Gilbert Crombez

Non-monotoneous parallel iteration for solving convex feasibility problems

Kybernetika, Vol. 39 (2003), No. 5, [547]--560

Persistent URL: http://dml.cz/dmlcz/135554

KYBERNETIKA — VOLUME 39 (2003), NUMBER 5, PAGES 547–560

NON–MONOTONEOUS PARALLEL ITERATION FOR SOLVING CONVEX FEASIBILITY PROBLEMS

GILBERT CROMBEZ

The method of projections onto convex sets to find a point in the intersection of a finite number of closed convex sets in an Euclidean space, sometimes leads to slow convergence of the constructed sequence. Such slow convergence depends both on the choice of the starting point and on the monotoneous behaviour of the usual algorithms. As there is normally no indication of how to choose the starting point in order to avoid slow convergence, we present in this paper a non-monotoneous parallel algorithm that may eliminate considerably the influence of the starting point.

Keywords: inherently parallel methods, convex feasibility problems, projections onto convex sets, slow convergence

AMS Subject Classification: 65Y05, 65B99, 47H09

1. INTRODUCTION

The method of projections onto convex sets (abbreviated as POCS) is often very well suited to solve the so-called "convex feasibility problem". In the m-dimensional Euclidean space $\mathbb{R}^m$, this problem may be described as follows: given a finite number of closed convex sets $\{C_j\}_{j=1}^n$ in $\mathbb{R}^m$ with nonempty intersection $C^* \equiv \bigcap_{j=1}^{n} C_j$, find a point in C^*. When the individual sets C_j are such that for each of them its corresponding shortest-distance projection operator P_j $(j = 1, \ldots, n)$ is explicitly known, by the POCS-method a sequence is constructed that converges to a point in C^*; depending on the number r $(1 \leq r \leq n)$ of projections used at each step to construct such sequence, sometimes one speaks of a sequential method $(r = 1)$, or a block-iterative method $(1 < r < n)$, or a (fully) parallel method $(r = n)$. An overview of general problems and methods may be found in [1], and in the books [12] and [4]. More specific sequential, block-iterative and parallel methods have been described in [9], [2], [5], [6] and [10]. Examples of the use of convex feasibility problems in applied domains (as for instance in image processing) may be found in [12].

As often has been remarked, however, the sequence that is constructed by the POCS-methods sometimes converges very slowly. The following combined facts may be responsible for this slow convergence: the mutual position of the involved convex sets (i.e., the given problem), and the algorithm used to reach a feasible point. Usually we can only interfere into the used algorithm. To see what facts in a common iteration algorithm may be responsible for slow convergence, we refer to Table 1 accompanying Example 1 at the end of this paper. In this example, 12 disks (having nonempty intersection) with corresponding projection operators $P_1, P_2, \dots, P_{12}$ are given; we use the sequential algorithm indicated as PP, in which we put $T \equiv P_{12}P_{11}\dots P_2P_1$, and in which the iteration sequence $\{\mathbf{x}_k\}_{k=0}^{+\infty}$ is constructed by $\mathbf{x}_{k+1} = T\mathbf{x}_k$. From Table 1 we see that, for some starting points, a point in the intersection is obtained after one application of T, while for other starting points the same method leads to a very slowly converging sequence (the same is true in the parallel case, although for different starting points). The following explanation for this different speed of convergence seems acceptable : for some starting points, the converging sequence $\{\mathbf{x}_k\}_{k=0}^{+\infty}$ may enter some narrow "corridor" between two or more convex sets; the *monotoneous* way of convergence that is present by using a common algorithm is then responsible for very small steps towards the limit point, leading to slow convergence. This monotoneous behaviour for the constructed sequence $\{\mathbf{x}_k\}_{k=0}^{+\infty}$ in $\mathbb{R}^m$ that converges to a point of the intersection C^* is expressed as $|\mathbf{x}_{k+1} - \mathbf{w}| \leq |\mathbf{x}_k - \mathbf{w}|, \forall \mathbf{w} \in C^*$, for all k.

The observations explained above lead to the following conclusions. First of all (and certainly for nonlinear projections), applicable theoretical results about the rate of convergence may be difficult to find. But what is even more important is that the real way out of the difficulties is not situated in finding some algorithm that in all circumstances is the fastest (as such algorithm probably does not exist), but in finding an algorithm that, *independent of the starting point*, leads to an acceptable speed of convergence. Otherwise said, returning to Example 1, it is well acceptable that for some starting points algorithm PP leads to a faster convergence than the new algorithm we want to construct, but the new algorithm should not lead to extremely slow convergence, although this last fact can only be observed experimentally.

Again from our observations, we see that a possible way out of slow convergence could be by allowing (and provoking) nonmonotoneous behaviour of the iteration sequence, because we then have the possibility to leave the small corridor by taking big steps at several iteration points. In a recent paper [8], we already elaborated this idea, and we constructed a parallel algorithm that at different steps in the iteration caused an interruption of the monotoneous behaviour of convergence, and that led to a much faster convergence in those cases where the monotoneous procedure was slow, while keeping an acceptable speed of convergence in the other cases. There was, however, one less desirable fact in the method: part of the computations had to be done in the space $(\mathbb{R}^m)^n$ instead of in $\mathbb{R}^m$.

In this paper, we present a new algorithm that may interrupt the monotoneous behaviour of the iteration sequence $\{\mathbf{x}_k\}_{k=0}^{+\infty}$, but such that all computations may be done in $\mathbb{R}^m$. In Section 2, we explain the theoretical background and in Section 3 we prove convergence of the constructed sequence to a point of C^*. For ease of presentation, we give the construction in the fully parallel case, although the method seems to be equally valid in the block-iterative case. At the end of Section

3 we also give some comments on the fact that the constructed algorithm should be seen as a prototype, but that more flexibility can be incorporated in it. In Section 4, we present two examples to compare the number of iterations needed to obtain convergence for different algorithms.

2. CONSTRUCTION OF THE ALGORITHM

2.1. For ease of reference, we start with a short description of the Pierra method [11] for viewing a parallel iterative projection method in some space as a (semi-) sequential one in a suitable product space.

Let $\mathbb{R}^m$ be the m-dimensional Euclidean space with standard inner product $\langle\ ,\ \rangle$ and norm $|\ |$ derived from $\langle\ ,\ \rangle$; denote $(\mathbb{R}^m,\ \langle\ ,\ \rangle\ ,\ |\ |)$ for short by H. Elements of H are denoted by boldface letters.

Suppose that in H, n closed convex sets $\{C_j\}_{j=1}^n$ are given, having nonempty intersection $C^* \equiv \bigcap_{j=1}^n C_j$. Projection onto C_j is denoted as P_j. We want to obtain a point in C^* by a parallel iterative procedure.

Consider the n-fold product $(\mathbb{R}^m)^n$ of $\mathbb{R}^m$; elements of $(\mathbb{R}^m)^n$ are denoted by capital letters. We introduce an inner product $\langle\langle\ ,\ \rangle\rangle$ and norm $\|\ \|$ on $(\mathbb{R}^m)^n$, as follows: when $V \equiv (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ and $W \equiv (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n)$ are elements of $(\mathbb{R}^m)^n$, put $\langle\langle V, W\rangle\rangle = \frac{1}{n}\sum_{j=1}^n \langle \mathbf{v}_j, \mathbf{w}_j\rangle$, $\|V\|^2 = \frac{1}{n}\sum_{j=1}^n |\mathbf{v}_j|^2$. We denote $((\mathbb{R}^m)^n,\ \langle\langle\ ,\ \rangle\rangle\ ,\ \|\ \|)$ for short by $\mathcal{H}$.

In $\mathcal{H}$, we consider the subsets $\mathcal{D}$ and $\mathcal{F}$, defined as follows. $\mathcal{D}$ is the set of all n-tuples with equal components, i.e., for $\mathbf{v} \in H$ we have that $(\mathbf{v}, \mathbf{v}, \dots, \mathbf{v}) \in \mathcal{D} \subset \mathcal{H}$. $\mathcal{D}$ is the image of H under the canonical imbedding $q : H \to \mathcal{H}$, where for $\mathbf{v} \in H$ we put $q(\mathbf{v}) \equiv (\mathbf{v}, \mathbf{v}, \dots, \mathbf{v})$. $\mathcal{D}$ is a closed linear subspace of $\mathcal{H}$. Projection onto $\mathcal{D}$ is denoted as $P_{\mathcal{D}}$.

The subset $\mathcal{F}$ of $\mathcal{H}$ is defined as the n-fold cartesian product of the convex sets $\{C_j\}_{j=1}^n$ in H, i.e., $\mathcal{F} = C_1 \times C_2 \times \cdots \times C_n$. It is a closed convex set of $\mathcal{H}$, with corresponding projection operator $P_{\mathcal{F}}$.
Clearly, $C^* \neq \emptyset$ is equivalent to $\mathcal{F} \cap \mathcal{D} \neq \emptyset$, and, moreover, $q(C^*) = \mathcal{F} \cap \mathcal{D}$. Hence, obtaining a point in $C^* \subset H$ is equivalent to obtaining a point in $\mathcal{F} \cap \mathcal{D}$. In particular, when we construct a sequence $\{X_k\}_{k=0}^{+\infty}$ in $\mathcal{D} \subset \mathcal{H}$ that converges in $\mathcal{H}$ to a point in $\mathcal{F} \cap \mathcal{D}$, the corresponding sequence $\{\mathbf{x}_k\}_{k=0}^{+\infty}$ in H with $\mathbf{x}_k = q^{-1}(X_k)$ will be convergent in H to a point in C^*.

Use of the Pierra method mentioned above is based on the properties given in Lemma 1, and may be resumed as in Lemma 2. For proofs we refer to [11, Lemma 1.1] and [8, Lemma 1].

Lemma 1. Let $V \equiv (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n) \in \mathcal{H}$. Then

(i) $P_{\mathcal{F}}V = (P_1\mathbf{v}_1, P_2\mathbf{v}_2, \dots, P_n\mathbf{v}_n)$.

(ii) $P_{\mathcal{D}}V = q(\frac{1}{n}\sum_{j=1}^n \mathbf{v}_j)$.

Lemma 2. Suppose that, starting from some point $\mathbf{x}_0$ in H, a sequence $\{\mathbf{x}_k\}_{k=0}^{+\infty}$ in H is constructed by a parallel method, as follows

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_{k+1} \sum_{j=1}^{n} \frac{1}{n}(P_j \mathbf{x}_k - \mathbf{x}_k), \tag{1}$$

where λ_{k+1} denotes a (positive) variable relaxation coefficient. Then, under the natural imbedding q of H into $\mathcal{H}$, the sequence $\{\mathbf{x}_k\}_{k=0}^{+\infty}$ in H is equivalent to the sequence $\{X_k\}_{k=0}^{+\infty}$ in $\mathcal{D} \subset \mathcal{H}$ constructed as follows:

$$\begin{aligned} X_0 &= q(\mathbf{x}_0) \\ X_{k+1} &= X_k + \lambda_{k+1}(P_{\mathcal{D}}(P_{\mathcal{F}} X_k) - X_k). \end{aligned} \tag{2}$$

Hence, the parallel method in H, given by (1), is equivalent to a semi-sequential method in $\mathcal{H}$, given by (2) (we use the word semi-sequential to stress the fact that no relaxation with respect to the projection onto $\mathcal{D}$ is allowed). The procedure in (2) may also be split into two separate steps, as follows:

$$Y_{k+1} = X_k + \lambda_{k+1}(P_{\mathcal{F}} X_k - X_k) \tag{3i}$$

$$X_{k+1} = P_{\mathcal{D}}(Y_{k+1}). \tag{3ii}$$

We also remark that in (1) the same equal weight factors $\frac{1}{n}$ for each projection operator P_j have been used; the procedure works equally well for fixed but different weight factors. For the value of the variable relaxation coefficient λ_{k+1} used at each step in (1) or (2) to obtain convergence, several possibilities are available. In particular, when in going from X_k to X_{k+1} (or from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$), the following value of λ_{k+1} is used

$$\lambda_{k+1} = \frac{\|P_{\mathcal{F}} X_k - X_k\|^2}{\|P_{\mathcal{D}}(P_{\mathcal{F}} X_k) - X_k\|^2} = \frac{\sum_{j=1}^{n} \frac{1}{n} |\mathbf{x}_k - P_j \mathbf{x}_k|^2}{|\mathbf{x}_k - \sum_{j=1}^{n} \frac{1}{n} P_j \mathbf{x}_k|^2}, \tag{4}$$

it has been shown in [7, Formulas (8) and (9)] that the following inequalities are true for each V in $\mathcal{F} \cap \mathcal{D}$:

$$\langle\langle X_{k+1} - V, X_{k+1} - X_k \rangle\rangle \leq 0, \tag{5}$$

$$\|X_{k+1} - V\|^2 \leq \|X_k - V\|^2 - \|X_k - X_{k+1}\|^2. \tag{6}$$

It is also true that $\lambda_{k+1} \geq 1$, for each k.
Moreover, again from [7] we deduce that the following orthogonality relation is true:

$$\|X_k - X_{k+1}\|^2 = \|X_k - P_{\mathcal{F}} X_k\|^2 + \|P_{\mathcal{F}} X_k - X_{k+1}\|^2. \tag{7}$$

2.2. Inequality (6) implies in particular that, when at each iteration step the value of λ_{k+1} given in (4) is used, the resulting sequence $\{X_k\}_{k=0}^{+\infty}$ will have a monotoneous behaviour. We now present an algorithm that, by using suitable values of the relaxation coefficients at regular steps, may lead to a non-monotoneous behaviour, but still will result in a sequence that is convergent to a point of $\mathcal{F} \cap \mathcal{D}$. We first explain the method for one possible interruption point with index $k+1$; afterwards we enumerate all possible interruption points.

Let N be a given positive integer ($N > 2$). Suppose that, starting from some point X_0 in $\mathcal{D} \subset \mathcal{H}$, we obtained by using procedure (2) with the corresponding relaxation coefficients as given in (4), the points $X_1, X_2, \ldots, X_k$ with $k \geq N$ (in particular, for those intermediate points up to and including X_k the properties corresponding to (5) and (6) are true). Now, in order to find X_{k+1}, we will use a relaxation coefficient such that, although it may no longer be true that $\|X_{k+1} - V\| \leq \|X_k - V\|$ for all $V \in \mathcal{F} \cap \mathcal{D}$, it will nevertheless be true that $\|X_{k+1} - V\| < \|X_{k+1-N} - V\|$, for each $V \in \mathcal{F} \cap \mathcal{D}$. Put another way, the monotoneous behaviour of the sequence $\{X_k\}_{k=0}^{+\infty}$ that we are going to construct may be interrupted at X_{k+1} with respect to X_k, but it is repaired with respect to X_{k+1-N}.

To this end, with γ_{k+1} denoting a positive but not yet determined number and with λ_{k+1} as given by (4), put

$$X_{k+1} = X_k + (\lambda_{k+1} + \gamma_{k+1})(P_{\mathcal{D}}(P_{\mathcal{F}} X_k) - X_k). \tag{8}$$

Putting

$$W_{k+1} = X_k + \lambda_{k+1}(P_{\mathcal{D}}(P_{\mathcal{F}} X_k) - X_k), \tag{9}$$

we know that for the couple (W_{k+1}, X_k) the properties corresponding to (5) and (6) are true (replacing in (5) and (6) X_{k+1} by W_{k+1}), and moreover we see that (8) may be rewritten as

$$X_{k+1} = W_{k+1} + \gamma_{k+1}(P_{\mathcal{D}}(P_{\mathcal{F}} X_k) - X_k). \tag{10}$$

For any V in $\mathcal{F} \cap \mathcal{D}$, we derive from (10)

$$\begin{aligned} \|X_{k+1} - V\|^2 &= \|W_{k+1} - V\|^2 + \gamma_{k+1}^2 \|P_{\mathcal{D}}(P_{\mathcal{F}} X_k) - X_k\|^2 \\ &+ 2\gamma_{k+1} \langle\langle W_{k+1} - V, P_{\mathcal{D}}(P_{\mathcal{F}} X_k) - X_k \rangle\rangle. \end{aligned}$$

In view of (9), we may replace $P_{\mathcal{D}}(P_{\mathcal{F}} X_k) - X_k$ in the right-hand-side of the foregoing equality by $\frac{1}{\lambda_{k+1}}(W_{k+1} - X_k)$. Hence, the former equality may be rewritten as

$$\begin{aligned} \|X_{k+1} - V\|^2 &= \|W_{k+1} - V\|^2 + \frac{\gamma_{k+1}^2}{\lambda_{k+1}^2} \|W_{k+1} - X_k\|^2 \\ &+ 2\frac{\gamma_{k+1}}{\lambda_{k+1}} \langle\langle W_{k+1} - V, W_{k+1} - X_k \rangle\rangle. \end{aligned}$$

As we already remarked, the inner product in the last term is non-positive (corresponding to (5)). We conclude that, irrespective of the positive value of γ_{k+1}, we have that

$$\|X_{k+1} - V\|^2 \le \|W_{k+1} - V\|^2 + \frac{\gamma_{k+1}^2}{\lambda_{k+1}^2}\|W_{k+1} - X_k\|^2. \tag{11}$$

The inequality (6), in which we replace X_{k+1} by W_{k+1}, is valid. Taking into account this new inequality in the first term on the right-hand-side of (11) we obtain:

$$\|X_{k+1} - V\|^2 \le \|X_k - V\|^2 - \|X_k - W_{k+1}\|^2 + \frac{\gamma_{k+1}^2}{\lambda_{k+1}^2}\|W_{k+1} - X_k\|^2,$$

which leads to

$$\|X_{k+1} - V\|^2 \le \|X_k - V\|^2 + \left(\frac{\gamma_{k+1}^2}{\lambda_{k+1}^2} - 1\right)\|W_{k+1} - X_k\|^2. \tag{12}$$

Now, the inequality (6) is also always true for the following couples of points: $(X_k, X_{k-1}), (X_{k-1}, X_{k-2}), \dots, (X_{k+1-(N-1)}, X_{k+1-N})$. Hence, repeatedly bounding (by using the inequalities corresponding to (6)) each new first term on the right-hand-side of the expressions obtained from (12), we get

$$\begin{aligned}
&\|X_{k+1} - V\|^2 \\
&\le \|X_{k-1} - V\|^2 - \|X_{k-1} - X_k\|^2 + (\tfrac{\gamma_{k+1}^2}{\lambda_{k+1}^2} - 1)\|W_{k+1} - X_k\|^2 \\
&\le \|X_{k-2} - V\|^2 - \|X_{k-2} - X_{k-1}\|^2 - \|X_{k-1} - X_k\|^2 + (\tfrac{\gamma_{k+1}^2}{\lambda_{k+1}^2} - 1)\|W_{k+1} - X_k\|^2 \\
&\le \dots \\
&\le \|X_{k+1-N} - V\|^2 - \|X_{k+1-N} - X_{k+1-(N-1)}\|^2 - \dots - \|X_{k-2} - X_{k-1}\|^2 \\
&\quad - \|X_{k-1} - X_k\|^2 + (\tfrac{\gamma_{k+1}^2}{\lambda_{k+1}^2} - 1)\|W_{k+1} - X_k\|^2.
\end{aligned}$$

Putting for short

$$M_{k+1} \equiv \|X_{k+1-N} - X_{k+1-(N-1)}\|^2 + \dots + \|X_{k-2} - X_{k-1}\|^2 + \|X_{k-1} - X_k\|^2, \tag{13}$$

the obtained inequality may be written as

$$\|X_{k+1} - V\|^2 \le \|X_{k+1-N} - V\|^2 - M_{k+1} + \left(\frac{\gamma_{k+1}^2}{\lambda_{k+1}^2} - 1\right)\|W_{k+1} - X_k\|^2. \tag{14}$$

Now let α be a given positive number, $0 < \alpha < 1$. In order to be sure that the newly obtained point X_{k+1} is closer to each point V of $\mathcal{F} \cap \mathcal{D}$ than X_{k+1-N} (i.e., in order to repair the monotony when considering X_{k+1} and the points with indices preceding and including $k+1-N$), it is sufficient to choose γ_{k+1} such that the following is true:

$$\left(\frac{\gamma_{k+1}^2}{\lambda_{k+1}^2} - 1\right)\|W_{k+1} - X_k\|^2 = \alpha M_{k+1},$$

which leads to

$$\gamma_{k+1} = \lambda_{k+1}\sqrt{1 + \frac{\alpha M_{k+1}}{\|W_{k+1} - X_k\|^2}}. \tag{15}$$

The just described procedure to construct a single possible interruption point X_{k+1} will now be applied for a specific subsequence of indices, as follows. Let J be a positive integer with $J > N$. Suppose that, starting from some point X_0 in $\mathcal{D}$ and up to the index J, we determine the points $X_1, X_2, \ldots, X_J$ according to the procedure (2) with λ_{k+1} as given by (4). However, for determining $X_{J+1}, X_{J+1+N}, X_{J+1+2N}, \ldots, X_{J+1+pN}, \ldots$ (p a nonnegative integer) we use procedure (8) with the corresponding γ-value as given by (15), while for all intermediate points between X_{J+1+pN} and $X_{J+1+(p+1)N}$ (for all nonnegative integers p) again procedure (2) is used. Then we obtain in $\mathcal{D}$ the sequence $\{X_k\}_{k=0}^{+\infty}$, having the subsequence $\{X_{J+1+pN}\}_{p=0}^{+\infty}$ for which it is true that:

$$\|X_{J+1+pN} - V\|^2 \leq \|X_{J+1+(p-1)N} - V\|^2 - (1-\alpha)M_{J+1+pN}, \text{ for all } V \in \mathcal{F} \cap \mathcal{D}. \tag{16}$$

We want to stress the fact that, contrary to the algorithm in [8], all computations now are done in $\mathcal{D}$ (and hence in H). The only supplementary computational effort, when comparing to a monotoneous parallel method, is to keep a list of the $N-1$ intermediate points between X_{J+1+pN} and $X_{J+1+(p+1)N}$, together with the point X_{J+1+pN} itself; these points are needed in order to compute $X_{J+1+(p+1)N}$.

Before giving the proof of convergence of the constructed sequence, we resume the former result in the following algorithm, stated in the space $\mathcal{H}$.

Algorithm.

Let $\mathcal{H}$ be some Euclidean space with inner product $\langle\langle\ ,\ \rangle\rangle$ and norm $\|\ \|$ derived from it, $\mathcal{D}$ a closed linear subspace of $\mathcal{H}$ and $\mathcal{F}$ a closed convex subset of $\mathcal{H}$ such that $\mathcal{F} \cap \mathcal{D} \neq \emptyset$. Let N and J be positive integers with $J > N$, and let α be a real number between 0 and 1. Starting from some point X_0 in $\mathcal{D}$, construct the sequence $\{X_k\}_{k=0}^{+\infty}$ in $\mathcal{D}$ as follows:

(i) When X_k has been obtained, and $k \notin \{J + mN\}_{m=0}^{+\infty}$, let

$$\lambda_{k+1} = \frac{\|P_{\mathcal{F}}X_k - X_k\|^2}{\|P_{\mathcal{D}}(P_{\mathcal{F}}X_k) - X_k\|^2}, \text{ and compute } X_{k+1} \text{ by}$$

$$X_{k+1} = X_k + \lambda_{k+1}(P_{\mathcal{D}}(P_{\mathcal{F}}X_k) - X_k).$$

(ii) When X_k has been obtained, and $k \in \{J + mN\}_{m=0}^{+\infty}$, let λ_{k+1} be determined as before; put

$$\begin{aligned} W_{k+1} &= X_k + \lambda_{k+1}(P_{\mathcal{D}}(P_{\mathcal{F}}X_k) - X_k), \\ M_{k+1} &= \|X_{k-1} - X_k\|^2 + \|X_{k-2} - X_{k-1}\|^2 + \cdots + \|X_{k+1-N} - X_{k+1-(N-1)}\|^2, \\ \gamma_{k+1} &= \lambda_{k+1}\sqrt{1 + \tfrac{\alpha M_{k+1}}{\|W_{k+1} - X_k\|^2}}, \end{aligned}$$

and compute X_{k+1} by

$$X_{k+1} = W_{k+1} + \gamma_{k+1}(P_{\mathcal{D}}(P_{\mathcal{F}}X_k) - X_k).$$

Then the constructed sequence $\{X_k\}_{k=0}^{+\infty}$ converges to a point of $\mathcal{F} \cap \mathcal{D}$.

3. CONVERGENCE OF THE CONSTRUCTED SEQUENCE

In this section we prove that the sequence $\{X_k\}_{k=0}^{+\infty}$, constructed in $\mathcal{D} \subset \mathcal{H}$ as described in the algorithm, is convergent to a point $A \in \mathcal{D} \cap \mathcal{F}$. By construction, the sequence $\{X_k\}_{k=0}^{+\infty}$ contains in particular the subsequence $\{X_{J+1+pN}\}_{p=0}^{+\infty}$, that we often will denote shortly as $\{X_{n_p}\}_{p=0}^{+\infty}$. The proof of convergence of the sequence $\{X_k\}_{k=0}^{+\infty}$ consists of the following three parts:

(i) The subsequence $\{X_{n_p}\}_{p=0}^{+\infty}$ contains a subsequence, denoted (for simplicity) as $\{X'_s\}_{s=0}^{+\infty}$, that converges to a point $A \in \mathcal{D} \cap \mathcal{F}$.

(ii) Each converging subsequence of $\{X_{n_p}\}_{p=0}^{+\infty}$ converges to the same point A.

(iii) The sequence $\{X_k\}_{k=0}^{+\infty}$ converges to A.

Proof of (i).

As a consequence of inequality (16), the subsequence $\{X_{n_p}\}_{p=0}^{+\infty}$ has the Fejér monotony property, i.e., $\|X_{J+1+pN} - V\| \leq \|X_{J+1+(p-1)N} - V\|$, for all V in $\mathcal{F} \cap \mathcal{D}$. In particular, this leads to the conclusion that the subsequence $\{X_{n_p}\}_{p=0}^{+\infty}$ is bounded, and hence it contains a subsequence, denoted as $\{X'_s\}_{s=0}^{+\infty}$, that converges to some point $A \in \mathcal{H}$. As the original sequence $\{X_k\}_{k=0}^{+\infty}$ belongs to $\mathcal{D}$ and as $\mathcal{D}$ is closed, the point A certainly belongs to $\mathcal{D}$. We now show that A also belongs to $\mathcal{F}$.
Again from the inequality (16) we deduce recursively that, when X_{n_p} and X_{n_q} with $q > p$ denote successive terms appearing in the subsequence denoted as $\{X'_s\}_{s=0}^{+\infty}$, then the following inequality is also true for all V in $\mathcal{F} \cap \mathcal{D}$

$$\|X_{n_q} - V\|^2 \leq \|X_{n_p} - V\|^2 - (1-\alpha)[M_{n_{p+1}} + M_{n_{p+2}} + \cdots + M_{n_q}]. \tag{17}$$

Hence, letting n_p and n_q both tend to infinity, we conclude that the sequence $\{\|X_{n_q} - V\|\}_{n_q=0}^{+\infty}$ tends to some nonnegative number $d(V)$, and that the expression $M_{n_{p+1}}$ tends to zero when $n_p \to +\infty$. In particular, we see from (13) that the expression $M_{n_{p+1}}$ contains the part $\|X_{n_p} - X_{n_p+1}\|^2$, that also tends to zero when $n_p \to +\infty$. This result, combined with the orthogonality relation (7) which states that $\|X_{n_p} - X_{n_p+1}\|^2 = \|X_{n_p} - P_{\mathcal{F}}X_{n_p}\|^2 + \|P_{\mathcal{F}}X_{n_p} - X_{n_p+1}\|^2$, leads to the conclusion that $\|X_{n_p} - P_{\mathcal{F}}X_{n_p}\|^2 \to 0$ when $n_p \to +\infty$. Finally, when X_{n_p} again is used as the general term X'_s of the subsequence $\{X'_s\}_{s=0}^{+\infty}$ that converges to A, and when we write:

$$\|P_{\mathcal{F}}X_{n_p} - A\| \leq \|P_{\mathcal{F}}X_{n_p} - X_{n_p}\| + \|X_{n_p} - A\|,$$

we know that both terms on the right-hand-side tend to zero when n_p (or s) tends to infinity. Hence, also the subsequence $\{P_{\mathcal{F}}X_{n_p}\}_{n_p}$ (also denoted as $\{P_{\mathcal{F}}X'_s\}_{s=0}^{+\infty}$) tends to A. As the sequence $\{P_{\mathcal{F}}X'_s\}_{s=0}^{+\infty}$ belongs to $\mathcal{F}$, and as $\mathcal{F}$ is closed, also A belongs to $\mathcal{F}$. Hence, $A \in \mathcal{F} \cap \mathcal{D}$.

Proof of (ii).

Let us suppose that the subsequence $\{X_{n_p}\}_{p=0}^{+\infty}$ contains another converging subsequence, denoted (again for simplicity) as $\{X'_t\}_{t=0}^{+\infty}$, but that this subsequence converges to a point A'. With a proof as in (i) above it will follow that also A' belongs to $\mathcal{F} \cap \mathcal{D}$. We now prove that $A' = A$.

The subsequence $\{X'_s\}$ is convergent to the point A. Writing $X'_s - A'$ as $X'_s - A + A - A'$, and developing, leads to:

$$\|X'_s - A'\|^2 - \|X'_s - A\|^2 = 2 < X'_s - A, A - A' > + \|A - A'\|^2. \tag{18}$$

In the same way, we obtain:

$$\|X'_t - A\|^2 - \|X'_t - A'\|^2 = 2 < X'_t - A', A' - A > + \|A' - A\|^2. \tag{19}$$

As remarked in the proof of (i) above, taking into account that both A and A' belong to $\mathcal{F} \cap \mathcal{D}$, the number sequences $\{\|X_{n_p} - A'\|\}_{p=0}^{+\infty}$ and $\{\|X_{n_p} - A\|\}_{p=0}^{+\infty}$ are convergent, with respective limits $d(A')$ and $d(A)$, and of course the same is true for their respective subsequences $\{\|X'_s - A'\|\}_{s=0}^{+\infty}$, $\{\|X'_t - A'\|\}_{t=0}^{+\infty}$, $\{\|X'_s - A\|\}_{s=0}^{+\infty}$ and $\{\|X'_t - A\|\}_{t=0}^{+\infty}$. Taking in (18) and (19) the limit, respectively for $s \to +\infty$ and for $t \to +\infty$, we obtain:

$$d(A')^2 - d(A)^2 = 0 + \|A - A'\|^2$$

and

$$d(A)^2 - d(A')^2 = 0 + \|A' - A\|^2.$$

Hence, $d(A) = d(A')$, and from this it easily follows that $A = A'$.

Proof of (iii).

The complete sequence $\{X_k\}_{k=0}^{+\infty}$ contains the specific subsequence $\{X_{n_p}\}_{p=0}^{+\infty}$ that already converges to the point A in $\mathcal{F} \cap \mathcal{D}$. Let now j be any index of the sequence $\{X_k\}_{k=0}^{+\infty}$, i.e., $j \in \mathbb{Z}^+$. Then there exist successive indices n_p and n_q of the subsequence $\{X_{n_p}\}_{p=0}^{+\infty}$ (with $n_p < n_q$) such that $n_p < j \leq n_q$. For $j \equiv n_q$, there is nothing to prove. For $j < n_q$, we know from the way of construction that for the successive points $X_{n_p}, X_{n_p+1}, \ldots, X_{j-1}, X_j$ the Fejér monotony property is valid; in particular, we have that $\|X_j - A\| \leq \|X_{j-1} - A\| \leq \cdots \leq \|X_{n_p+1} - A\| \leq \|X_{n_p} - A\|$, and we know that $\|X_{n_p} - A\| \to 0$ when $n_p \to +\infty$. Hence, when $j \to +\infty$ we also have that $X_j \to A$. □

We summarize the foregoing convergence result of the algorithm in the following theorem, stated in the original space $H \equiv (\mathbb{R}^m, \langle\ ,\ \rangle, |\ \ |)$.

Theorem. Suppose that in H, n closed convex sets $\{C_i\}_{i=1}^n$ with corresponding projection operators $\{P_i\}_{i=1}^n$ and with nonempty intersection $\bigcap_{i=1}^n C_i$ are given. Let N and J be positive integers, $N > 2$ and $J > N$, and let α be a real number with $0 < \alpha < 1$. Suppose that, starting from some point $\mathbf{x}_0$ in H, the point $\mathbf{x}_k$ of the sequence $\{\mathbf{x}_k\}_{k=0}^{+\infty}$ has been obtained, and that the next iteration point $\mathbf{x}_{k+1}$ is constructed according to the following rule:

(i) When $k \notin \{J + mN\}_{m=0}^{+\infty}$, put :

$$\lambda_{k+1} = \frac{\sum_{j=1}^{n} \frac{1}{n}|\mathbf{x}_k - P_j\mathbf{x}_k|^2}{|\mathbf{x}_k - \sum_{j=1}^{n} \frac{1}{n}P_j\mathbf{x}_k|^2},$$

and construct $\mathbf{x}_{k+1}$ by:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_{k+1} \sum_{j=1}^{n} \frac{1}{n}(P_j\mathbf{x}_k - \mathbf{x}_k).$$

(ii) When $k \in \{J + mN\}_{m=0}^{+\infty}$, determine λ_{k+1} as in (i). Put:

$$\mathbf{w}_{k+1} = \mathbf{x}_k + \lambda_{k+1} \sum_{j=1}^{n} \frac{1}{n}(P_j\mathbf{x}_k - \mathbf{x}_k),$$

$$M_{k+1} = |\mathbf{x}_{k-1} - \mathbf{x}_k|^2 + |\mathbf{x}_{k-2} - \mathbf{x}_{k-1}|^2 + \cdots + |\mathbf{x}_{k+1-N} - \mathbf{x}_{k+1-(N-1)}|^2,$$

$$\gamma_{k+1} = \lambda_{k+1}\sqrt{1 + \frac{\alpha M_{k+1}}{|\mathbf{w}_{k+1} - \mathbf{x}_k|^2}},$$

and construct $\mathbf{x}_{k+1}$ by:

$$\mathbf{x}_{k+1} = \mathbf{w}_{k+1} + \gamma_{k+1} \sum_{j=1}^{n} \frac{1}{n}(P_j\mathbf{x}_k - \mathbf{x}_k).$$

Then, although for the sequence $\{\mathbf{x}_k\}_{k=0}^{+\infty}$ the monotoneous behaviour may be interrupted at some (or all) indices $k \in \{J + mN\}_{m=0}^{+\infty}$, the sequence converges to a point of $\bigcap_{i=1}^{n} C_i$.

Remark concerning the flexibility of the algorithm.

As stated in the introduction, the algorithm as described above should be seen as a prototype of a class of algorithms allowing more flexibility. Without going into details, the following extensions seem to be possible:

a. Instead of using equal weight factors $\frac{1}{n}$ in the Pierra method, a family of fixed but non-equal weight factors $\{\omega_i\}_{i=1}^{n}$ with $0 < \omega_i < 1$ for each i and $\sum_{i=1}^{n} \omega_i = 1$ may be used.

b. In the given algorithm, we try to provoke an interruption of the monotony at every a priori fixed number of N iterations, by taking a well-determined big step at each of these a priori fixed moments. The algorithm could also be adapted such that a big step is taken after each "variable" number N_p of iterations, where N_p changes between two fixed integers M_1 and M_2 (e.g., $M_1 = 5$ and $M_2 = 100$), and where N_p plays the role of N above. From a practical point of view, the explanation is as follows: it seems necessary to take a big step when in a number of foregoing iterations very small steps have been taken; hence, the user of the algorithm keeps record of the distances $|\mathbf{x}_k - \mathbf{x}_{k-1}|$ of successive iteration points. From some index k on, we proceed as follows : when the sum of M_1 such distances is "too small", the algorithm should provoke an interruption of the monotoneous behaviour; on the other hand, if this sum of M_1 distances is "big enough", then for each number of steps between M_1 and M_2 the user decides whether and where an interruption has to be created, by comparing the sum of each number (between M_1 and M_2) of distances with a list of thresholds; at the number N_p between M_1 and M_2 where the sum of the distances is for the first time smaller than the wanted threshold, the algorithm should force an interruption of the monotoneous behaviour; and finally, when no interruption has been forced "between M_1 and M_2", and when the sum of M_2 distances is still big enough, the user provokes nevertheless an interruption at M_2.

4. EXAMPLES AND CONCLUDING REMARKS

In this last section we consider two examples to compare the results of the algorithm given in Section 2 with the ones corresponding to some classical methods.

In our first example, we take twelve disks in the plane as closed convex sets $\{C_j\}_{j=1}^{12}$; these disks are given by the following expressions (with respect to an orthonormal system of axes):

$$\left(x - \cos\left(\frac{j\pi}{12}\right)\right)^2 + \left(y - \sin\left(\frac{j\pi}{12}\right)\right)^2 \leq 1, \text{ for } j = 1, \ldots, 12.$$

(here, (x, y) denotes a generic point in the plane).

Clearly, their intersection (in fact determined by C_1 and C_{12}) is nonempty; in particular, (0,0) is a point in their intersection, but it contains more points. Explicit expressions for the associated projection operators $\{P_j\}_{j=1}^{12}$ may be found in [3]. Starting from some given point in the plane, and using those projection operators, we want to find a point in the intersection. We use the following algorithms:

PP: The method of pure projections in a sequentially composed manner, i.e., when $\mathbf{x}_k$ is the current iteration point, and when we put $T \equiv P_{12}P_{11} \cdots P_2P_1$, then the update $\mathbf{x}_{k+1}$ is given by $\mathbf{x}_{k+1} = T\mathbf{x}_k$. It is a monotoneous procedure.

PAR: The parallel projection method given in (1) and (2), with fixed equal weights $\left(\frac{1}{n} \equiv \frac{1}{12}\right)$ at each step, and with λ_{k+1} as given by (4). Again, this leads to a monotoneous way of convergence.

NMPAR (0.9,5,10): The non-monotoneous parallel procedure developed in this paper. As in PAR, use has been made of fixed equal weights at each step and of the value of λ_{k+1} as given by (4). Besides, it contains the following parameters as explained in the Algorithm:

α: The real number between 0 and 1 that may be responsible for interrupting the monotoneous behaviour; in the example we put $\alpha = 0.9$.

N: The number that determines the period to repeat the use of the adapted relaxation coefficient; we took $N = 5$ in the example.

J: The first index where the adapted relaxation coefficient is used; in the example we put $J = 10$.

In Table 1 at the end of this paper we have given, for eight different starting points ((-3,0),...) either the number of iterations needed to obtain a point in the intersection (this is a positive integer), or the sum of the distances of the current iteration point to the twelve sets C_j after 25 and 50 iterations respectively. From this table the following conclusions may be made.

– PP: the influence of the choice of the starting point on the speed of convergence is very clear; sometimes convergence is obtained after one step, while in other cases there is a very slow way of convergence. This completely unpredictable behaviour of convergence makes it rather unlikely to obtain practical useful results in a theoretical manner. In applied problems it is not at all clear what guess of starting point to make in order to assure fast convergence. Together with a bad choice of starting point, the monotoneous behaviour of the iteration sequence seems to be responsible for slow convergence.

– Method PAR: the same remarks as for PP can be made; for some starting points there is a quick convergence, in other cases convergence is slow.

– Method NMPAR (0.9,5,10): the results of this method seem to confirm what had already been observed by use of another non-monotoneous method in [8]: for those cases where the monotoneous parallel method gives a fast convergence, the same is true for the non-monotoneous method (in our example, the number of iterations in those cases is equal, due to the choice of our parameters); but, for those cases where either PP or PAR lead to slow convergence, NMPAR leads to convergence in an acceptable number of steps. The explanation for this phenomenon seems to be that, by interruption of the monotoneous behaviour, the newly obtained iteration point has left some small corridor which is responsible for small steps in the iteration, and as a consequence the intersection of the convex sets is approached from another direction.

Finally, we want to mention the influence of the parameter α on the speed of convergence. From the theoretical investigation in the construction of the algorithm it is clear that, as α becomes closer to 1, the corresponding step length in the iteration is bigger, and this increases the probability of interrupting the monotoneous behaviour.

In our second example, we consider 15 sets C_i in $\mathbb{R}^2$ ($i = 1, \dots, 15$) having the following form:

$$C_i = \{\mathbf{x} \in \mathbb{R}^2 : b_{i,1} \leq \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_{i,2}\},$$

with $\{b_{i,1}\}_{i=1}^{15}$ and $\{b_{i,2}\}_{i=1}^{15}$ sets of real numbers, and with $\{\mathbf{a}_i\}_{i=1}^{15}$ a set of 15 given points in $\mathbb{R}^2$. These data are chosen such that $\bigcap_{i=1}^{15} C_i \neq \emptyset$.

The following algorithms have been used:

PP1RO: the sequential iteration scheme where at each iteration step only one pure projection onto a set C_j, chosen at a Random Order, has been used; i.e., when $\mathbf{x}_k$ is the current iteration point, then $\mathbf{x}_{k+1} = P_j\mathbf{x}_k$, where P_j is the projection operator onto a set C_j choosen at random. Moreover, for this algorithm and for each given starting point, the algorithm was run 30 times. The number appearing in Table 2, corresponding to algorithm PP1RO and to a given starting point, is the average number of iterations needed to obtain convergence.

Table 1.

Starting point	(-3,0)	(10,-10)	(3,4)
PP	1	3.279208×10^{-3}	3.661634×10^{-3}
		5.000838×10^{-4}	5.49556×10^{-4}
PAR	9.972098×10^{-3}	4	1.129448×10^{-2}
	3.128052×10^{-3}		3.427267×10^{-3}
NMPAR (0.9,5,10)	22	4	22

Starting point	(-17,12)	(-2,1)	(-100,-50)
PP	3.601907×10^{-3}	3.202676×10^{-3}	1
	5.419265×10^{-4}	4.89951×10^{-4}	
PAR	1.185358×10^{-2}	9.768488×10^{-3}	8.859039×10^{-3}
	3.548027×10^{-3}	3.080129×10^{-3}	2.859947×10^{-3}
NMPAR (0.9,5,10)	22	22	24

Starting point	(2,-4)	(0,2)
PP	3.005983×10^{-3}	3.694175×10^{-3}
	4.637248×10^{-4}	5.537283×10^{-4}
PAR	5	9.757404×10^{-3}
		3.077506×10^{-3}
NMPAR (0.9,5,10)	5	25

PP1λRO: the sequential iteration scheme where at each iteration step one relaxed projection onto a randomly chosen set C_j is used; i.e., when $T_j = \mathbf{1} + \lambda_j(P_j - \mathbf{1})$ with $\mathbf{1}$ the identity operator on H and with λ_j a positive real number, then $\mathbf{x}_{k+1} = T_j\mathbf{x}_k$. In the example, each λ_j had the value 1.5. As in PP1RO, the number appearing in Table 2 corresponding to PP1λRO and to a given starting point is the average number of iterations needed to obtain convergence.

PAR: the common parallel algorithm as in example 1, but now for 15 sets (n=15).

NMPAR (0.9,5,10) and NMPAR (0.9,20,25): the non-monotoneous parallel procedures developed in this paper, but with parameters $\alpha = 0.9, N = 5, J = 10$ and $\alpha = 0.9, N = 20, J = 25$ respectively.

For PAR and for NMPAR (, ,), the numbers figuring in Table 2 give the number of iterations that was necessary to obtain convergence. From Table 2, we may conclude that, also for this example, the new algorithm has better convergence properties than the traditional ones.

Table 2.

Starting point	(0,0)	(-10,10)	(9,2)
PP1RO	1046	1565	398
PP1λRO	395	716	236
PAR	884	888	3
NMPAR (0.9,5,10)	82	89	3
NMPAR (0.9,20,25)	187	69	3

Starting point	(-3,6)	(5,-1)	(7,8)
PP1RO	769	1347	775
PP1λRO	398	574	400
PAR	329	923	326
NMPAR (0.9,5,10)	40	142	30
NMPAR (0.9,20,25)	50	89	51

(Received February 3, 2003.)

REFERENCES

[1] H. Bauschke and J. Borwein: On projection algorithms for solving convex feasibility problems. SIAM Rev. *38* (1996), 367–426.

[2] D. Butnariu and Y. Censor: On the behaviour of a block-iterative projection method for solving convex feasibility problems. Internat. J. Computer Math. *34* (1990), 79–94.

[3] D. Butnariu and S. D. Flam: Strong convergence of expected-projection methods in Hilbert spaces. Numer. Funct. Anal. Optim. *16* (1995), 601–636.

[4] Y. Censor and S.A. Zenios: Parallel Optimization. Theory, Algorithms, and Applications. Oxford University Press, New York 1997.

[5] P. L. Combettes: Construction d'un point fixe commun à une famille de contractions fermes. C. R. Acad. Sci. Paris Sér. I Math. *320* (1995), 1385–1390.

[6] G. Crombez: Viewing parallel projection methods as sequential ones in convex feasibility problems. Trans. Amer. Math. Soc. *347* (1995), 2575–2583.

[7] G. Crombez: Solving convex feasibility problems by a parallel projection method with geometrically defined parameters. Appl. Anal. *64* (1997), 277–290.

[8] G. Crombez: Improving the speed of convergence in the method of projections onto convex sets. Publ. Math. Debrecen *58* (2001), 29–48.

[9] L. G. Gubin, B. T. Polyak, and E. V. Raik: The method of projections for finding the common point of convex sets. U.S.S.R. Comput. Math. and Math. Phys. *7* (1967), 1–24.

[10] K. Kiwiel: Block-iterative surrogate projection methods for convex feasibility problems. Linear Algebra Appl. *215* (1995), 225–259.

[11] G. Pierra: Decomposition through formalization in a product space. Math. Programming *28* (1984), 96–115.

[12] H. Stark and Y. Yang: Vector Space Projections. Wiley, New York 1998.

Dr. Gilbert Crombez, Department of Applied Mathematics and Computer Science, Ghent University, Krijgslaan 281/S9, B-9000 Gent. Belgium.
e-mail: Gilbert.Crombez@UGent.be