Numer. Math. 49, 367–378 (1986)

Numerische Mathematik

Convergence Results for an Accelerated Nonlinear Cimmino Algorithm

Alfredo N. Iusem [1] and Alvaro Rodolfo De Pierro [2]

[1] Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, Brasil
[2] Instituto de Matemática, Universidade Federal do Rio de Janeiro, CP 68530, Rio de Janeiro, Brasil

Summary. We present an accelerated version of Cimmino's algorithm for solving the convex feasibility problem in finite dimension. The algorithm is similar to that given by Censor and Elfving for linear inequalities. We show that the nonlinear version converges locally to a weighted least squares solution in the general case and globally to a feasible solution in the consistent case. Applications to the linear problem are suggested.

Subject Classifications. AMS(MOS): 52A05, 65F10, 90C25; CR: G1.6.

1. Introduction

The idea of calculating centroids of systems of masses originated with Cimmino's iterative algorithm for the solution of a system of linear equations (see [4], [7, p. 160] and [8, p. 119]). The original method of Cimmino was generalized for solving integral equations of the first kind by Kammerer and Nashed [9]. This generalization and others produced algorithms for the solution of the convex feasibility problem, which consists in finding a point in the intersection of a finite family of closed convex sets. Nashed [12] studied the behavior of such algorithms in the non consistent case, i.e. when the intersection of the family is empty.

There exist several other iterative schemes for solving convex feasibility problems. A recent review of these methods may be found in [1].

We remark three main advantages of our method. The first, shared by all Cimmino-like methods, is that each iterate can be easily implemented in parallel processors, which is not the case of sequential methods as [17]. On the other hand, as it will explained in Sect. 6, it avoids possibly slow convergence due to a large number of satisfied constraints specially near the limit. The third advantage is that in many real applications, errors in measures do not guarantee existence of feasible solutions; therefore approaches which do not rely on consistency of the system and algorithms like this, which are capable of handling inconsistent systems, are desirable (see e.g. [16, 18]).

In this paper we study the behavior of the nonlinear version of Cimmino's algorithm presented by Censor and Elfving in [2] for solving linear inequalities. In [3] the same authors generalize the algorithm for solving the convex feasibility problem in finite dimension, but adding a factor in the relaxation parameter that might cause slow convergence when the number of equations is large. In both works the nonfeasible case is not studied. Here, we give a local convergence proof, which is global in the feasible case, to a weighted least squares solution using original proof techniques. Finally, we suggest some applications of the method to the solution of interval inequalities.

2. The Algorithm

Let $C_1, C_2, \ldots, C_m$ be closed convex sets in $\mathbb{R}^n$. Take $0<\lambda_i \in \mathbb{R}$ $(1 \leqq i \leqq m)$ such that $\sum_{i=1}^{m} \lambda_i = 1$. Let P_i: $\mathbb{R}^n \to C_i$ be the projection on C_i, i.e.

$$P_i x = \arg\min_{y \in C_i} \|x-y\|, \tag{1}$$

where $\|\cdot\|$ denotes the norm induced by the standard inner product in $\mathbb{R}^n$, $\langle\ ,\ \rangle$. Define P: $\mathbb{R}^n \to \mathbb{R}^n$ as

$$Px = \sum_{i=1}^{m} \lambda_i P_i x. \tag{2}$$

Since the P_i's are continuous, P is continuous.

For $x \in \mathbb{R}^n$, let $I(x) = \{i\colon x \notin C_i\}$ and $C(x)$ the cardinal of the set $I(x)$. Define:

$$\mu(x) = \begin{cases} \left(\sum_{i \in I(x)} \lambda_i\right)^{-1} & \text{if } C(x) \geqq 2 \\ 1 & \text{otherwise.} \end{cases} \tag{3}$$

Let:

$$\bar{P}x = x + \mu(x)(Px - x). \tag{4}$$

Consider $v_i(x) = \mu(x)\lambda_i$. Observe that, if $C(x) \geqq 2$

$$v_i(x) > 0 \qquad \forall\, i \in I(x) \tag{5}$$

$$\sum_{i \in I(x)} v_i(x) = 1. \tag{6}$$

It follows from (2), (3), (4) and (6) that

$$\bar{P}x = \begin{cases} \sum_{i \in I(x)} v_i(x) P_i x & \text{if } C(x) \geqq 2 \\ Px & \text{otherwise.} \end{cases} \tag{7}$$

The algorithm is now defined by the sequence: (for a given initial point x^0),

$$x^{k+1} = \bar{P}x^k. \tag{8}$$

3. Auxiliary Results

In this section we prove some facts regarding the projections P_i and the operators P and $\bar{P}$.

Proposition 1. Let $x, y \in \mathbb{R}^n$. Then, for any i:

i) $\|P_i x - P_i y\| \leqq \|x - y\|$,

ii) $\|P_i x - P_i y\| = \|x - y\|$ implies that $P_i x - P_i y = x - y$.

Proof. By the Convex Separation Theorem (see, e.g. [10]):

$$\langle x - P_i x, P_i y - P_i x \rangle \leqq 0, \tag{9}$$

$$\langle P_i y - y, P_i y - P_i x \rangle \leqq 0. \tag{10}$$

Adding (9) and (10),

$$\|P_i x - P_i y\|^2 \leqq \langle x - y, P_i x - P_i y \rangle \leqq \|x - y\| \, \|P_i x - P_i y\|. \tag{11}$$

So (i) holds. If $\|P_i x - P_i y\| = \|x - y\|$, then the right hand side of (11) is equal to $\|P_i x - P_i y\|^2$ meaning that equality holds throughout (11). Hence, there exists $\alpha \geqq 0$ such that $P_i x - P_i y = \alpha(x - y)$. Using again the hypothesis of (ii), $\alpha = 0$ or $\alpha = 1$; in both cases ii) holds. □

Proposition 2. *Let* $x, y \in \mathbb{R}^n$, *then*

i) $\|Px - Py\| \leqq \|x - y\|$,

ii) $\|Px - Py\| = \|x - y\|$ *implies that* $Px - Py = x - y$.

Proof. i)

$$\|Px - Py\| = \left\| \sum_{i=1}^{m} \lambda_i (P_i x - P_i y) \right\| \leqq \sum_{i=1}^{m} \lambda_i \|P_i x - P_i y\|$$
$$\leqq \left(\sum_{i=1}^{m} \lambda_i \right) \|x - y\| = \|x - y\|,$$

using Proposition 1.i).

ii) If $\|Px - Py\| = \|x - y\|$, we have equality throughout the last chain of inequalities, so $\sum_{i=1}^{m} \lambda_i \|P_i x - P_i y\| = \|x - y\|$. Applying Proposition 1.i) conclude that

$$\|P_i x - P_i y\| = \|x - y\| \quad \text{for } 1 \leqq i \leqq m.$$

By Proposition 1.ii), $P_i x - P_i y = x - y$. Then, $Px - Py = x - y$. □

Let now define

$$F = \{x \in \mathbb{R}^n : Px = x\},$$

i.e., F is the set of fixed points of the operator P (possibly empty).

Proposition 3. *Let* $x \in \mathbb{R}^n$, $z \in F$. *Then,*

i) $\|Px - z\| \leqq \|x - z\|$,

ii) $\|Px - z\| = \|x - z\|$ *implies that* $x \in F$.

Proof. Apply Proposition 2, observing that $z = Pz$. □

Proposition 4. *Let* $y, z \in F$. *Then* $P_i z - z = P_i y - y$ $(1 \leqq i \leqq m)$.

Proof. Since $\|Py - Pz\| = \|y - z\|$, conclude, as in the proof of Proposition 2.ii) that $P_i y - P_i z = y - z$ $(1 \leqq i \leqq m)$. □

The next proposition needs the following:

Lemma 1. *For any* $x, y \in \mathbb{R}^n$,

$$\langle Px - Py, y - Py \rangle \leqq \sum_{i=1}^{m} \sum_{j=i}^{m} \lambda_i \lambda_j \langle (P_i x - P_j x) - (P_i y - P_j y), P_i y - P_j y \rangle.$$

Proof. Using (9) for any i,

$$\langle P_i x - P_i y, y - P_i y \rangle \leqq 0 \quad \text{implies} \quad \langle P_i x - P_i y, y \rangle \leqq \langle P_i x - P_i y, P_i y \rangle,$$

then

$$\langle Px - Py, y \rangle \leqq \sum_{i=1}^{m} \lambda_i \langle P_i x - P_i y, P_i y \rangle,$$

summing on i.

Substracting $\langle Px - Py, y - Py \rangle$ from the last inequality, we obtain:

$$\begin{aligned}\langle Px - Py, y - Py \rangle &\leqq \sum_{i=1}^{m} \lambda_i \langle P_i x - P_i y, P_i y \rangle - \langle Px - Py, Py \rangle \\ &= \left(\sum_{j=1}^{m} \lambda_j \right) \sum_{i=1}^{m} \lambda_i \langle P_i x - P_i y, P_i y \rangle \\ &\quad - \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j \langle P_i x - P_i y, P_j y \rangle \\ &= \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j \langle P_i x - P_i y, P_i y - P_j y \rangle \\ &= \sum_{i=1}^{m} \sum_{j=i}^{m} \lambda_i \lambda_j \langle (P_i x - P_j x) - (P_i y - P_j y), P_i y - P_j y \rangle. \quad \square\end{aligned}$$

Proposition 5. *For any* $z \in F$, $x \in \mathbb{R}^n$, $\langle z - Px, x - Px \rangle \leqq 0$.

Proof. Apply Lemma 1 with $x = z$ and $y = x$, then,

$$\begin{aligned}\langle z - Px, x - Px \rangle &= \langle Pz - Px, x - Px \rangle \\ &\leqq \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \lambda_j \langle (P_i z - P_j z) - (P_i x - P_j x), P_i x - P_j x) \rangle. \end{aligned} \tag{12}$$

Apply Lemma 1 with $y = z$, then,

$$0 = \langle Px - Pz, z - Pz \rangle \leqq \sum_{i=1}^{m} \sum_{j=i}^{m} \lambda_i \lambda_j \langle (P_i x - P_j x) - (P_i z - P_j z), P_i z - P_j z \rangle. \tag{13}$$

Adding (12) and (13) together:

$$\langle z - Px, x - Px \rangle \leqq - \sum_{i=1}^{m} \sum_{j=i}^{m} \lambda_i \lambda_j \|(P_i x - P_j x) - (P_i z - P_j z)\|^2 \leqq 0. \quad \square$$

Take now $J \subset \{1, 2, \ldots, m\}$ and let Q_J: $\mathbb{R}^n \to \mathbb{R}^n$ be defined as

$$Q_J = \sum_{i \in J} v_i(J) P_i x, \tag{14}$$

where $v_i(J) = \frac{\lambda_i}{\sum_{j \in J} \lambda_j}$. So $\sum_{i \in J} v_i(J) = 1$. Let $F_J = \{x \in \mathbb{R}^n\colon Q_J x = x\}$. It is clear that we can substitute in Proposition 5 F by F_J and P by Q_J, so

$$\text{for all } x \in \mathbb{R}^n \text{ and } z \in F_J, \text{ we have that } \langle x - Q_J x, z - Q_J x \rangle \leqq 0. \tag{15}$$

Proposition 6. $\{x \in \mathbb{R}^n\colon \bar{P} x = x\} = F$.

Proof. From (4), since $\mu(x) \neq 0$. $\forall x \in \mathbb{R}^n$. □

It follows from Proposition 4 that the set $I(z)$ is the same for any $z \in F$. Let I be such set.

Proposition 7. *If $J \neq \emptyset$ and $I \subset J$, then $F \subset F_J$.*

Proof. Take $z \in F$ and let $K = \{1, 2, \ldots, m\} - J$. Since $I \subset J$, for $i \in K$, $P_i z = z$. So

$$\begin{aligned} z = Pz &= \sum_{i \in J} \lambda_i P_i z + \sum_{i \in K} \lambda_i P_i z = (\sum_{i \in J} \lambda_i) Q_J z + (\sum_{i \in K} \lambda_i) z \\ &= (\sum_{i \in J} \lambda_i) Q_J z + (1 - \sum_{i \in J} \lambda_i) z, \quad \text{then } Q_J z = z. \quad \square \end{aligned}$$

4. Convergence Results

In this section we first characterize F as the set of minimizers of the weighted average (with the λ_i's) of the squares of the distances to the sets C_i. After this, we obtain a convergence result for the sequence defined by the operator P and finally we arrive to our main convergence result for the algorithm defined by the operator $\bar{P}$.

Define the positive function f: $\mathbb{R}^n \to \mathbb{R}$ as

$$f(x) = \sum_{i=1}^{m} \lambda_i \|P_i x - x\|^2. \tag{16}$$

Let G be the set of minimizers of f. Being a convex combination of distances to convex sets, f is clearly a convex function. It follows that G is convex (if non empty).

We define also the sequence:

$$y^{k+1} = P y^k \quad (k = 0, 1, 2, \ldots). \tag{17}$$

Lemma 2. *For any $y \in \mathbb{R}^n$, $f(Py) \leqq f(y) - \|Py - y\|^2$.*

Proof. Since $P_i P y$ is the closest point to Py in C_i, we have

$$\|P_i P y - P y\|^2 \leqq \|P_i y - P y\|^2 = \|P_i y - y\|^2 + \|y - P y\|^2 - 2\langle P_i y - y, P y - y \rangle.$$

Summing on i:

$$f(Py)=\sum_{i=1}^{m}\lambda_i\|P_iPy-Py\|^2\leqq\sum_{i=1}^{m}\lambda_i\|P_iy-y\|^2+\|y-Py\|^2-2\|Py-y\|^2$$
$$=f(y)-\|Py-y\|^2. \quad \square$$

As an immediate consequence, we have,

Corollary 1. *$f(y^k)$ is a decreasing sequence.*

Lemma 3. *For any $y^0\in\mathbb{R}^n$, $\|Py^k-y^k\|$ tends to zero as $k\to\infty$.*

Proof. From Lemma 2, we have that $\|Py^k-y^k\|\leqq f(y^k)-f(y^{k-1})$. By Corollary 1, $f(y^k)$ is a positive decreasing sequence, hence convergent. It follows that $f(y^k)-f(y^{k-1})\to 0$ as $k\to\infty$. $\square$

Theorem 1. *$F=G$.*

Proof. i) $G\subset F$. From Lemma 2, $0\leqq\|Py-y\|^2\leqq f(y)-f(Py)$. If $y\in G$, $f(y)-f(Py)\leqq 0$. So $Py=y$ and $y\in F$.

ii) $F\subset G$. Take $z\in F$. Assume, by contradiction, that $z\notin G$. So there exists $y\in\mathbb{R}^n$ such that $f(y)<f(z)$. Consider the set

$$A=\{x\in\mathbb{R}^n\colon f(x)\leqq f(y)\}.$$

By continuity and convexity of f, A is closed and convex. Let y^0 be the (unique) closest point to z in A. By Lemma 2, $f(Py^0)\leqq f(y^0)$, so $Py^0\in A$. The definition of y^0 implies now $\|Py^0-z\|\geqq\|y^0-z\|$. On the other hand $\|Py^0-z\|=\|Py^0-Pz\|\leqq\|y^0-z\|$ by Proposition 2i). By uniqueness of y^0, conclude that $Py^0=y^0$. So $Py^0-Pz=y^0-z$. From Proposition 4, $P_iy^0-y^0=P_iz-z$ for any i. Hence $f(y^0)=f(z)$. But $f(y^0)\leqq f(y)<f(z)$, a contradiction. Therefore $f(z)\leqq f(y)$ for any $y\in\mathbb{R}^n$, that is to say, $z\in G$. $\square$

Corollary 2. *If $C=\bigcap_{i=1}^{m}C_i\neq\emptyset$, then $F=C$.*

Proof. $z\in C$ implies that $z\in C_i$, $\forall i$, i.e., $P_iz=z$, $\forall i$, equivalent to $f(z)=0$. So $C=G=F$. $\square$

Theorem 2. *For any $y^0\in R^n$, the sequence defined by* (17) *converges if and only if $F\neq\emptyset$. When it does converge the limit point belongs to F.*

Proof. If the sequence converges, by continuity of the operator P, it is clear that the limit point is a fixed point. Reciprocally, if $F\neq\emptyset$, suppose $z\in F$, then by Proposition 2i)

$$\|y^{k+1}-z\|\leqq\|y^k-z\|. \tag{18}$$

Therefore $\|y^k-z\|$ is a decreasing (hence bounded) sequence, then (y^k) has a convergent subsequence $y^{k_j}\to y$. By continuity of P, $Py^{k_j}\to Py$ and $\|Py^{k_j}-y^{k_j}\|^2\to\|Py-y\|^2$. From Lemma 3, we get $\|Py-y\|=0$, i.e., $y\in F$. Take for any $\varepsilon>0$, an l such that $\|y^{j_l}-y\|<\varepsilon$. Now for $k>j_l$ and taking into account

(18) we have

$$\|y^k - y\| \leqq \|y^{j_l} - y\| < \varepsilon,$$

i.e., $y^k \to y$. □

For the sequence defined by (8) we will prove a local convergence theorem (similar to Theorem 2) which becomes global when $C \neq \emptyset$, in which case $F = C$. First we show that if the sequence defined by (8) converges, the limit point belongs to F. This is not an immediate consequence of Proposition 6 (as would be the case for the sequence defined by (17)) because the operator $\bar{P}$ is not continuous: note that μ takes only a finite number of values. From now on (x^k) will always refer to the sequence defined by (8).

Lemma 4. *If* $x^k \xrightarrow[k\to\infty]{} x$, *then* $x \in F$.

Proof. Since $\mu(x) \geqq 1$ for all $x \in \mathbb{R}^n$:

$$\|Px^k - x^k\| \leqq \mu(x^k)\|Px^k - x^k\| = \|x^{k+1} - x^k\|. \tag{19}$$

Taking limits throughout (19) as $k \to \infty$ and remembering that P is continuous, get

$$\|Px - x\| \leqq 0, \quad \text{i.e., } x \in F. \quad \square$$

Assume now $F \neq \emptyset$ and let $\rho = \frac{1}{2}\min_{i\in I}\{\|P_i z - z\|\}$ for some $z \in F$. By Proposition 4, ρ is independent of $z \in F$. By convention, $\rho = \infty$ if $I = \emptyset$. Let $B = \{x \in \mathbb{R}^n \colon \exists\, z \in F$ such that $\|x - z\| \leqq \rho\}$.

Lemma 5. *If* $x \in B$, $z \in F$, *then* $\|\bar{P}x - z\| \leqq \|Px - z\| \leqq \|x - z\|$.

Proof. Take $x \in B$. By definition of B, there exists $z^0 \in F$ such that $\|x - z^0\| \leqq \rho$. Note that if $i \in I$, $\|x - z^0\| < \|P_i z^0 - z^0\|$, then $x \notin C_j$, i.e., $I \subset I(x)$. By definition of $\bar{P}$, $\bar{P}x$ is equal either to $Q_{I(x)}x$ or to Px. By Proposition 7, $z \in F_{I(x)}$. Also Px lies in the segment between $\bar{P}x$ and x because $\mu(x) \geqq 1$. Since the angle between x, $\bar{P}x$ and z is obtuse by (15) or Proposition 5, the statement of the lemma is true. □

Theorem 3. *If* $F \neq \emptyset$ *and* $x^0 \in B$, *then* x^k *converges to a point in* F.

Proof. Take $z \in F$ such that $\|x^0 - z\| \leqq \rho$. By Lemma 5:

$$\|x^1 - z\| = \|\bar{P}x^0 - z\| \leqq \|x^0 - z\| \leqq \rho, \qquad \text{so, } x^1 \in B.$$

Applying recursively the same argument, we conclude that

i) $x^k \in B$, for all k, (20)

ii) $\|x^k - z\|$ decreases (21)

It follows that (x^k) is bounded, so there is a convergent subsequence

$$x^{j_k} \xrightarrow[k\to\infty]{} x.$$

Since $x^{j_k} \in B$, by (20), we apply again Lemma 5, together with (21):

$$\|x^{j_{k+1}} - z\| \leqq \|x^{j_k+1} - z\| = \|\bar{P}x^{j_k} - z\| \leqq \|Px^{j_k} - z\| \leqq \|x^{j_k} - z\|. \tag{22}$$

Taking limits as $k \to \infty$ throughout (22), remembering that P is continuous:

$$\|x-z\| \leqq \|Px-z\| \leqq \|x-z\|, \quad \text{then } \|Px-z\| = \|x-z\|.$$

By Proposition 3, $x \in F$. Apply now Lemma 5. In view of (20):

$$\|x^{k+1}-x\| = \|\bar{P}x^k - x\| \leqq \|x^k - x\|.$$

So $\|x^k - x\|$ decreases in k. Since for the subsequence (x^{j_k}) we have $\|x^{j_k}-x\| \xrightarrow[k\to\infty]{} 0$, we conclude that $\|x^k - x\| \xrightarrow[k\to\infty]{} 0$. So the whole sequence converges to $x \in F$. □

Taking into account Theorem 1 and Corollary 2, we may rephrase Theorem 3 as

Corollary 3. *If $F \neq \emptyset$ and $x^0 \in B$, the sequence defined by* (8) *converges to a point which minimizes a convex combination* (*with coefficients λ_i*) *of the squares of the distances to the sets C_i.*

Observe that when $I = \emptyset$, we have $B = \mathbb{R}^n$. Since $I = \emptyset$ is equivalent to $P_i z = z$ $(1 \leqq i \leqq m)$, i.e., $C \neq \emptyset$, by Corollary 2, in the consistent case Theorem 3 can be restated as:

Corollary 4. *If $C \neq \emptyset$, the sequence defined by* (8) *converges to a point in C from any starting point $x^0 \in \mathbb{R}^n$.*

So, the local convergence result of Theorem 2 becomes global in the consistent case.

5. Applications

Although the method studied in the preceeding sections may be used for solving general feasibility problems, it supposes that orthogonal projections are easy to compute, which is not always a realistic hypothesis. Nevertheless, general proofs for convex sets provide also a theoretical foundation for the convergence of the algorithm when applied to the solution of interval inequalities.

Let us consider the problem of finding $x \in \mathbb{R}^n$ such that

$$b^1 \leqq Ax \leqq b^2, \tag{23}$$

where A is an $m \times n$ matrix (without zero rows), b^1 and b^2 being m-vectors, such that $b^1 < b^2$. Suppose also that A, b^1 and b^2 have the form

$$A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{pmatrix}, \quad b^1 = \begin{pmatrix} b_1^1 \\ b_2^1 \\ \vdots \\ b_p^1 \end{pmatrix}, \quad b^2 = \begin{pmatrix} b_1^2 \\ b_2^2 \\ \vdots \\ b_p^2 \end{pmatrix} \tag{24}$$

where $A_1, A_2, \ldots, A_p$ are orthogonal blocks (i.e., $A_i A_i^t = D_i$, D_i a diagonal matrix, for $i=1,2,\ldots,p$) and $b_1^j, b_2^j, \ldots. b_p^j$ (for $j=1,2$), the correspondent block vectors. We remark that this is a practical situation for large and sparse problem, as can be seen in [6, 14].

In this case, orthogonal projections on the convex sets defined by each block are very easy to compute. This is shown in the following,

Proposition 8. *If B is an $s \times t$ orthogonal matrix, i.e., $(b^i)^t b^j = 0$ for $i \neq j$, being b^i, $(i=1,\ldots,s)$ the nonnull rows of B, c^1 and c^2 s-vectors such that $c^1 < c^2$, then*

i) $M = \{x \in \mathbb{R}^t \colon c^1 \leqq Bx \leqq c^2\} \neq \emptyset$.

ii) *The orthogonal projection of a point $x \in R^t$ on M is given by*

$$\bar{x} = x - B^t D^{-2} z, \tag{25}$$

where D is the diagonal matrix $(\|b^1\|, \ldots, \|b^s\|)$ and z is the vector defined by

$$z_i = \begin{cases} (b^i)^t x - c_i^2 & \text{if } (b^i)^t x \geqq c_i^2 \\ 0 & \text{if } c_i^1 \leqq (b^i)^t x \leqq c_i^2 \\ (b^i)^t x - c_i^1 & \text{if } (b^i)^t x \leqq c_i^1 \end{cases} \tag{26}$$

Proof. i) By Gale's theorem (see, e.g., [11, p. 33]), $M \neq \emptyset$ if and only if the system:

$$(B^t, -B^t)\begin{pmatrix} y^1 \\ y^2 \end{pmatrix} = 0 \quad \text{and} \quad (c^2)^t y^2 - (c^1)^t y^1 = 1 \tag{27}$$

has no solution with $y^1, y^2 \geqq 0$.

But (27) means that

$$B^t(y^1 - y^2) = 0, \tag{28}$$

and applying B on both sides of (28) we have that $y^1 = y^2$. Therefore by (27),

$$(c^2 - c^1)^t y^1 = -1.$$

But $c^2 - c^1 > 0$ by hypothesis, then, if $y^1 \geqq 0$, we arrive to a contradiction.

ii) Observe now that, taking into account (26),

$$B\bar{x} - c^2 = Bx - c^2 - z \leqq 0 \tag{29}$$

and

$$B\bar{x} - c^1 = Bx - c^1 - z \geqq 0. \tag{30}$$

On the other hand

$$\bar{x} - x = -B^t w + B^t \bar{w} \tag{31}$$

where w and $\bar{w}$ are s-vectors defined by

$$w_i = \begin{cases} c_i^1 - (b^i)^t x & \text{if } (b^i)^t x = c_i^2 \\ 0 & \text{otherwise,} \end{cases} \qquad \bar{w}_i = \begin{cases} (b^i)^t x - c_i^2 & \text{if } (b^i)^t \bar{x} = c_i^2 \\ 0 & \text{otherwise} \end{cases} \tag{32}$$

Now, (29), (30) and (31) are the Kuhn-Tucker conditions for the quadratic problem:

$$\min \tfrac{1}{2}\|y-x\|^2 \quad \text{subject to } c^1 \leqq By \leqq c^2,$$

and the thesis follows. □

With the above results, we can apply algorithm (8) to the problem (23) with

$$P_i x = x - A_i^t D_i^{-2} z^i \quad \text{for } i=1, \dots, p,$$

z^i as in (26).

For this particular case, it is very easy to prove that the function f is piecewise quadratic on a finite number of polyhedral sets; then it achieves its minimum as a consequence of Frank-Wolfe's Theorem ([13], Corollary 27.3.1). This fact implies that the set F of Theorem 1 is not empty and convergence is guaranteed by Theorem 3.

6. Final Remarks

The main advantages of the algorithm defined by (8) with respect to simpler implementations of Cimmino's method (i.e. with relaxations parameters which do not depend upon the point x) lies in the acceleration effect achieved when $\mu(x)$ attains a large value, as compared with the original Cimmino's algorithm, defined by the operator P. A first insight into this effect follows directly from Lemma 5. Assume x^k belongs to B and let $x^{k+1} = Px^k$, $\bar{x}^{k+1} = \bar{P}x^k$. Lemma 5 indicates that $\bar{x}^{k+1}$ will be closer to any point in F then x^{k+1}. By a slight refinement of Lemma 5 we can get a quantitative estimate of this acceleration effect. Take for instance the case of equal weights, i.e. $\lambda_i = \frac{1}{m}$ $(1 \leqq i \leqq m)$. In this case $\mu(x) = \frac{m}{C(x)}$. For $x \in B$, the obtuse angle property (15) applies and, since Px lies in the segment between x and $\bar{P}x$, we get, for any $z \in F$:

$$\|\bar{P}x - z\|^2 \leqq \|Px - z\|^2 - \|Px - \bar{P}x\|^2 = \|Px - z\|^2 - \left(\frac{m}{C(x)} - 1\right) \|Px - x\|^2.$$

So our algorithm will generate from x a point which is closer to any point in F (in the sense of the square of the norm) than the one generated from the same x by the original Cimmino's method by an ammount at least as large as $\left(\frac{m}{C(x)} - 1\right) \|Px - x\|^2$. The factor $\frac{m}{C(x)} - 1$ is strictly positive unless all constraints are violated, in which case both algorithms produce the same next iterate.

If m is very large (in applications such as computerized tomography [1] m can be as large as 300,000) and $C(x)$ is much smaller than m, which is likely to happen near convergence for inequality constraints, $\frac{m}{C(x)} - 1$ becomes very large as well and the acceleration effect is more significant.

Since the matrices in such applications often present no detectable structure, it can be expected that the number of active constraints at the solution point be small.

On the other hand, in algorithms with point-independent relaxation parameters, like:

$$x^{k+1} = x^k + \alpha_k (P x^k - x^k) \tag{33}$$

convergence is guaranteed only when the relaxation parameters α_k satisfy:

$$0 < \varepsilon_1 \leqq \alpha_k \leqq \varepsilon_2 < 2 \qquad (\varepsilon_1, \varepsilon_2 > 0). \tag{34}$$

In fact, global convergence for (33)-(34) can be established following a line similar to our convergence proof for the algorithm defined by (17) using the obtuse angle property of Proposition 5 (see detailed proofs in [5] for the linear case and in [15] for the non linear case).

Let $x^k \in B$, $\bar{x}^{k+1} = \bar{P} x^k$ as before and x^{k+1} as defined in (33) with α_k satisfying (34). Following the same argument we get, for any $z \in F$:

$$\|\bar{x}^{k+1} - z\|^2 \leqq \|x^{k+1} - z\|^2 - \left(\frac{m}{C(x)} - 2\right) \|P x^k - x^k\|^2 \tag{35}$$

i.e. our algorithm produces points closer to F than those generated by (33) as long as less than half of the constraints are violated, and much closer when the numbers of violated constraints is much smaller than the total number of constraints. By the way, the algorithm in [3], although its relaxation parameters are point-dependent, behaves in the same way as the algorithm defined by (33) as compared with our algorithm, when m is large.

These remarks can be extended to the case of different weights, just by counting each constraint with its weight. For instance, if x^{k+1} and $\bar{x}^{k+1}$ are as in (35), $\bar{x}^{k+1}$ will be closer to F than x^{k+1} as long as the sum of the weights of the violated constraints is less than one half.

The algorithms defined by (8) and (33) can be combined in different ways. One possibility is to merge them into:

$$x^{k+1} = x^k + \alpha_k (\bar{P} x^k - x^k) \tag{36}$$

with α_k as in (34) but this variation is unlikely to produce a significant improvement over (8).

Another possibility is to start with (33) and switch to (8) when $\dfrac{m}{C(x^k)}$ becomes large and/or the algorithm seems to approach convergence, i.e. $\|x^{k+1} - x^k\|$ is small. This strategy may be effective when C is empty, since the acceleration effect of our algorithm (and its own convergence) is guaranteed in such a case only when $x^k \in B$, i.e. when the sequence gets close to F.

Regarding global convergence for the algorithm defined by (8) in the inconsistent case, the following example shows that further refinement is required (see Fig. 1).

In the situation below, the sequence generated by (8) oscillates between the points x^k and $\bar{P} x^k$. We conjecture global convergence if $C(x)$ in (3) is substituted by the cardinal number of the set $\bar{I}(x)$ defined as $\bar{I}(x) = \{P_i x \colon P_i x \neq x\}$.

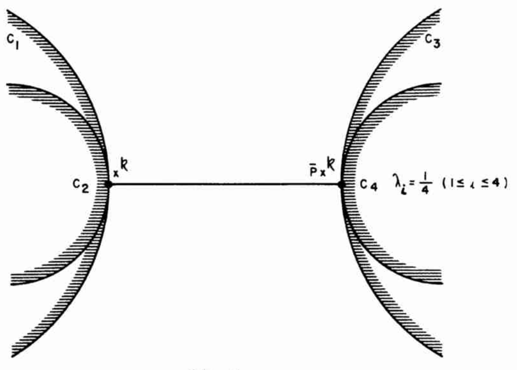

Fig. 1

References

1. Censor, Y.: Row-action methods for huge and sparse systems and their applications. SIAM Rev. **23**, 444–466 (1981)
2. Censor, Y., Elfving, T.: New methods for linear inequalities. Linear Algebra Appl. **42**, 199–211 (1982)
3. Censor, Y., Elfving, T.: A nonlinear Cimmino algorithm. Report 33, Department of Mathematics, University of Haifa 1981
4. Cimmino, G.: Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. Ric. Sci. **1**, 326–333 (1938)
5. De Pierro, A., Iusem, A.: A Simultaneous Projection Method for Linear Inequalities. Linear Algebra Appl. **64**, 243–253 (1985)
6. Elfving, T.: Block iterative methods for consistent and inconsistent linear equations. Numer. Math. **33**, 1–12 (1980)
7. Gastinel, N.: Analise Numérique Linéaire. Paris: Hermann 1970
8. Householder, A.S.: The Theory of Matrices in Numerical Analysis. New York: Dover Publications 1964
9. Kammerer, W.J., Nashed, M.Z.: A generalization of a matrix iterative method of G. Cimmino to best approximate solution of linear integral equation of the first kind. Rend. Accad. Naz. dei Lincei **51**, 20–25 (1971)
10. Luenberger, D.G.: Optimization by Vector Space Methods. New York: John Wiley 1969
11. Mangasarian, O.L.: Nonlinear Programming. New York: McGraw-Hill 1969
12. Nashed, M.Z.: Continuous and semicontinuous analogues of iterative methods of Cimmino and Kaczmarz with applications to the inverse Radon Transform. In: Mathematical Aspects of Computerized Tomography (G.T. Herman and F. Notherer, Eds.). Lect. Notes Med. Inf. **8**, 160–178 (1981)
13. Rockafeller, R.T.: Convex Analysis. Cambridge: University Press 1970
14. Tewarson, R.P.: Projection method for solving sparse linear systems. Comput. J. **12**, 77–80 (1969)
15. De Pierro, A., Iusem, A.: A Parallel Projection Method of Finding a Common Point of a Family of Convex Sets. Pesquisa Operacional, 1–20 (1985)
16. Censor, Y., Elfving, T., Herman, G.T.: Regularized least squares solution of linear inequalities. Technical Report No. MIPG97. Department of Radiology, University of Pennsylvania 1985
17. Gubin, L.G., Polyak, B.T., and Raik, E.V.: The method of projections for finding the common point of convex sets. USSR Comput. Math. Math. Phys. **7**, 1–24 (1967)
18. Herman, G.T., Lent, A., Hurwitz, H.: A storage-efficient algorithm for finding the regularized solution of a large inconsistent system of equations. J. Inst. Math. Appl. **25**, 361–366 (1980)

Received May 8, 1984 / November 11, 1985