A Simultaneous Projections Method for Linear Inequalities

Alvaro R. De Pierro*
Instituto de Matemática
Universidade Federal do Rio de Janeiro
C. P. 68530
Rio de Janeiro, Brazil

and

Alfredo N. Iusem
Instituto de Matemática Pura e Aplicada
Estrada Dona Castorina 110
Rio de Janeiro, Brazil

Submitted by Richard A. Brualdi

ABSTRACT

An iterative method for solving general systems of linear inequalities is considered. The method, a relaxed generalization of Cimmino's scheme for solving linear systems, was first suggested by Censor and Elfving. Each iterate is obtained as a convex combination of the orthogonal projections of the previous iterate on the half spaces defined by the linear inequalities. The algorithm is particularly suitable for implementation on computers with parallel processors. We prove convergence from any starting point for both consistent and nonconsistent systems (to a feasible point in the first case, and to a weighted least squares type solutions in the second).

1. INTRODUCTION

Recent applications which require the solution of huge systems of linear equations have provoked a renewed interest in iterative methods for general linear systems (e.g. linear programming [4, 7], image reconstruction from projections [6]).

*This research was partially supported by CNPq under grant No. 30.1699/81.

52 Vanderbilt Ave., New York, NY 10017 0024-3795/85/$3.30

Several sequential methods (derived mostly from Kaczmarz's) have been suggested [1]. In these methods each iterate is obtained from the previous one by considering only one equality or inequality. Convergence results for the inequality case are limited to consistent systems.

Cimmino [3] devised an iterative scheme for the solution of a finite system of linear algebraic equations. The method starts with an arbitrary point in $\mathbb{R}^n$ as an initial approximation, and then calculates at each step the centroid of a system of masses placed at the reflections of the previous iterate with respect to the hyperplanes defined by the system of equations. This centroid is taken as a new iterate. Cimmino's method is discussed in [5]. See [8] for a lucid geometric description of both Kaczmarz's and Cimmino's methods, including considerations on the behavior of the latter in the inconsistent case.

Censor and Elfving [2] generalized Cimmino's method to linear inequalities and gave convergence proofs only for the feasible case. In this paper we consider a special case of Censor and Elfving's method which converges from any starting point to a solution of the system of linear inequalities in the consistent case and to a weighted least squares solution in the general case. The study of the behavior of the algorithm in the practically important situation when the problem is infeasible is made possible by our approach to the convergence proof, which differs from Censor and Elfving's.

Geometrically, each iterate lies in the half line determined by the previous one and a convex combination of its orthogonal projections on all the half spaces defined by the inequalities.

These simultaneous algorithms (as opposed to successive ones) are particularly suitable for implementation on computers with parallel processors.

2. THE ALGORITHM

Consider the system of linear inequalities:

$$\langle a^i, x \rangle \leqslant b_i \qquad (1 \leqslant i \leqslant r), \tag{1}$$

where $a^i, x \in \mathbb{R}^n$ $(a^i \neq 0)$, $b_i \in \mathbb{R}$, $2 \leqslant r$, and $\langle\ ,\ \rangle$ is the Euclidean inner product.

Let $\lambda_1, \dots, \lambda_r$ be real numbers such that

$$\sum_{i=1}^{r} \lambda_i = 1, \quad 0 < \lambda_i \qquad (1 \leqslant i \leqslant r). \tag{2}$$

Define, for $x \in \mathbb{R}^n$, $1 \leqslant i \leqslant r$,

$$c_i(x) \triangleq \min\left\{0, \frac{b_i - \langle a^i, x\rangle}{\|a^i\|^2}\right\} \tag{3}$$

where $\|\ \|$ stands for the Euclidean norm in $\mathbb{R}^n$. Then

$$P_i x = x + c_i(x) a^i \tag{4}$$

defines the orthogonal projection of the point x on the closed half space

$$C_i \triangleq \{x \in \mathbb{R}^n : \langle a^i, x\rangle \leqslant b_i\}.$$

Now, define $P: \mathbb{R}^n \to \mathbb{R}^n$ by

$$P = \sum_{i=1}^{r} \lambda_i P_i \tag{5}$$

and, for any $\alpha \in \mathbb{R}$,

$$P_\alpha = I + \alpha(P - I), \tag{5$'$}$$

where I is the identity function.

With these definitions at hand, we state the simultaneous projection algorithm for solving linear inequalities.

ALGORITHM. Take an arbitrary $x^0 \in \mathbb{R}^n$. Define inductively

$$x^{k+1} = P_\alpha x^k \tag{6}$$

with $0 < \alpha < 2$.

In the next section we introduce some properties of the operators P and P_α.

3. AUXILIARY RESULTS

We start with two elementary properties of orthogonal projections, which follow easily from Theorem 11.2 in [9].

For any $x, y \in \mathbb{R}^n$, $1 \leqslant i \leqslant r$

$$\langle P_i y - P_i x, x - P_i x \rangle \leqslant 0, \tag{7}$$

$$\|P_i y - P_i x\| \leqslant \|y - x\|, \tag{8}$$

$$\|P_i x - P_i y\| = \|x - y\| \quad \Rightarrow \quad P_i x - P_i y = x - y. \tag{8'}$$

LEMMA 1. *For any $x, y \in \mathbb{R}^n$, $\|Py - Px\| \leqslant \|y - x\|$.*

Proof.

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

$$\leqslant \|y - x\|$$

[using (8) and (2)]. ∎

LEMMA 2. *Let $v, w \in \mathbb{R}^n$. Then*

$$\langle Pw - Pv, v - Pv \rangle \leqslant \sum_{i=1}^{r} \sum_{j=i}^{r} \lambda_i \lambda_j$$

$$\times \left(\|P_i v - P_j v\| \|P_i w - P_j w\| - \|P_i v - P_j v\|^2 \right).$$

Proof. From (7), for any i,

$$\langle P_i w - P_i v, v - P_i v \rangle \leqslant 0 \quad \Rightarrow \quad \langle P_i w - P_i v, v \rangle \leqslant \langle P_i w - P_i v, P_i v \rangle$$

$$\Rightarrow \langle Pw - Pv, v \rangle \leqslant \sum_{i=1}^{r} \lambda_i \langle P_i w - P_i v, P_i v \rangle.$$

Substracting $\langle Pw - Pv, Pv \rangle$ from the last inequality, get

$$\begin{aligned}
\langle Pw - Pv, v - Pv \rangle &\leqslant \sum_{i=1}^{r} \lambda_i \langle P_i w - P_i v, P_i v \rangle - \langle Pw - Pv, Pv \rangle \\
&= \left(\sum_{j=1}^{r} \lambda_j \right) \sum_{i=1}^{r} \lambda_i \langle P_i w - P_i v, P_i v \rangle \\
&\quad - \sum_{i=1}^{r} \sum_{j=1}^{r} \lambda_i \lambda_j \langle P_i w - P_i v, P_j v \rangle \\
&= \sum_{i=1}^{r} \sum_{j=1}^{r} \lambda_i \lambda_j \langle P_i w - P_i v, P_i v - P_j v \rangle \\
&= \sum_{i=1}^{r} \sum_{j=i}^{r} \lambda_i \lambda_j \left(\langle P_i w - P_j w, P_i v - P_j v \rangle - \|P_i v - P_j v\|^2 \right) \\
&\leqslant \sum_{i=1}^{r} \sum_{j=i}^{r} \lambda_i \lambda_j \left(\|P_i w - P_j w\| \, \|P_i v - P_j v\| - \|P_i v - P_j v\|^2 \right) \quad \blacksquare
\end{aligned}$$

Consider the (possibly empty up to now) set F of fixed points of P, i.e. $F = \{ z \in \mathbb{R}^n : Pz = z \}$. Then we have the following lemma:

Lemma 3. *For any $z \in F$, $x \in \mathbb{R}^n$, $\langle z - Px, x - Px \rangle \leqslant 0$.*

Proof. Define $a_{ij} = \|P_i x - P_j x\|$, $b_{ij} = \|P_i z - P_j z\|$. Apply Lemma 2 with $w = z$, $v = x$:

$$\langle z - Px, x - Px \rangle = \langle Pz - Px, x - Px \rangle \leqslant \sum_{i=1}^{r} \sum_{j=i}^{r} \lambda_i \lambda_j \left(b_{ij} a_{ij} - a_{ij}^2 \right). \tag{9}$$

Apply Lemma 2 with $w = x$, $v = z$:

$$0 = \langle Px - Pz, z - Pz \rangle \leqslant \sum_{i=1}^{r} \sum_{j=i}^{r} \lambda_i \lambda_j \left(b_{ij} a_{ij} - b_{ij}^2 \right). \tag{10}$$

Add (9) and (10) together:

$$\langle z - Px, x - Px \rangle \leqslant - \sum_{i=1}^{r} \sum_{j=i}^{r} \lambda_i \lambda_j (a_{ij} - b_{ij})^2 \leqslant 0. \qquad \blacksquare$$

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

$$f(x) = \sum_{i=1}^{r} \lambda_i \|P_i x - x\|^2 \tag{11}$$

and let G be the set of minimizers of f. It is clear that f is a convex function, since it is a positive combination of distances to closed convex sets. See e.g. [8, pp. 28, 32].

We proceed now to show that $F = G$. We need the following lemma:

LEMMA 4. *For any $x \in \mathbb{R}^n$,*

$$f(P_\alpha x) \leqslant f(x) - \left(\frac{2}{\alpha} - 1\right) \|P_\alpha x - x\|^2.$$

Proof. Since $P_i x$ is the closest point to x in C_i,

$$\begin{aligned} \|P_i P_\alpha x - P_\alpha x\|^2 &\leqslant \|P_i x - P_\alpha x\|^2 \\ &= \|P_i x - x\|^2 + \|x - P_\alpha x\|^2 - 2\langle P_i x - x, P_\alpha x - x \rangle. \end{aligned}$$

Summing on i,

$$\begin{aligned} f(P_\alpha x) &= \sum_{i=1}^{r} \lambda_i \|P_i P_\alpha x - P_\alpha x\|^2 \\ &\leqslant \sum_{i=1}^{r} \lambda_i \|P_i x - x\|^2 + \|x - P_\alpha x\|^2 - 2\langle Px - x, P_\alpha x - x \rangle \\ &= f(x) - \left(\frac{2}{\alpha} - 1\right) \|P_\alpha x - x\|^2. \qquad \blacksquare \end{aligned}$$

It follows that f is a descent function for the algorithm (6) (with $0 < \alpha < 2$) and, as a special case for $\alpha = 1$,

$$f(Px) \leqslant f(x) - \|Px - x\|^2. \tag{12}$$

Let $g_\alpha : \mathbb{R}^n \to \mathbb{R}_{\geqslant 0}$ be defined as $g_\alpha(x) = \|P_\alpha x - x\|^2$.

THEOREM 1. $F = G$.

Proof.

(i) $G \subset F$. Take $x \in G$, so $f(x) - f(Px) \leqslant 0$. From (12), $0 \leqslant g_1(x) \leqslant f(x) - f(Px)$. Therefore $g_1(x) = 0$ and $Px = x$.

(ii) $F \subset G$. Take $z \in F$, $x \in \mathbb{R}^n$. Assume, by negation, $f(x) < f(z)$. Consider the level set

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

A is closed and convex, because of the continuity and convexity of f. Let y^0 be the closest point to z in A. By (12), $f(Py^0) \leqslant f(y^0)$, i.e. $Py^0 \in A$. The definition of y^0 implies now $\|Py^0 - z\| \geqslant \|y^0 - z\|$. From Lemma 1, $\|Py^0 - z\| = \|Py^0 - Pz\| \leqslant \|y^0 - z\|$. So $\|Py^0 - z\| = \|y^0 - z\|$. Thus $Py^0 = y^0$. Therefore $\|y^0 - z\| = \|Py^0 - Pz\| \leqslant \sum_{i=1}^r \lambda_i \|P_i y^0 - P_i z\| \leqslant \|y^0 - z\|$ [using (8)]. Then $\|P_i y^0 - P_i z\| = \|y^0 - z\|$ for any i. From (8′) get $P_i z - z = P_i y^0 - y^0$. So $f(z) = f(y^0) \leqslant f(x)$, a contradiction. So $f(z) \leqslant f(x)$ for any $x \in \mathbb{R}^n$, that is to say $z \in G$. ■

Let now F_α be the set of fixed points of P_α. It is immediate, from the definition of P_α that $F_\alpha = F$ for any $\alpha > 0$. So we have

COROLLARY 1. *$F_\alpha = G$ for any $\alpha > 0$.*

Observe that if the sequence defined in (6) converges, the limit point belongs to F_α. We conclude that whenever the algorithm converges, it converges to a point which minimizes the weighted average (with the λ_i's) of the squares of the distances to the C_i's, i.e. a weighted least squares solution.

The next proposition shows that in the feasible case F is exactly the set of feasible points.

PROPOSITION 1. *If* $C = \bigcap_{i=1}^{r} C_i \neq \varnothing$, *then* $F = C$.

Proof. Obviously $C \subset F$. Take $z \in F$, $x \in C$. Since $x \in C$, $P_i x = x$ and $\|P_i x - P_j x\| = 0$ for any i, j. Now, as in Lemma 3,

$$0 = \langle x - Pz, z - Pz \rangle = \langle Px - Pz, z - Pz \rangle$$

$$\leqslant - \sum_{i=1}^{r} \sum_{j=i}^{r} \lambda_i \lambda_j \|P_i z - P_j z\|^2.$$

So $P_i z = P_j z$ for all i, j. Then $z = \sum_{i=1}^{r} \lambda_i P_i z = P_j z$ for any j, i.e., $z \in C_j$ for any j, implying $z \in C$. ∎

4. CONVERGENCE RESULTS

In this section we give convergence results for the algorithm defined in (6). The main one is that the algorithm converges for any initial point $x^0 \in \mathbb{R}^n$, whether the system (1) is consistent or not. From now on, let $\{x^k\}$ be the sequence defined by (6).

LEMMA 6. *If* $F \neq \varnothing$, *then* $\|x^k - z\|$ *decreases for any* $z \in F$, $x^0 \in \mathbb{R}^n$, *i.e.,* $\{x^k\}$ *is Fejer-monotone with respect to the set* F.

Proof.

$$\|x^{k+1} - z\|^2 = \|x^k - z\|^2 + \alpha^2 \|Px^k - x^k\|^2 + 2\alpha \langle x^k - z, Px^k - x^k \rangle$$

$$= \|x^k - z\|^2 + \alpha(\alpha - 2)\|Px^k - x^k\|^2$$

$$+ 2\alpha \langle Px^k - x^k, Px^k - z \rangle. \tag{13}$$

The second term of (13) is negative because $\alpha \in (0,2)$, and the third is nonpositive by Lemma 3. ∎

THEOREM 2. *If* $\{x^k\}$ *defined by* (6) *is bounded, then it converges for any* $x^0 \in \mathbb{R}^n$, *and* $F \neq \varnothing$.

Proof. If $\{x^k\}$ is bounded there exists a convergent subsequence

$$x^{k_j} \underset{j\to\infty}{\to} x.$$

So

$$g_\alpha(x^{k_j}) \underset{j\to\infty}{\to} g_\alpha(x).$$

By Lemma 4,

$$g_\alpha(x^{k_j}) \leqslant \frac{\alpha}{2-\alpha}\left[f(x^{k_j}) - f(x^{k_j+1})\right].$$

Since $f(x^k)$ is decreasing and bounded below by 0, $f(x^k) - f(x^{k+1})$ tends to 0. So $g_\alpha(x) = 0 \Rightarrow P_\alpha x = x \Rightarrow x \in F$. So $F \neq \varnothing$. Given $\epsilon > 0$, take j so that $\|x^{k_j} - x\| < \epsilon$. Now for any $m > k_j$, apply Lemma 6: $\|x^m - x\| \leqslant \|x^{k_j} - x\| < \epsilon$. So $x^k \to x$. ■

COROLLARY 2. *If $F \neq \varnothing$, then $\{x^k\}$ converges for any $x^0 \in \mathbb{R}^n$ to a point in F.*

Proof. Take $z \in F$. Then $\|x^k\| \leqslant \|x^k - z\| + \|z\| \leqslant \|x^0 - z\| + \|z\|$ (Lemma 6). So $\{x^k\}$ is bounded. Since g_α is continuous, the limit point belongs to $F_\alpha = F$. ■

COROLLARY 3. *If $C = \bigcap_{i=1}^r C_i \neq \varnothing$, then $\{x^k\}$ converges for any $x^0 \in \mathbb{R}^n$ to a point in C.*

Proof. Immediate from Corollary 2 and Proposition 1. ■

There is one last step to obtain a general convergence theorem for our algorithm: to prove that F is always nonempty.

PROPOSITION 2. $F \neq \varnothing$.

Proof. In view of (3) and (5) the function $P_i x - x$ is piecewise affine. In fact there are just two pieces: the two closed half spaces $\langle a_i, x\rangle \leqslant b_i$ and $\langle a_i, x\rangle \geqslant b_i$. So f is a piecewise quadratic function, and there is a covering of $\mathbb{R}^n$ by closed polyhedra (intersections of closed half spaces associated with the equations $\langle a_i, x\rangle = b_i$) such that f coincides with an affine function on each one of them. Since f is bounded below by 0, we apply Frank-Wolfe's theorem [9, Corollary 27.3.1] and conclude that f attains its minimum on each of these polyhedra. Since there are finitely many of them, f attains its global minimum. So $F = G \neq \varnothing$. ■

THEOREM 3. *For any starting point $x^0 \in R^n$ the sequence $\{x^k\}$ generated by* (6) *converges. If the system* (1) *is consistent, the limit point is a feasible point for* (1). *Otherwise, the limit point minimizes $f(x) = \sum_{i=1}^{r} \lambda_i \|P_i x - x\|^2$, i.e., it is a weighted (with the λ_i's) least squares solution to* (1).

Proof. Immediate from Corollaries 2 and 3 and Proposition 2.

5. THE CASE OF A VARIABLE RELAXATION PARAMETER

Consider now the following modification of the algorithm:

$$x^{k+1} = x^k + \alpha_k (Px^k - x^k) \tag{14}$$

with $\alpha_k \in (\epsilon_1, 2 - \epsilon_2)$, $\epsilon_1, \epsilon_2 > 0$. This algorithm has the same convergence properties as the algorithm defined by (6). In effect, substituting x^k for x, x^{k+1} for $P_\alpha x$, and α_k for α in the proof of Lemma 4, we get

$$f(x^{k+1}) \leqslant f(x^k) - \left(\frac{2}{\alpha_k} - 1\right)\|x^{k+1} - x^k\|^2, \tag{15}$$

where $\{x^k\}$ is defined by (14). Lemma 6 also holds for $\{x^k\}$ as in (14) with α_k substituting for α. We only need to establish Theorem 2 for this case.

LEMMA 8. *Theorem 2 holds for $\{x^k\}$ as in* (14).

Proof. Again there is a convergent subsequence

$$x^{k_j} \underset{j}{\to} x.$$

By (15) f is a descent function for x^k. So $f(x^k)$ is convergent and $\lim[f(x^k) - f(x^{k+1})] = 0$. Since

$$\|x^{k+1} - x^k\|^2 \leqslant \frac{\alpha_k}{2 - \alpha_k}\left[f(x^k) - f(x^{k+1})\right] \leqslant \frac{2}{2 - \epsilon_2}\left[f(x^k) - f(x^{k+1})\right],$$

it follows that $\lim\|x^k - x^{k-1}\| = 0$. By continuity of P

$$\lim_j Px^{k_j} = Px, \tag{16}$$

$$\|Px^{k_j} - x^{k_j}\| = \frac{1}{\alpha_k}\|x^{k_j+1} - x^{k_j}\| < \frac{1}{\epsilon_1}\|x^{k_j-1} - x^{k_j}\| \underset{j}{\to} 0. \tag{17}$$

From (16) and (17) conclude that $Px = x$. Hence $x \in F$ and $F \neq \varnothing$. Using Lemma 6, conclude as before that the whole sequence tends to x. ■

It follows that Corollaries 2 and 3, and therefore Theorem 3, also hold for the case of a variable relaxation parameter.

In [2] a more general relaxation scheme is considered, where the parameter, which depends on x^k, may be greater than 2. We conjecture that our proof for the infeasible case may be extended also to this algorithm.

REFERENCES

1 Y. Censor, Row-action methods for huge and sparse systems and their applications, *SIAM Rev.* 23:444–466 (1981).
2 Y. Censor and T. Elfving, New methods for linear inequalities, *Linear Algebra Appl.* 42:199–211 (1982).
3 G. Cimmino, Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari, in *La Ricerca Scientifica*, Roma, 1938, XVI, Ser. II. Anno IX, 1:326–333.
4 R. W. Cottle and Y. S. Pang, On solving linear complementarity problems as linear programs, *Math. Programming Stud.* 7:88–107 (1978).
5 N. Gastinel, *Analyse Numérique Linéaire*, Hermann, Paris, 1966.
6 G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic, New York, 1980.
7 O. L. Mangasarian, Linear complementarity problems solvable by a single linear programm, *Math. Programming* 10:263–270 (1976).
8 M. N. Nashed, 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. Natterer, Eds.), Lectures Notes in Medical Informatics, 8:160–178, Springer, Berlin, 1981.
9 R. T. Rockafeller, *Convex Analysis*, Cambridge U. P., 1970.

Received 19 September 1983; revised 2 February 1984