Alfredo Noel Iusem; Álvaro Rodolfo De Pierro

On the set of weighted least squares solutions of systems of convex inequalities

Commentationes Mathematicae Universitatis Carolinae, Vol. 25 (1984), No. 4, 667--678

Persistent URL: `http://dml.cz/dmlcz/106333`

COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE
25,4 (1984)

ON THE SET OF WEIGHTED LEAST SQUARES SOLUTIONS OF SYSTEMS OF CONVEX INEQUALITIES

A. IUSEM, A. DE PIERRO*

Abstract: This paper studies the set of fixed points of a convex combination of projections on m fixed convex sets, or equivalently the set of weighted least squares solutions of a system of convex inequalities. It is proved that such set is the intersection of translates of the convex sets and that its interior is empty when the convex sets have empty intersection. For the case of a system of linear inequalities, the behavior of the set as a function of the right hand side and the coefficients of the convex combination is discussed.

Key words and phrases: Linear inequalities, convex inequalities, iterative algorithms for linear systems.

Classification: 52A05, 65F10, 90C25

---

1. Introduction

Let $C_1, C_2, \dots, C_m$ be closed convex sets in a Hilbert space H. Let $P_i : H \to C_i$ be the projection over C_i (i.e. $P_i x = \arg\min_{z \in C_i} \|x-z\|$). Let $S = \{\lambda \in \mathbb{R}^m$ s.t. $\sum_{i=1}^{m} \lambda_i = 1,\ \lambda_i > 0$ $(1 \le i \le m)\}$. Take $\lambda \in S$ and define $P: H \to H$ as

$$P_x = \sum_{i=1}^{m} \lambda_i P_i x \qquad (1)$$

---

*The work of this author was partially supported by CNPq, under grant nº 301699/81.

Let $F(\lambda)$ be the set of fixed points of P:

$$F(\lambda) = \{x \in H : Px=x\} \tag{2}$$

Consider now the function $f_\lambda : H \to \mathbb{R}$ defined by

$$f_\lambda(x) = \sum_{i=1}^{m} \lambda_i \|P_i x - x\|^2 \tag{3}$$

and let $G(\lambda)$ be the set of minimizers of f_λ. Let $C = \bigcap_{i=1}^{m} C_i$. In [3] it was proved that

i) $F(\lambda) = G(\lambda) \quad \forall \lambda \in S$ (4)

ii) If $C \neq \emptyset$, then $F(\lambda) = G(\lambda) = C \quad \forall \lambda \in S$ (5)

iii) If $z_1, z_2 \in F(\lambda)$ then $P_i z_1 - z_1 = P_i z_2 - z_2 \; (1 \leq i \leq m)$ (6)

The set $F(\lambda)$, in view of (4), can be seen as the set of weighted (with the λ_i's) least squares solutions to the problem of finding a point in C. An important particular case results when the sets C_i are of the form

$C_i = \{x: g_i(x) \leq 0\}$ with $g_i : H \to \mathbb{R}$ convex and continuous. (7)

In this case, the task of finding a point in C is equivalent to solving the system

$$g_i(x) \leq 0 \qquad (1 \leq i \leq m) \tag{8}$$

In [2,3,4] it was shown that several iterative algoritms for solving problems such as (8) have the property that they converge (whenever the set $F(\lambda)$ is non empty) to a point in $F(\lambda)$ i.e. to a solution of (8) when it is feasible and to a weighted (with the λ_i's) least squares solution of (8) otherwise.

These algorithms, which fall under the category of "row action methods" introduced by Censor [1], are widely used in practice for applications in the area of computerized tomography and image reconstruction from projections [5,6]. An example of such algorithms consists in taking an arbitrary $x^0 \in H$ and defining

$$x^{k+1} = Px^k.$$

Thus, the study of the sets $F(\lambda)$ has interesting consequences on the understanding of the behaviour of these iterative algorithms.

In section 2 of this paper two results are established, namely:

i) The set $F(\lambda)$ is the intersection of translates of the sets $C_1, \dots, C_m$.

ii) If $C = \emptyset$, the set $F(\lambda)$ has empty interior.

In section 3 we consider the case when the functions g_i of (7) are affine and $H = \mathbb{R}^n$. In this case (8) becomes

$$Ax \le b \tag{9}$$

with $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. We are interested in the behaviour of $F(\lambda)$ as a function of b. The main result is the following: If b is replaced by $\bar{b} < b$ (i.e. $\bar{b}_i < b_i$ for all i) then for each $\lambda \in S$ there is a $\mu \in S$ such that $F(\mu)$ for the problem with $\bar{b}$ is contained in $F(\lambda)$ for the problem with b.

2. Some results on $F(\lambda)$ for general convex sets

We start with a lemma related to the formulation given by (7), i.e. we consider convex sets of the form $C_i = \{x: g_i(x) \le 0\}$ with g_i convex.

Lemma 1. Assume $F(\lambda) \neq \phi$. Take $z \in F(\lambda)$. Define $y_i =$ $= z-P_i z$. Then

$$F(\lambda) = \{x \in H: g_i(x-y_i) \leq 0\}.$$

Proof: First observe that, by virtue of (6), the vector y_i is independent of the chosen point $z \in F(\lambda)$

i) $\subset$). Take $z \in F(\lambda)$

$$y_i = z-P_i z \Rightarrow z-y_i = P_i z \Rightarrow g_i(z-y_i) = g_i(P_i z) \leq 0$$

because $P_i z \in C_i$

ii) $\supset$) Take x such that $g_i(x-y_i) \leq 0$ $(1 \leq i \leq m)$. So:

$$x-y_i \in C_i \Rightarrow \|x-P_i x\|^2 \leq \|x-(x-y_i)\|^2 = \|y_i\|^2 = \|z-P_i z\|^2.$$

From the definition of f_λ it follows that $f_\lambda(x) \leq f_\lambda(z)$. Applying (4) and the definition of $G(\lambda)$ conclude that $x \in G(\lambda) = F(\lambda)$. ∎

Now, given a vector $y \in H$, define $C_i+y = \{x \in H:$ $x = x_i+y$ with $x_i \in C_i\}$. i.e. C_i+y is the translate of C_i by the vector y. We go back to the original formulation where C_i is just a closed convex set in H.

Theorem 1. There exist vectors $y_i \in H$ $(1 \leq i \leq m)$ such that $F(\lambda) = \bigcap_{i=1}^{m} (C_i+y_i)$. The vectors y_i can be taken as $y_i = z-P_i z$ where z is any vector in $F(\lambda)$.

Proof: Consider the functions $g_i(x) = \|P_i x-x\|$. Being distances to closed convex sets, the functions g_i are convex (see [7, pp.28,32]). Since $g_i(x) \leq 0$ iff $x \in C_i$ we conclude that $C_i = \{x: g_i(x) \leq 0\}$. Apply Lemma 1:

$F(\lambda) = \{x \in H: \|x-y_i - P_i(x-y_i)\| \leq 0\} = \{x \in H: x-y_i = P_i(x-y_i)\}$ $= \{x \in H: x-y_i \in C_i\} = \bigcap_{i=1}^{m} (C_i+y_i)$. The second statement of the theorem also follows from Lemma 1. ∎

Let, for a set $B \subset H$, $\mathring{B}$ denote the interior of B.

Theorem 2. $C = \emptyset \Rightarrow F^o(\lambda) = \emptyset \quad \forall \lambda \in S$.

Proof: Suppose $F^o(\lambda) \neq \emptyset$. Take $z_1 \in F^o(\lambda)$. So $\exists \epsilon > 0$ such that

$$\|z-z_1\| < \epsilon \Rightarrow z \in F(\lambda). \tag{10}$$

Since $C = \emptyset$, $\exists j$ such that $z_1 \notin C_j \Rightarrow P_j z_1 - z_1 \neq 0$. (11)

Let $\gamma = \min \{1, \frac{\epsilon}{\|P_j z_1 - z_1\|}\}$. Take any $\beta \in (0,\gamma)$. (12)

Define $z_2 = z_1 + \beta(P_j z_1 - z_1)$. So

$$\|z_2-z_1\| = \beta\|P_j z_1 - z_1\| < \epsilon \Rightarrow z_2 \in F(\lambda) \quad \text{(from (10))}.$$

On the other hand, since z_2 lies in the segment between z_1 and $P_j z_1$, $P_j z_2 = P_j z_1 \Rightarrow z_2 - P_j z_2 = (1-\beta)(z_1 - P_j z_1) \neq$ $\neq z_1 - P_j z_1$ (using (11) and (12)). This contradicts (6), so $F^o(\lambda) = \emptyset$. ∎

Corollary 1. Consider a system like (8) with g_i strictly convex ($1 \le i \le m$) and continuous. If the system is infeasible and $F(\lambda) \neq \emptyset$, then $F(\lambda)$ is a singleton, i.e. there is a unique weighted least squares solution of (8).

Proof. Take $v,w \in F(\lambda)$. Applying Lemma 1,

$$g_i(v-y_i) \le 0, \quad g_i(w-y_i) \le 0 \qquad (1 \le i \le m)$$

If $v \neq w$, $g_i(\frac{v+w}{2} - y_i) < 0$ $(1 \le i \le m)$, because of strict convexity.

Since g_i is continuous, for each i, there is a neighborhood of $\frac{v+w}{2}$ contained in $C_i + y_i$. From Theorem 1, $\frac{v+w}{2} \in F^o(\lambda)$, in contradiction with Theorem 2. So $v = w$.

If H is finite dimensional, the hypothesis that $F(\lambda) \neq \emptyset$ is redundant. A strictly convex function g in finite dimension has the property that $\lim_{\|x\|\to\infty} g(x) = \infty$. So $\{x: g(x) \le 0\}$ is bounded. Then all the sets C_i are bounded. In [3] it was proved that if at least one of the C_i's is bounded, $F(\lambda) \neq \emptyset$.

3. Some results on $F(\lambda)$ in the linear case

Consider now a system like (9). Take $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$. Let $a^i \neq 0$ $(1 \le i \le m)$ be the rows of A. So $C_i = \{x \in \mathbb{R}^n: \langle a^i, x\rangle \le b_i\}$ and $C = \{x \in \mathbb{R}^n: Ax \le b\}$.

Let us perturbate the right hand side b to $\bar{b} = b - \varepsilon$ $(\varepsilon \in \mathbb{R}^m,\ \varepsilon \ge 0)$. We are interested in the behaviour of the set $F(\lambda)$ as a function of ε. If $\bar{P}$ is the operator P with $\bar{b}$ substituting for b (same for $\bar{P}_i$) let $F(\lambda,\varepsilon)$ be the set of fixed points of $\bar{P}$ and $C(\varepsilon) = \{x \in \mathbb{R}^n: Ax \le b-\varepsilon\}$. With this notation $F(\lambda)$ becomes $F(\lambda,0)$ and C becomes $C(0)$.

It is clear that if $C(\varepsilon) \neq \emptyset$. $C(\varepsilon) \subset C(0)$. It follows from (5) that in such a case

$$\forall\ \lambda,\mu \in S \quad F(\mu,\varepsilon) \subset F(\lambda,0) \tag{13}$$

We want to extend this result to the case $C(\varepsilon) = \emptyset$. In this case, an arbitrary $\mu \in S$ will not satisfy (13). In fact, it will be shown that given λ and ε there exists a $\mu \in S$ (in general depending on λ and ε) which makes (13) true.

We start with another characterization of the set $F(\mu,\varepsilon)$. For $x \in \mathbb{R}^n$ define $x^+ \in \mathbb{R}^n$ as

$$x_i^+ = \begin{cases} x_i & \text{if} \quad x_i \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The projection P_i on the half space C_i has the well known formula

$$P_i x = x - \frac{(\langle a^i, x\rangle - b_i)^+}{\|a^i\|^2} a^i \tag{14}$$

From (4):

$$F(\mu,\varepsilon) = \arg\min_x \sum_{i=1}^m \mu_i \|\bar{P}_i x - x\|^2 =$$

$$= \arg\min_x \sum_{i=1}^m \frac{\mu_i}{\|a^i\|^2} [(\langle a^i, x\rangle - b_i + \varepsilon_i)^+]^2 .$$

This minimization problem is equivalent to

$$\min_{x,y} \sum_{i=1}^m \frac{\mu_i}{\|a^i\|^2} y_i^2 \quad , \qquad \text{s.t.} \quad Ax \leq b - \varepsilon + y, \quad y \geq 0 \tag{15}$$

The feasible set of (15) is non empty, because the system is feasible for big enough y. So (15) consists in the minimization of a quadratic function bounded below on a polyhedron. Frank-Wolfe's theorem (see [7, Cor. 27.3.1]) insures the existence of a solution. Because the minimand is strictly convex in y, the y part of the solution is unique. Let us call it $y(\mu,\varepsilon)$. It follows that:

Proposition 1

i) $F(\mu,\varepsilon) \neq \emptyset \quad \forall \mu \in S, \ \varepsilon \geq 0$

ii) $F(\mu,\varepsilon) = \{x \in \mathbb{R}^n : Ax \leq b - \varepsilon + y(\mu,\varepsilon)\}$ where $y(\mu,\varepsilon)$ is the solution of (15).

Let Q be the projection of the feasible set of (15) on the y coordinates, i.e.

$$Q = \{y \in \mathbb{R}^m,\ y \geq 0 \text{ and } \exists x \text{ s.t. } Ax \leq b+y-\epsilon\}.$$

Take $y \in Q$, $\bar{y} \geq y$. Clearly any x feasible for (15) remains feasible if $\bar{y}$ substitutes for y. We rephrase this fact as

Proposition 2. $y \in Q,\ \bar{y} \geq y \Rightarrow \bar{y} \in Q$.

Lemma 2. $\exists\ D \in \mathbb{R}^{s\times m}$, $D \geq 0$ and $c \in \mathbb{R}^s$ (for some s) such that

$$Q = \{y \in \mathbb{R}^m\colon Dy \geq c,\ \ y \geq 0\}. \tag{16}$$

Proof: The feasible set of (15) is a polyhedron in $\mathbb{R}^{m+n}$. So its projection Q is a polyhedron in $\mathbb{R}^m$ (see [7,Th.19.3]) i.e. $Q = \{y \in \mathbb{R}^m\colon Dy \geq c,\ y \geq 0\}$ for some D, c. We still need to show that $D \geq 0$. If some entry d_{ij} were negative take any $y \in Q$ and define $\bar{y} = y+Me^j$ where $e^j \in \mathbb{R}^m$ is defined as $e^j_j = 1$, $e^j_i = 0$ for $i \neq j$. $\bar{y} \in Q$ for all $M \geq 0$ because of Proposition 2, but the i-th constraint is violated for big enough M. ∎

We need some results on systems of inequalities like (16). Let $T \in \mathbb{R}^{s\times m}$ with rows t^i and entries $t_{ij} \geq 0$. Let $E = \{z \in \mathbb{R}^m\colon Tz \geq u,\ \ z \geq 0\}$. For $\nu \in \mathbb{R}^m$, $\nu > 0$ consider the problem

$$\begin{aligned} &\min \sum_{i=1}^{m} \nu_i z_i^2 \\ &\text{s.t.}\ \ Tz \geq u \\ &\qquad\ \ z \geq 0 \end{aligned} \tag{17}$$

Assume (17) is feasible. Again, by strict convexity and Frank-Wolfe's theorem, (17) has a unique solution $z(\nu)$. Take any $z^o \in E$, $z^o > 0$. Let $\bar{E} = \{z \in E\colon z \leq z^o\}$ and

$$z^* = \arg\min_{z\in\bar{E}} \|z\| \tag{18}$$

Let $I = \{i: \langle t^i, z^* \rangle = u_i\}$, $J = \{j: z_j^* = 0\}$, $K = \{j: z_j^* > 0\}$.

Lemma 3. $\forall\, k \in K \;\exists\, i \in I$ s.t. $t_{ik} > 0$ and $t_{ij} = 0 \;\forall_j \in J$.

Proof: Take $k \in K$. Certainly there exists $i \in I$ such that $t_{ik} > 0$. Otherwise $\bar{z} = z^* - \eta e^k$ belongs to $\bar{E}$ for small enough η in contradiction with (18). Let $\bar{I}_k = \{i \in I: t_{ik} > 0\}$. Assume, by negation, that

$$\forall\, i \in \bar{I}_k \;\exists\, j \in J \text{ s.t. } t_{ij} > 0 \tag{19}$$

Define

$$\sigma = \min_{1 \le j \le m} \{z_j^o\} \tag{20}$$

$$\xi = \min_{i \in \bar{I}} \left\{ \frac{\sum_{j \in J} t_{ij}}{t_{ik}} \right\} \tag{21}$$

$$\delta = \min \{z_k^*, \sigma\xi\}$$

$\delta > 0$ because of (19). Define $\hat{z}$ as

$$\hat{z}_j = \begin{cases} z_j^* & \text{if } j \in K - \{k\} \\ \sigma & \text{if } j \in J \\ z_k^* - \delta & \text{if } j = k \end{cases}$$

From (21) $\hat{z} \in E$ (in the system $Tz \ge u$ the increase in the columns in J is greater than the decrease in the k-th column). From (20) $\hat{z} \in \bar{E}$. So $\tilde{z} = z^* + \alpha(\hat{z} - z^*) \in E$ for $\alpha \in (0,1)$. If r is the cardinal of J, $\|\tilde{z}\|^2 = \|z^*\|^2 - \alpha[2\delta z_k^* - \alpha(r\sigma^2 + \delta^2)]$. Hence $\|\tilde{z}\| < \|z^*\|$ for $\alpha < \dfrac{2\delta z_k^*}{r\sigma^2 + \delta^2}$ in contradiction with (18).

Lemma 4. Given $z^o \in E$, $z^o > 0$. $\exists\, \nu \in \mathbb{R}^m$, $\nu > 0$ such that $z(\nu) \le z^o$.

Proof: Take z^* as in (18). Use Lemma 3 to select, for each $k \in K$, a row $i(k) \in I$ such that $t_{i(k),k} > 0$ and $t_{i(k),j} = 0$

$\forall\ j \in J$. Let $L = \{i(k): k \in K\}$. Let $\rho = \sum_{\ell\in L} t^{\ell}$. By construction $\rho_j = 0$ if $j \in J$ and $\rho_j > 0$ if $j \in K$. Also

$$\forall\ z \in E \qquad \langle \rho, z^* \rangle = \sum_{\ell\in L} u_\ell \le \langle \rho, z \rangle \tag{22}$$

Define

$$\nu_j = \begin{cases} \dfrac{\rho_j}{z_j^*} & \text{if} \quad j \in K \\ 1 & \text{if} \quad j \in J \end{cases}$$

Take any $z \in E$. From (22)

$$\langle \rho, z^* \rangle \le \langle \rho, z \rangle \Rightarrow \sum_{j\in K} \nu_j z_j^{*2} \le \sum_{j\in K} \nu_j z_j^* z_j \Rightarrow$$

$$2 \sum_{j=1}^{m} \nu_j z_j^{*2} \le 2 \sum_{j=1}^{m} \nu_j z_j^* z_j \le \sum_{j=1}^{m} \nu_j (z_j^{*2} + z_j^2) \Rightarrow$$

$$\sum_{j=1}^{m} \nu_j z_j^{*2} \le \sum_{j=1}^{m} \nu_j z_j^2 .$$

So for such ν, $z^* = z(\nu)$. Since $z^* \in E$ the lemma is established. ∎

We prove now the main result of this section:

__Theorem 3__. $\forall\ \lambda \in S$, $\varepsilon > 0$ $\exists\ \mu \in S$ such that $F(\mu, \varepsilon) \subset \subset F(\lambda, 0)$.

__Proof__: By Proposition 1.ii)

$$F(\lambda, 0) = \{x \in \mathbb{R}^n: Ax \le b + y(\lambda, 0)\}. \tag{25}$$

Also $y(\lambda, 0) + \varepsilon$ is a feasible y for system (15); i.e. $y(\lambda,0)+\varepsilon \in Q$. Since $\varepsilon > 0$, $y(\lambda,0)+\varepsilon > 0$ and we may apply Lemma 4 with $z^0 = y(\lambda, 0) + \varepsilon$, $T = D$ and $u = c$. Conclude that there exists $\nu > 0$ such that the solution $z(\nu)$ of $\min_{z\in Q} \sum_{j=1}^{m} \nu_j z_j^2$ satisfies

$$z(\nu) \le y(\lambda, 0) + \varepsilon . \tag{26}$$

Take now

$$\mu_i = \frac{\nu_i \|a^i\|^2}{\sum_{j=1}^{m} \nu_j \|a^j\|^2}$$

So $\mu \in S$ and $z(\nu)$ solves $\min_{z \in Q} \sum_{i=1}^{m} \frac{\mu_i z_i^2}{\|a^i\|^2}$. It follows from Proposition 1.ii) that

$$F(\mu,\epsilon) = \{x \in \mathbb{R}^n : Ax \leq b - \epsilon + z(\nu)\}. \tag{27}$$

Take any $x \in F(\mu,\epsilon)$. From (26) and (27) $Ax \leq b + y(\lambda,0)$. From (25) $x \in F(\lambda,0)$. So $F(\mu,\epsilon) \subset F(\lambda,0)$. ∎

Geometrically, the theorem states that by a suitable change of weights the set of weighted least squares solutions of the tighter perturbed problem is included in the set of weighted least squares solutions of the original one, extending the inclusion relationship between the feasible sets of both problems. Observe that the result also holds when $C(\epsilon) = \emptyset$ and $C(0) \neq \emptyset$. In such a case the vector μ of the theorem does not depend on λ (only on ϵ), since $F(\lambda,0) = C(0)$ $\forall \lambda \in S$.

References

[1] Y. CENSOR: Row-action methods for huge and sparse systems and their application. <u>SIAM Rev</u>. Vol. 23: 444-466 (1981).

[2] A. DE PIERRO and A.N. IUSEM: A Simultaneous Projections Method for Linear Inequalities. To be published in Linear Algebra and its Applications.

[3] A. DE PIERRO and A.N. IUSEM: A Parallel Projection Method for Finding a Common Point of a Family of Convex Sets. Informes de Matemática, Série B-021/84, IMPA, Rio de Janeiro, (1984).

[4] A. DE PIERRO and A.N. IUSEM: A Simultaneous Iterative Method for Computing Projections on Polyhedra. To be published.

[5] G.T. HERMAN: Image Reconstruction from Projections: The Fundamentals of Computerized Tomography. Academic Press, New York, 1980.

[6] G.T. HERMAN and A. LENT: Iterative Reconstruction Algorithms. Comput. Biol. Med. 6: 273-294 (1976).

[7] R.J. ROCKAFELLER: Convex Analysis. Cambridge Univ. Press. Cambridge, Mass. (1970).

IMPA - Estrada Dona Castorina 110
Jardim Botânico - Rio de Janeiro - Brasil

Instituto de Matemática - UFRJ
Rio de Janeiro, Brasil.

(Oblatum 10.5. 1984)