remark 2.1. If $P_n \equiv P$ does not depend on n, then, for each k, the translated basic method coincides with the basic one which is therefore the iteration method for P.

Definition 2.4. *Let us assume that X is a vector space. Let some basic method with iteration mappings P_k be given as also a sequence of "errors" e_k. The* **approximate** *basic method is defined by the iteration scheme*

$$\begin{cases} x_k &= P_k\,x_{k-1} + e_k, \quad k = 1, 2, \ldots \\ x_0 &= x \in X. \end{cases}$$

The following lemma is substantially in [5] (proof of Remark 14) in a particular context and not considered on its own.

Lemma 2.1. *Let us consider some basic method with iteration mappings P_n and the associated approximate basic method with errors e_k.*

Let us assume that X is a Banach space with norm $\|.\|$ and topological dual $X^\star$, that τ is a Hausdorff locally convex topology on X compatible with the duality $X, X^\star$, that S is a closed subset of X, that, for all $n \in \mathbb{N}$, $P_n : X \to X$ is nonexpansive, i. e.

$$\forall x, y \in X, \quad \|P_n x - P_n y\| \leq \|x - y\|,$$

and that $\sum_{k=1}^{+\infty} \epsilon_k < +\infty$ with $\epsilon_k := \|e_k\|$.

If, for all $k \in \mathbb{N}$, the translated basic method is τ-convergent, then the approximate basic method is τ-convergent too.

Proof. From [5].

Let $\{x_k\}$ be a sequence generated by the approximate basic method and let $\{\xi_n(k)\}$ be the sequence generated by the translated basic method with $\xi_0(k) := x_k$. By assumption, $\forall k, \xi(k) := \tau - \lim_{n \to +\infty} \xi_n(k)$ exists and is in S.

We have

(1) $$\forall n \geq 0, \quad \forall k > 0, \|\xi_n(k) - \xi_{n+1}(k-1)\| \leq \epsilon_k$$

This results from the nonexpansiveness of the $P_n's$. Indeed we have

$$\|P_{k+n}\xi_{n-1}(k) - P_{k-1+n+1}\xi_n(k-1)\| \leq \|\xi_{n-1}(k) - \xi_n(k-1)\|.$$

Therefore,

$$\|\xi_n(k) - \xi_{n+1}(k-1)\| \le \|\xi_0(k) - \xi_1(k-1)\| = \|x_k - P_k x_{k-1}\| \le \epsilon_k.$$

Passing to the lower limit as $n \to +\infty$ in (1), thanks to the τ-lower semi-continuity of the norm, we get that $\{\xi(k)\}$ is a Cauchy sequence then norm convergent and, as S is closed, its limit x_∞ is in S.

Writing (1) successively with $n := 0$ and k, with $n := 1$ and $k-1$, and so on until n and $k-n$, for $k > n$, and adding, we get

$$\|x_k - \xi_{n+1}(k-n-1)\| \le \epsilon_k + \cdots + \epsilon_{k-n}$$

Changing k into $k+n$, we get

$$(2) \qquad \forall n \ge 0, \; \|x_{k+n} - \xi_{n+1}(k-1)\| \le \epsilon_{k+n} + \cdots + \epsilon_k \le \sum_{i=k}^{+\infty} \epsilon_i.$$

Now split the difference $x_{k+n} - x_\infty$ in three parts:

$$d_1 := x_{k+n} - \xi_{n+1}(k-1), d_2 := \xi_{n+1}(k-1) - \xi(k-1), d_3 := \xi(k-1) - x_\infty$$

Let V be a neighbourhood of the origin for the topology τ and therefore for the norm topology. Thanks to (2) and the convergence of $\xi(k)$ to x_∞ we have $d_1 + d_3 \in V/2$ for all n and some $k := K$ large enough. Then, by definition of $\xi(K-1)$, there exists N such that, for all $n \ge N$, $d_2 \in V/2$. Finally, for all $k \ge K+N$, $x_k - x_\infty \in V$, that is, x_k $\tau-$converges to x_∞. □

Remark 2.2. In the assumptions of lemma 2.1, the convergence of the translated basic method may seem more restrictive than the convergence of the sole basic method. Actually, these are equivalent if the iteration mapping does not depend on the iteration index (see Remark 2.1) or (as usual in the applications) if the dependance is through a sequence of parameters and if the property needed on this sequence for the convergence of the basic method is invariant by translation, that is, if any translated sequence satisfies this property.

Proposition 2.1. *Let us consider some basic method with iteration mappings* P_n, *and the associated perturbed method with iteration mappings* Q_k.

Let us assume that X *is a Banach space with norm* $\|.\|$ *and topological dual* $X^\star$, *that* τ *is a Hausdorff locally convex topology on* X *compatible with the duality* $X, X^\star$, *that,* $\forall k \in \mathbb{N}$, $\forall x, y \in X$, $\forall \rho \ge 0$,

P_k *is nonexpansive* ,

$\overline{x}$ *is a fixed point of* P_k,

$$\|Q_k x - Q_k y\| \leq (1+\epsilon_k)\|x-y\|, \text{ where } \epsilon_k \geq 0, \sum_{k=1}^{+\infty} \epsilon_k < +\infty,$$

$$\sum_{k=1}^{+\infty} \Delta_{k,\rho} < +\infty, \text{ where } \Delta_{k,\rho} := \sup_{\|x\|\leq\rho} \|Q_k x - P_k x\|.$$

If, for all $k \in \mathbb{N}$, the translated basic method is τ-convergent (cf. remark 2.2), then the perturbed method is τ-convergent too.

Proof. Let $\{x_k\}$ be a sequence generated by the perturbed method.

First we shall prove that $\{x_k\}$ is bounded. We have

$$\|x_k - \overline{x}\| \leq \|Q_k x_{k-1} - Q_k \overline{x}\| + \|Q_k \overline{x} - P_k \overline{x}\| \leq (1+\epsilon_k)\|x_{k-1} - \overline{x}\| + \Delta_{k,\|\overline{x}\|}.$$

So,

$$\|x_k - \overline{x}\| \leq \prod_{i=1}^{k}(1+\epsilon_i)(\|x_0 - \overline{x}\| + \sum_{i=1}^{k} \Delta_{i,\|\overline{x}\|}) < +\infty$$

since, as the ϵ_k-series is convergent, the $(1+\epsilon_k)$-infinite product is finite. Therefore $\{x_k\}$ is bounded in norm by some positive ρ. Then we have

$$\|x_k - P_k x_{k-1}\| = \|Q_k x_{k-1} - P_k x_{k-1}\| \leq \Delta_{k,\rho}.$$

In other words, $\{x_k\}$ is generated by the approximate basic method associated with the given basic method for some "error" term e_k satisfying $\|e_k\| \leq \Delta_{k,\rho}$. Therefore the result follows from Lemma 2.1. □

3. Gradient-prox method for monotone inclusions. Let X be a real Hilbert space endowed with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\|.\|$, let A, B be maximal monotone operators on X, A single valued everywhere defined. We are concerned with the problem of finding a zero of $A+B$, that is, the monotone inclusion

$$0 \in (A+B)x$$

or, what is equivalent for any $\lambda > 0$, the fixed point problem

$$x = J_\lambda^B (I - \lambda A)x$$

where $J_\lambda^B := (I + \lambda B)^{-1}$ is the resolvent (or prox mapping) of B with parameter λ.

In all what follows we assume that the set valued inverse A^{-1} is α-strongly monotone, which is equivalent to the firm nonexpansiveness of αA, that is, $\alpha > 0$ and

$$\forall x, y \in X, \ \ \langle x - y, Ax - Ay \rangle \geq \alpha \|Ax - Ay\|^2.$$

It should be noted that this implies that A is $1/\alpha$-Lipschitz continuous (and therefore $A+B$ maximal monotone), the converse being true if A is the gradient of a convex function ([3]).

Associated with this problem we consider the basic iterative method (gradient-prox method) defined by

$$P_n := J^B_{\lambda_n} \circ (I - \lambda_n A), \qquad 0 < \underline{\lambda} \leq \lambda_n \leq \overline{\lambda} < 2\alpha.$$

A perturbed version of this method has already been studied (as a particular case) in [6, 9] in the context of convex minimization, that is, with A the gradient (resp. B the subdifferential) of a proper closed convex function on X.

We show in the following two lemmas that P_n is c-firmly nonexpansive for a suitable positive c, a crucial property for the convergence of the method.

Lemma 3.1. *Let M_i be c_i-firmly nonexpansive mappings, $i = 1, 2$, that is, $c_i > 0$ and*

$$\forall x, y \in X, \ \ \|M_i x - M_i y\|^2 \leq \|x - y\|^2 - c_i \|(I - M_i)x - (I - M_i)y\|^2.$$

Then $M_1 \circ M_2$ is c-firmly nonexpansive with $c := \min\{c_1, c_2\}/2$.

P r o o f.

$$\|M_1M_2x - M_1M_2y\|^2 \leq \|M_2x - M_2y\|^2 - c_1\|(I - M_1)M_2x - (I - M_1)M_2y\|^2 \leq$$

$$\|x - y\|^2 - c_2\|x - y - (M_2x - M_2y)\|^2 - c_1\|M_2x - M_2y - (M_1M_2x - M_1M_2y)\|^2$$

$$\leq \|x - y\|^2 - \min\{c_1, c_2\}/2\|x - y - (M_1M_2x - M_1M_2y)\|^2 \qquad \square$$

Lemma 3.2.

$$\forall n \in \mathbb{N}, \ \ \forall x, y \in X, \ \ \|P_n x - P_n y\|^2 \leq \|x - y\|^2 - c\|(I - P_n)x - (I - P_n)y\|^2$$

with $c := \min\{1, 2\alpha/\overline{\lambda} - 1\}/2 > 0$.

P r o o f. First we show that, for all n, $I - \lambda_n A$ is $(2\alpha/\overline{\lambda} - 1)$-firmly nonexpansive. Indeed we have

$$\forall x, y \in X, \ \|(I - \lambda_n A)x - (I - \lambda_n A)y\|^2 \leq \|x - y\|^2 + \|\lambda_n Ax - \lambda_n Ay\|^2 - 2\lambda_n\alpha\|Ax - Ay\|^2$$

$$\leq \|x - y\|^2 - (2\alpha/\overline{\lambda} - 1)\|\lambda_n Ax - \lambda_n Ay\|^2.$$

Then, as the composition of the 1-firmly nonexpansive mapping $J^B_{\lambda_n}$ ([5]) with the $(2\alpha/\overline{\lambda} - 1)$-firmly nonexpansive mapping $I - \lambda_n A$, P_n is c-firmly nonexpansive with $c := \min\{1, 2\alpha/\overline{\lambda} - 1\}/2$, thanks to Lemma 3.1. $\square$

Proposition 3.1. *If $S := (A+B)^{-1}(0) \neq \emptyset$, then the (translated) gradient-prox method is weakly convergent (strongly if S has a non empty interior or if $A+B$ is asymptotically well behaved, that is, $d(0,(A+B)u_n) \to 0$ implies $d(u_n, S) \to 0$).*

Proof. Thanks to Remark 2.2 it is sufficient to prove the convergence of the basic method ($k = 0$). From Lemma 3.2 we get

$$\forall \overline{x} \in S, \ \ \|\xi_n - \overline{x}\|^2 \leq \|\xi_{n-1} - \overline{x}\|^2 - c\|\xi_{n-1} - \xi_n\|^2 \tag{3}$$

That ξ_n converges weakly to some point in S is rather standard from (3) and the Lipschitz property of A ([9]). Indeed, $\{\xi_n\}$ is S-Féjer monotone, that is,

$$\forall \overline{x} \in S, \ \ \forall n, \ \ \|\xi_n - \overline{x}\| \leq \|\xi_{n-1} - \overline{x}\|.$$

Therefore $\{\xi_n\}$ is bounded, for all $\overline{x} \in S$, $\|\xi_n - \overline{x}\|$ is convergent and $\|\xi_{n-1} - \xi_n\| \to 0$. Then $\{\xi_n\}$ is stationary for $A+B$, that is,

$$\xi_n^\star := (\xi_{n-1} - \xi_n)/\lambda_n + A\xi_n - A\xi_{n-1} \in (A+B)\xi_n$$

with $\xi_n^\star \to 0$. As the graph of $A+B$ is weak-strong closed, any weak limit point is in S. The uniqueness of such limit point is standard (see for instance [15], proof of Theorem 1). Proof of strong convergence can be found in [9] (proof of Proposition 5.1) in the first case and in [8] (proof of Proposition 5.7) in the second one. □

Remark 3.1. By the way, it is shown in [10] and in [4] as a special case, that when $T(:= A+B$ here) is the subdifferential of a proper closed convex function f on a Banach space X, the asymptotical well behaviour of T is equivalent to (generalized to non uniqueness cases) Tykhonov well posedness of f for minimization. On the other hand, it is shown in [11] that, if f is Gateaux differentiable on a non empty closed convex subset K of X then $f + \delta_K$ is Tykhonov well posed for minimization iff the associated variational inequality is well posed in a suitable sense.

Actually the equivalence between asymptotical well behaviour of $T := T_0 + \partial\varphi$ (with T_0 a set valued operator from X into its dual $X^\star$ and φ a proper convex function on X) and (generalized) Luchetti-Patrone well posedness of the variational inequality (VI):

$$\overline{x} \in X, \ \ \overline{y} \in T_0(\overline{x}), \ \ \forall x \in X, \ \ \langle \overline{y}, x - \overline{x} \rangle + \varphi(x) - \varphi(\overline{x}) \geq 0$$

holds true for a general normed space X. More precisely the Luchetti-Patrone's notion of approximate solution to VI coincides with the notion of approximate solution to the inclusion $0 \in T(x)$ (equivalent statement) that induces the notion of stationary sequence for T. Indeed, let ϵ be a positive real and $x_\epsilon \in X$, an ϵ-approximate solution to VI in the sense of Luchetti-Patrone, that is,

$$\exists y_\epsilon \in T_0(x_\epsilon), \ \ \forall x \in X, \ \ \langle y_\epsilon, x - x_\epsilon \rangle + \varphi(x) - \varphi(x_\epsilon) \geq -\epsilon\|x - x_\epsilon\|.$$

This is equivalent to

$$\inf_{x\in\operatorname{dom}\varphi}\max_{y\in B_\star(0,\epsilon)}\langle y_\epsilon+y,x-x_\epsilon\rangle+\varphi(x)-\varphi(x_\epsilon)\geq 0,$$

(where $B_\star(0,\epsilon)$ denotes the dual closed ball of radius ϵ), which in turn, thanks to Moreau's $\max - \inf$ theorem ([12]) is equivalent to

$$\max_{y\in B_\star(0,\epsilon)}\inf_{x\in\operatorname{dom}\varphi}\langle y_\epsilon+y,x-x_\epsilon\rangle+\varphi(x)-\varphi(x_\epsilon)\geq 0,$$

that is,

$$\exists\overline{y}_\epsilon\in B_\star(0,\epsilon),\ \ \overline{y}_\epsilon\in T(x_\epsilon)$$

or

$$d_\star(0,T(x_\epsilon)\leq\epsilon.$$ □

Now we consider the perturbed method associated with the gradient-prox method defined as follows.

B_k is a maximal monotone operator on X, $A_k := A + G_k$ with G_k a single valued operator on X satisfying

$$\|G_kx-G_ky\|\leq\eta_k\|x-y\|,\ \ \eta_k\geq 0,\exists x_0\in X,\ \ \forall k\in\mathbb{N},\ \ G_k(x_0)=0$$

$$Q_k := J^{B_k}_{\lambda_k}\circ(I-\lambda_kA_k).$$

Lemma 3.3.

$$\forall x,y\in X,\ \ \|Q_kx-Q_ky\|\leq(1+\lambda_k\,\eta_k)\|x-y\|$$

Proof. Straightforward from the nonexpansiveness of $J^{B_k}_{\lambda_k}$ and $(I-\lambda_kA)$ and the definition of A_k. □

Lemma 3.4.

$$\Delta_{k,\rho}\leq\lambda_k\eta_k(\rho+\|x_0\|)+\delta_{\lambda_k,\rho'}(B_k,B)$$

where

$$\rho' := \rho+\overline{\lambda}\|A0\|,\ \ \delta_{\lambda,\rho}(B_k,B) := \sup_{\|x\|\leq\rho}\|J^{B_k}_\lambda x-J^B_\lambda x\|\quad([1,\ 16])$$

Proof. Let $x\in X$ with $\|x\|\leq\rho$. Then, as $I-\lambda_kA$ is non expansive, we have $\|(I-\lambda_kA)x\|\leq\rho+\overline{\lambda}\|A0\|$. Therefore,

$$\|Q_kx-P_kx\|\leq\lambda_k\|G_kx\|+\|J^{B_k}_{\lambda_k}(I-\lambda_kAx)-J^B_{\lambda_k}(I-\lambda_kAx)\|$$

$$\leq \lambda_k \eta_k(\rho + \|x_0\|) + \delta_{\lambda_k,\rho'}(B_k, B) \qquad \square$$

As a direct consequence of Proposition 3.1, Lemmas 3.3 and 3.4, and Proposition 2.1, we get the expected result of stability of the gradient-prox method:

Proposition 3.2. *If* $\sum \lambda_k \eta_k < +\infty$, *and* $\forall \rho \geq 0$, $\sum \delta_{\lambda_k,\rho}(B_k, B) < +\infty$, *then the gradient-prox method is weakly (strongly under the additional assumptions in Proposition 3.1) stable.*

Acknowledgments. The author would like to thank the referee for valuable remarks concerning the presentation.

REFERENCES

[1] H. Attouch, R. J. B. Wets. Isometries for the Legendre-Fenchel transform. *Trans. Amer. Math. Soc.* **296** (1984), 33-60.

[2] P. Alart, B. Lemaire. Penalization in non-classical convex programming via variational convergence. *Math. Programming* **51** (1991), 307-331.

[3] J. B. Baillon, G. Haddad. Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones. *Israel J. Math.* **26** (1977), 137-150.

[4] M. A. Bahraoui, B. Lemaire. Convergence of diagonally stationary sequences in convex optimization. *Set-Valued Analysis* **2** (1994), 1-13.

[5] H. Brezis, P. L. Lions. Produits infinis de résolvantes. *Israel J. Math.* **29**, *4* (1978), 329-345.

[6] B. Lemaire. Coupling optimization methods and variational convergence. In: Trends in Mathematical Optimization, International Series of Numerical Mathematics, vol. **84**(c), Birkhäuser Verlag, 1988, pp. 163-179.

[7] B. Lemaire. About the convergence of the proximal method. In: Advances in Optimization, Proceedings Lambrecht 1991, Lecture Notes in Econom. and Math. Systems vol. **382**, Springer-Verlag, 1992, pp. 39-51.

[8] B. Lemaire. Bounded diagonally stationary sequences in convex optimization. *J. Convex Analysis* **1** (1994), 75-86.

[9] B. Lemaire. On the convergence of some iterative method for convex minimization. In: Recents Developments in Optimization (eds. R. Durier, C. Michelot), Lectures Notes in Econom. and Math. Systems vol. **429**, Springer-Verlag, 1995, pp. 252-268.

[10] B. Lemaire. Bonne position, conditionnement, et bon comportement asymptotique, Séminaire d'Analyse Convexe, Montpellier, 1992, exposé $n^0 5$.

[11] R. Luchetti, F. Patrone. A characterization of Tyhonov well posedness for minimum problems with applications to variational inequalities. *Numer. Funct. Anal. Optim.* **3** (1981), 461-476.

[12] J. J. Moreau. Théorèmes inf-sup. *C. R. Acad. Sci. Paris* **258** (1964), 2720-2722.

[13] K. Mouallif. Convergence variationnelle et méthodes perturbées pour les problèmes d'optimisation et de point selle. Thèse d'Etat, Université de Liège, 1989.

[14] A. Moudafi. Thèse de Doctorat, Université des Sciences et Techniques du Languedoc, 1991.

[15] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **14**, No *5* (1976), 877-898.

[16] P. Tossings. Sur les zéros des opérateurs maximaux monotones et applications. Thèse de Doctorat, Université de Liège, 1990.

B. Lemaire
Dépt. de Mathématiues
Université de Montpellier II
Place Eugéne Bataillon
34095 Montpellier Cedex 05
France

Received January 30, 1996
Revised April 15, 1996