JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **190**, 129–141 (1995)

Pareto-Optimality in Linear Regression

E. CARRIZOSA, E. CONDE, F. R. FERNANDEZ, M. MUÑOZ, AND J. PUERTO

Facultad de Matemáticas, Tarfia s/n, 41012 Seville, Spain

Submitted by L. A. Baxter

Received March 2, 1993; revised February 7, 1994

In this paper the linear regression problem is studied in the context of vector optimization theory. The set of Pareto-optimal solutions is represented as the set of optimal solutions to certain optimization problems, and is geometrically characterized as a finite union of bounded polyhedra. © 1995 Academic Press, Inc.

1. INTRODUCTION

Let $\mathbb{O} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, each x_i being a $d \times 1$ matrix ($d \geq 2$) and each $y_i \in \mathbb{R}$. In the linear regression model, it is assumed that $\mathbb{O}$ is a sample drawn from a certain multivariate random variable (X, Y) in $\mathbb{R}^d \times \mathbb{R}$, such that the *response* variable Y is a linear function of the *predictive* variable X, affected by a perturbation term ε. In other words, it is assumed that $\mathbb{O}$ is a sample from a multivariate random variable (X, Y) that verifies

$$Y = X'\beta + \varepsilon,$$

where β is an unknown parameter and $X'\beta$ denotes the usual scalar product in $\mathbb{R}^d$. The aim is to determine

(i) The parameter β^* such that the hyperplane h_{β^*}: $y = x'\beta^*$ gives the *best* fit to the sample.

(ii) For each $x \in \mathbb{R}^d$, the value $x'\beta^*$ (*prediction* at x), where β^* is the parameter obtained in (i).

The set $\mathbb{O}$ induces a vector function (*error function*) ε: $\mathbb{R}^d \to \mathbb{R}$,

0022-247X/95 $6.00

$$\varepsilon(\beta) = (\varepsilon_1(\beta), \ldots, \varepsilon_n(\beta)) = (|y_1 - x_1'\beta|, \ldots, |y_n - x_n'\beta|),$$

from which we want to determine the value β^* for β that minimizes (in some sense) the vector function ε and, for each $x \in \mathbb{R}^d$, the prediction $x'\beta^*$.

This problem, seeking the simultaneous minimization of the errors ε_i $(i = 1, \ldots, n)$ is in nature a vector optimization problem, which can be stated as

$$\text{(VOP)} \min_{\beta \in \mathbb{R}^d} \varepsilon(\beta),$$

where the minimization must be understood in the vectorial sense (see, e.g., [3, 16, 17]).

Given the vectorial essence of problem (VOP), the most commonly used strategy to face problem (i) has consisted of scalarizing (VOP) through a certain globalizing function ϕ.

Hence, instead of "solving" (VOP), one solves a scalarized version (P_ϕ),

$$(\mathrm{P}_\phi) \min_{\beta \in \mathbb{R}^d} \phi(\varepsilon(\beta))$$

For $\phi = \|\cdot\|_2$ (the Euclidean norm), (P_ϕ) leads to the least sum of squares regression, which, although the most popular with researchers, is not the only globalizing function in the literature (see, e.g., [1, 4, 10]): one should mention, among others, $\phi = \|\cdot\|_1$ (least absolute deviations regression), $\phi = \|\cdot\|_\infty$ (Chebychev regression), $\phi = \|\cdot\|_p$, $1 < p < \infty$ (L_p regression [10]), convex combinations of the former [11], and ϕ = median [15]. It is intuitively clear that these different approaches lead to different optimal parameters β. What is not so evident is the range of variation for this set of optimal parameters. In this paper we address this question by studying the linear regression problem through the vector optimization problem (VOP) defined above.

First (in Section 2) we discuss the concept of a Pareto-optimal solution, and the set $\mathcal{S}$ of Pareto-optimal solutions to (VOP) is identified with a set of optimal solutions to optimization problems.

Section 3 is devoted to a geometrical characterization of $\mathcal{S}$ in terms of the *elementary convex sets* associated with (VOP). The results obtained also enable determination of the set of *optimal predictions*, discussed in Section 4.

2. Pareto-Optimal Solutions

Given a family $\{f_a : a \in A\}$ of functions f_a: $\Omega \subset \mathbb{R}^p \to \mathbb{R}$ and a set $Z \subset \Omega$, a point $z^* \in Z$ is said to be a *Pareto-optimal* solution to problem $\min_{z\in Z} (f_a(z))_{a\in A}$ iff there exists no $z \in Z$ such that

$$f_a(z) \leq f_a(z^*) \text{ for all } a \in A, \qquad f_a(z) < f_a(z^*) \text{ for some } a \in A.$$

Consider the vector optimization problem (VOP),

$$\text{(VOP)} \min_{\beta\in\mathbb{R}^d} \varepsilon(\beta),$$

and denote by $\mathcal{S}$ the set of Pareto-optimal solutions to (VOP).

In this section we state some properties of $\mathcal{S}$ that allow a representation of $\mathcal{S}$ as a set of optimal solutions to problems of the form $\min_{\beta\in\mathbb{R}^d} \phi(\varepsilon(\beta))$ when ϕ varies in a certain set $\mathcal{F}$ of functions.

For the sake of simplicity, throughout this paper the following assumption is made:

A: The matrix $X = (x_1, \ldots, x_n)$ has rank d.

Proposition 1. *For each $t = (t_1, t_2, \ldots, t_n) \in \mathbb{R}^n$, the level set $C = \{\beta \in \mathbb{R}^d\colon \varepsilon_i(\beta) \leq t_i \ \forall\, i = 1, \ldots, n\}$ is compact.*

Proof. Given $t \in \mathbb{R}^n$, the result is trivial if t has at least a negative component; hence, we can assume that $t_i \geq 0$ for all $i = 1, \ldots, n$. As each ε_i is a continuous function, it follows that C is closed. To prove that C is bounded, it suffices to show that C has no direction of recession (see [14, Theorem 8.4]). Suppose that there exist $\beta \in C$, $\overline{\beta} \neq 0$ such that $\beta + \lambda\overline{\beta} \in C$ for all $\lambda \geq 0$. Then, for every i, $|y_i - x_i' \beta - \lambda x_i'\overline{\beta}| \leq t_i$ for all $\lambda \geq 0$. Hence, $x_i'\overline{\beta} = 0$ for all i, thus $\overline{\beta} = 0$ (by assumption A). This is a contradiction with $\overline{\beta} \neq 0$, thus the result holds. ∎

The next proposition shows that $\mathcal{S}$ *dominates* $\mathbb{R}^d$, in the sense that, for any parameter $\beta \in \mathbb{R}^d$, there exists $\beta^* \in \mathcal{S}$ such that the hyperplane h_β^*: $y = x'\beta^*$ gives a better or equal (componentwise speaking) fit to $\mathbb{C}$ than the hyperplane h_β: $y = x'\beta$.

Proposition 2. *for all $\beta \in \mathbb{R}^d$ there exists $\beta^* \in \mathcal{S}$ such that*

$$\varepsilon_i(\beta^*) \leq \varepsilon_i(\beta) \qquad \forall\, i = 1, \ldots, n,$$

with at least one inequality strict if $\beta \notin \mathcal{S}$.

Proof. Define on $\mathbb{R}^d$ the quasiorder R, $\beta_1 R \beta_2$ iff $\varepsilon_i(\beta_1) \leq \varepsilon_i(\beta_2)$ $\forall i = 1, \ldots, n$. By the proposition above, the set $\{\beta \in \mathbb{R}^d: \beta R \beta^*\}$ is compact for all $\beta^* \in \mathbb{R}^d$. The result then follows by using Theorem 2.3.6 of [17]. ∎

The proposition above enables the statement of *localization theorems*, which are completely analogous to those obtained in [13] for the location parameter:

A function $\phi: \mathbb{R}^n \to \mathbb{R}$ is said to be *nondecreasing* iff

$$\phi(u) \leq \phi(v) \qquad \forall\, u, v \in \mathbb{R}^n \text{ such that } u_i \leq v_i\ \forall i$$

A function $\phi: \mathbb{R}^n \to \mathbb{R}$ is said to be *strictly increasing* iff

$$\phi(u) < \phi(v) \qquad \forall u, v \in \mathbb{R}^n \text{ such that } u_i \leq v_i\ \forall i,\ u_i < v_i \text{ for some } i.$$

PROPOSITION 3. (i) *If $\phi: \mathbb{R}^n \to \mathbb{R}$ is a strictly increasing function, then $\mathcal{S}$ contains all the optimal solutions to problem* (Pϕ),

$$\text{P}\phi\text{:} \min_{\beta \in \mathbb{R}^d} \phi(\varepsilon(\beta)).$$

(ii) *If $\phi: \mathbb{R}^n \to \mathbb{R}$ is lower semicontinuous and nondecreasing, then $\mathcal{S}$ contains at least one optimal solution to* Pϕ.

Proof. (i) Assume that there exists $\beta \notin \mathcal{S}$ such that β is an optimal solution to (Pϕ). By Proposition 2, there exists $\beta^* \in \mathcal{S}$ such that

$$\varepsilon_i(\beta^*) \leq \varepsilon_i(\beta) \text{ for all } i; \qquad \varepsilon_i(\beta^*) < \varepsilon_i(\beta) \text{ for some } i.$$

As ϕ is strictly increasing,

$$\phi(\varepsilon_1(\beta^*), \ldots, \varepsilon_n(\beta^*)) < \phi(\varepsilon_1(\beta), \ldots, \varepsilon_n(\beta)),$$

which contradicts the optimality of β.

Hence, $\mathcal{S}$ contains *all* the optimal solutions to Pϕ.

(ii) As, by Proposition 1, $\varepsilon(\cdot)$ has compact level sets, the lower semicontinuity of ϕ implies that (P_ϕ) has an optimal solution. In other words, there exists $\beta^1 \in \mathbb{R}^d$ such that

$$\phi(\varepsilon(\beta^1)) \leq \phi(\varepsilon(\beta)) \qquad \text{for all } \beta \in \mathbb{R}^d.$$

By Proposition 2, there exists $\beta^2 \in \mathcal{S}$ such that $\varepsilon_i(\beta^2) \leq \varepsilon_i(\beta^1)$ for all i. As ϕ is nondecreasing, it follows that β^2 is Pareto-optimal and also an optimal solution to (P_ϕ). ∎

Hence, as soon as ϕ is nondecreasing and lower semicontinuous, the search of optimal solutions to (P_ϕ) can (or should) be restricted to the points in $\mathcal{S}$. This is of particular interest when (P_ϕ) may have several optimal solutions (as occurs, e.g., in least absolute deviations or least-median regression). In such cases, it would be clear that, among all the optimal solutions to (P_ϕ), one should choose one which is also Pareto-optimal.

In order to gain insight into the problem, observe that the error function ε is a convex piecewise linear vector function, thus (VOP) can be transformed into an equivalent linear vector optimization problem. Indeed, let (Q) be the problem

$$
\begin{aligned}
\text{(Q)}\ \min\ & t = (t_1, \ldots, t_n) \\
& t_i \geq y_i - x_i' \beta \qquad \text{for all } i \\
& t_i \geq -(y_i - x_i' \beta) \qquad \text{for all } i \\
& (\beta, t) \in \mathbb{R}^d \times \mathbb{R}^n,
\end{aligned}
$$

and denote by $\mathcal{S}^0$ the set of Pareto-optimal solutions to (Q).

Let $\mathcal{S}^1 = \{\beta \in \mathbb{R}^d \colon (\beta, t) \in \mathcal{S}^0 \text{ for some } t \in \mathbb{R}^n\}$; i.e., $\mathcal{S}^1$ is the projection onto $\mathbb{R}^d$ of $\mathcal{S}^0$. One has

PROPOSITION 4. $\mathcal{S} = \mathcal{S}^1$.

Proof. We first show that $\mathcal{S} \subset \mathcal{S}^1$, by showing that $(\beta, \varepsilon(\beta)) \in \mathcal{S}^0$ for all $\beta \in \mathcal{S}$. Indeed, let $\beta \in \mathcal{S}$, and suppose that $(\beta, \varepsilon(\beta)) \notin \mathcal{S}^0$; then there exists (β^*, t^*) such that

$$\varepsilon(\beta^*) \leq t^* \leq \varepsilon(\beta) \qquad \text{and} \qquad t^* \neq \varepsilon(\beta).$$

Hence, $\varepsilon(\beta^*) \leq \varepsilon(\beta)$, and $\varepsilon(\beta^*) \neq \varepsilon(\beta)$, which implies that β is not Pareto-optimal to (VOP), which is a contradiction. Hence, $(\beta, \varepsilon(\beta)) \in \mathcal{S}^0$ which implies that $\beta \in \mathcal{S}^1$ for all $\beta \in \mathcal{S}$.

Conversely, let $\beta \in \mathcal{S}^1$; by definition, there exists $t \in \mathbb{R}^n$ such that $\varepsilon(\beta) \leq t$ and $(\beta, t) \in \mathcal{S}^0$. Suppose that $\beta \notin \mathcal{S}$; then, there exists β^* such that $\varepsilon(\beta^*) \leq \varepsilon(\beta)$, and $\varepsilon(\beta^*) \neq \varepsilon(\beta)$. Let $t^* = \varepsilon(\beta^*)$; then, $\varepsilon(\beta^*) = t^* \leq \varepsilon(\beta) \leq t$ and $t^* \neq t$, thus $(\beta, t) \notin \mathcal{S}^0$, which is a contradiction. ∎

The proposition above enables us to use the powerful tools of linear vector optimization. In particular, one can characterize the set $\mathcal{S}$ as the set of optimal solutions to problems of the form (P_ϕ) when ϕ varies in the set $\mathcal{F}$ of strictly increasing functions. As the functions ε_i are polyhedral, one has (see [5, Theorem 3.2])

PROPOSITION 5. *Let $\beta^* \in \mathbb{R}^d$. The following statements are equivalent*:

(i) $\beta^* \in \mathcal{S}$.

(ii) There exists $w \in W = \{w \in \mathbb{R}^n : w_i > 0 \text{ for all } i\}$ such that β^* is an optimal solution to problem $\min_\beta \sum_{i=1}^n w_i \varepsilon_i(\beta)$.

(iii) There exists a strictly increasing function $\phi: \mathbb{R}^n \to \mathbb{R}$ such that β^* is an optimal solution to (P_ϕ).

3. A Geometrical Characterization of $\mathcal{S}$

As Proposition 4 shows, (VOP) can be transformed into an equivalent linear vector optimization problem (Q). As the set $\mathcal{S}^0$ of Pareto-optimal solutions to (Q) can be obtained by means of existing methods ([3, 16, 18]), $\mathcal{S}$ can be obtained as follows:

(1) Find $\mathcal{S}^0$.

(2) Use Proposition 4.

However, we can exploit the special properties of the functions involved in (VOP) to describe geometrically $\mathcal{S}$.

Indeed, ε is a polyhedral (convex piecewise linear) function, thus $\mathbb{R}^d$ can be split into domains of linearity of ε, which will be called, following [7], *elementary convex sets*. This idea, successfully used for location problems ([6, 7, 8]), leads to a geometrical characterization of $\mathcal{S}$, as we show in this section.

For $i = 1, \ldots, n$, let

$$D_i(1) = \{\beta \in \mathbb{R}^d : y_i - x_i' \beta > 0\}$$

$$D_i(-1) = \{\beta \in \mathbb{R}^d : y_i - x_i' \beta < 0\}$$

$$D_i(0) = \{\beta \in \mathbb{R}^d : y_i - x_i' \beta = 0\}.$$

For each $k = (k_1, k_2, \ldots, k_n) \in \{-1, 0, 1\}^n$, let $C(k)$ be the polyhedron $C(k) = \bigcap_{i=1}^n D_i(k_i)$.

Definition 1. A nonempty set $C \subset \mathbb{R}^d$ is said to be an *elementary convex set* if $C = C(k)$ for some $k \in \{-1, 0, 1\}^n$.

Denote by $\mathcal{C}$ the family of elementary convex sets.

Evidently, the nonempty sets $C(k)$ are polyhedra that split $\mathbb{R}^d$ into domains of linearity of ε, i.e.,

- If $\beta \in D_i(1) \cup D_i(0)$, then $|y_i - x_i' \beta| = y_i - x_i' \beta$.
- If $\beta \in D_i(-1)$, then $|y_i - x_i' \beta| = -(y_i - x_i' \beta)$.

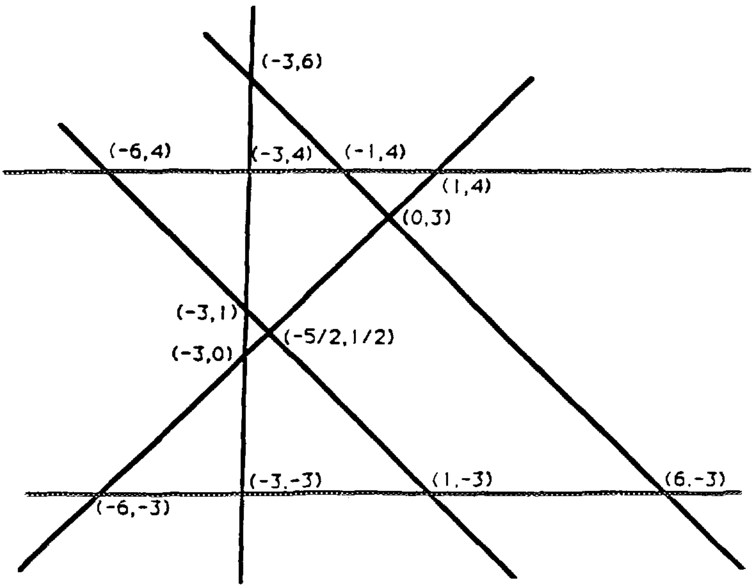

FIG. 1. Elementary convex sets.

As a consequence, for any elementary convex set C, there exists $k = (k_1, k_2, \ldots, k_n) \in \{-1, 1\}^n$ such that, for $i = 1, \ldots, n$, $\varepsilon_i(\beta) = k_i(y_i - x_i'\beta)$ for all $\beta \in C$.

As an illustration, consider the following example:

EXAMPLE 1. Let $d = 2$, $n = 6$, $x_1 = (0, 1)$, $y_1 = 4$; $x_2 = (1, 1)$, $y_2 = 3$; $x_3 = (0, 1)$, $y_3 = -3$; $x_4 = (-1, 1)$, $y_4 = 3$; $x_5 = (1, 0)$, $y_5 = -3$; and $x_6 = (-1, 1)$, $y_6 = -2$. As shown in Fig. 1, $\mathscr{C}$ consists of 65 different elementary convex sets, 20 of which have dimension 2, 32 of which have dimension 1, and 13 of which have dimension 0.

The elementary convex set $C(1, 1, -1, 1, -1, -1)$ is the compact polyhedron with vertices $(0, 3)$, $(6, -3)$, $(1, -3)$, and $(-5/2, 1/2)$.

Although the cardinality of $\mathscr{C}$ may be rather high, it can only increase polynomially in the sample size n. Indeed, one has

PROPOSITION 6. *$\mathscr{C}$ has cardinality $O(n^d)$ and can be described in $O(n^d)$ time with $O(n^d)$ space.*

Proof. One only has to see that $\mathscr{C}$ corresponds to the different pieces obtained with the arrangement associated with the family of hyperplanes $\mathscr{H} = \{y_i = x_i'\beta\}_{i=1}^n$, thus Theorem 3.3 in [9] applies. ∎

As shown below, the family $\mathscr{C}$ plays a crucial role in the characterization of $\mathscr{S}$; in fact, the main result in this paper is the following theorem.

THEOREM 1. *$\mathcal{S} = \bigcup_{C\in\mathcal{C}, C:\text{bounded}} C$, and $\mathcal{S}$ is connected.*

The proof is postponed after the statement of the following lemmas. Given a polyhedron C, denote by ri(C) its relative interior [2].

LEMMA 1. *Given $C \in \mathcal{C}$, the following statements are equivalent*:

(i) $C \subset \mathcal{S}$.

(ii) $\text{ri}(C) \cap \mathcal{S} \neq \emptyset$.

Proof. As C is convex and nonempty, $\text{ri}(C) \neq \emptyset$, thus (i) implies (ii). Let us show now the converse:

Let $\beta^* \in \text{ri}(C) \cap \mathcal{S}$. By Proposition 5, there exists $w \in W$ such that

$$\sum_{i=1}^{n} w_i\varepsilon_i(\beta^*) = \min_{\beta\in\mathbb{R}^d} \sum_{i=1}^{n} w_i\varepsilon_i(\beta)$$

Hence,

$$\sum_{i=1}^{n} w_i\varepsilon_i(\beta^*) = \min_{\beta\in C} \sum_{i=1}^{n} w_i\varepsilon_i(\beta)$$

As $C \in \mathcal{C}$, there exists $k \in \{-1, 1\}^n$ such that

$$\varepsilon_i(\beta) = k_i(y_i - x_i'\beta) \qquad \forall i = 1, \ldots, n, \forall \beta \in C.$$

Hence, $\sum_{i=1}^{n} w_i\varepsilon_i$ is a linear function (restricted to C), attaining its minimum at a point in ri(C).

It then follows that $\sum_{i=1}^{n} w_i\varepsilon_i$ is constant on C, thus

$$\sum_{i=1}^{n} w_i\varepsilon_i(\beta^*) = \min_{\beta\in\mathbb{R}^d} \sum_{i=1}^{n} w_i\varepsilon_i(\beta) \qquad \forall \beta^* \in C$$

Hence, by Proposition 5, it follows that $C \subset \mathcal{S}$. ▌

LEMMA 2. *Let $C \in \mathcal{C}$. Then,*

$$C \subset \mathcal{S} \qquad \textit{iff } C \textit{ is bounded.}$$

Proof. Let $C \subset \mathcal{S}$. We show (by contradiction) that C is bounded. Indeed, if C is not bounded, ri(C) contains a ray, i.e.,

$$\exists \beta_0, \beta_1 \in \mathbb{R}^d, \qquad \beta_1 \neq 0 / \beta_0 + \lambda\beta_1 \in C \qquad \forall\lambda \geq 0 \tag{i}$$

As ri(C) $\subset \mathcal{S}$, by Proposition 5, following the proof of Lemma 1, it is easily shown that there exists $w \in W$ such that

$$\min_{\beta \in \mathbb{R}^d} \sum_{i=1}^{n} w_i \varepsilon_i(\beta) = \sum_{i=1}^{n} w_i \varepsilon_i(\beta^*) = \alpha \qquad \forall \beta^* \in C. \tag{ii}$$

On the other hand (recall that C is an elementary convex set), one has

$$\exists k \in \{-1, 1\}^n \text{ such that } \varepsilon_i(\beta) = k_i(y_i - x_i' \beta) \qquad \forall \beta \in C, \forall i \tag{iii}$$

By (i),

$$k_i(y_i - x_i' \beta_0) - \lambda k_i x_i' \beta_1 \geq 0 \qquad \forall\, i = 1, \dots, n, \forall \lambda \geq 0,$$

thus

$$k_i x_i' \beta_1 \leq 0 \qquad \forall\, i = 1, \dots, n. \tag{iv}$$

By (ii) and (iii),

$$\alpha = \sum_{i=1}^{n} w_i k_i (y_i - x_i' \beta_0 - \lambda x_i' \beta_1) \qquad \forall \lambda \geq 0,$$

thus $\sum_{i=1}^{n} w_i k_i x_i' \beta_1 = 0$, which, by (iv), implies that

$$k_i x_i' \beta_1 = 0 \qquad \forall i = 1, \dots, n,$$

contradicting assumption A ($k \in \{-1, 1\}^n$ and $\beta_1 \neq 0$). Hence, C is bounded, as we wanted to show.

The converse can also be shown by contradiction: Suppose that C is a bounded elementary convex set which is not contained in $\mathcal{S}$. As C is a nonempty convex set, its relative interior ri(C) is not empty. Let $\beta_0 \in$ ri(C); by Lemma 1, ri(C) $\cap\, \mathcal{S} = \varnothing$, thus $\beta_0 \notin \mathcal{S}$. By Proposition 2, there exists $\beta_1 \in \mathcal{S}$ such that

$$\varepsilon_i(\beta_1) \leq \varepsilon_i(\beta_0)\ \forall i, \qquad \varepsilon_i(\beta_1) < \varepsilon_i(\beta_0) \text{ for some } i \tag{v}$$

As C is bounded, the ray $\{\beta_0 + \lambda(\beta_0 - \beta_1): \lambda \geq 0\}$ is not contained in C. Let $\lambda^* = \max\{\lambda \geq 0: \beta_0 + \lambda(\beta_0 - \beta_1) \in C\}$, and let $\beta^* = \beta_0 + \lambda^*(\beta_0 - \beta_1)$. It follows that $\beta^* \in C \backslash \text{ri}(C)$, thus there exists $j \in \{1, \dots, n\}$ such that $\varepsilon_j(\beta^*) = 0 < \varepsilon_j(\beta_0)$.

β_0 is an interior point of the segment with endpoints β_1 and β^*, i.e.,

$\exists t \in (0, 1)$ such that $\beta_0 = t\beta_1 + (1 - t)\beta^*$. By convexity of the function ε_j,

$$\varepsilon_j(\beta_0) \leq t\varepsilon_j(\beta_1) + (1 - t)\varepsilon_j(\beta^*) = t\, \varepsilon_j(\beta_1) < \varepsilon_j(\beta_1),$$

thus $\varepsilon_j(\beta_0) < \varepsilon_j(\beta_1)$, contradicting (v). Hence, $C \subset \mathcal{S}$, as asserted. ∎

Proof of Theorem 1. Evidently, for each $\beta \in \mathbb{R}^d$ there exists $C \in \mathscr{C}$ such that $\beta \in \mathrm{ri}(C)$. Hence, one has

$$\begin{aligned} \mathcal{S} &= \textstyle\bigcup_{C\in\mathscr{C}} \{\beta \in \mathcal{S}\colon \beta \in \mathrm{ri}(C)\} && \text{(by Lemma 1)} \\ &= \textstyle\bigcup_{C\in\mathscr{C}\colon C\subset\mathcal{S}} C && \text{(by Lemma 2)} \\ &= \textstyle\bigcup_{C\in\mathscr{C} \text{ is bounded}} C, \end{aligned}$$

as asserted.

Finally, connectedness of $\mathcal{S}$ can be proved as follows: As (Q) is a linear vector optimization problem, the set $\mathcal{S}^0$ of its Pareto-optimal solutions is connected [12]. Hence, as the projection function $(u, v) \in \mathbb{R}^d \times \mathbb{R}^n \to u \in \mathbb{R}^d$ is continuous, the set $\mathcal{S}^1 = \{\beta \in \mathbb{R}^d\colon (\beta, t) \in \mathcal{S}^0 \text{ for some } t\}$ is connected. By Proposition 5, $\mathcal{S}^1 = \mathcal{S}$, thus the result holds. ∎

Example 2. For the sample $\mathbb{O}$ in Example 1, Theorem 7 implies that $\mathcal{S}$ is the set enclosed by the polygonal with vertices $(-3, 6)$, $(-1, 4)$, $(0, 3)$, $(6, -3)$, $(-6, -3)$, $(-3, 0)$, $(-3, 1)$, $(-6, 4)$, depicted in Fig. 2.

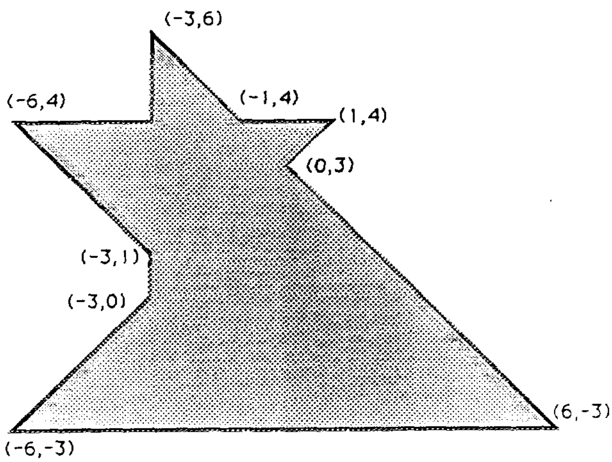

Fig. 2. The set of pareto-optimal parameters.

4. Optimal Predictions

The final aim of linear regression consists in predicting the response variable (y) by means of the predictor variable (x) following the model

$$Y = X' \beta + \varepsilon.$$

As we showed in Section 2, $\mathcal{S}$ corresponds to the set of hyperplanes which are optimal solutions to problems of the form (Pϕ), for some strictly increasing function ϕ.

In this section we characterize the set of predictions for x, when the parameter β varies in $\mathcal{S}$.

Definition 2. The point-to-set function Π: $x \in \mathbb{R}^d \rightarrow \{x' \beta: \beta \in \mathcal{S}\}$ is called the *prediction function*.

In order to determine the prediction function Π, some notation is needed:

- Let V denote the set of *vertices* of the elementary convex sets:

$$V = \{\beta: \{\beta\} \in \mathcal{C}\}.$$

- Let L: $\mathbb{R}^d \rightarrow \mathbb{R}$ and U: $\mathbb{R}^d \rightarrow \mathbb{R}$ be the functions defined as

$$L(x) = \min \{x' \beta: \beta \in V\}$$
$$U(x) = \max \{x' \beta: \beta \in V\}.$$

Observe that L (respectively, U) is concave (respectively, convex) piecewise linear. One then has

Theorem 2. *$\Pi(x) = [L(x), U(x)]$ for all $x \in \mathbb{R}^n$.*

Proof. Let $x \in \mathbb{R}^d$. The set $\Pi(x)$ is a compact connected subset of $\mathbb{R}$, because $\Pi(x)$ is the image of the compact connected $\mathcal{S}$ under the linear (thus continuous) function $\beta \in \mathbb{R}^d \rightarrow x' \beta$.

One then has

$$\Pi(x) = [\min_{z \in \Pi(x)} z, \max_{z \in \Pi(x)} z] \qquad \text{(vi)}$$

Consider the family $\mathcal{C}^*$ of bounded elementary convex sets. Evidently, every $C \in \mathcal{C}^*$ is a bounded polyhedron, whose extreme points are elements of V. Furthermore, every vertex $\beta \in V$ verifies $\{\beta\} \in \mathcal{C}^*$.

Hence, one has

$$\min_{z\in\Pi(x)} z = \min_{\beta\in\mathcal{S}} x'\beta = \min_{C\in\mathscr{C}^*}\min_{\beta\in C} x'\beta$$
$$= \min_{C\in\mathscr{C}^*}\min_{\beta\in C\cap V} x'\beta = \min_{\beta\in V} x'\beta = L(x).$$

The equality $\max_{z\in\Pi(x)} z = U(x)$ can be shown in the same way. By (vi), $\Pi(x) = [L(x), U(x)]$, as asserted. ∎

5. Concluding Remarks

In this paper we have considered the linear regression problem as a vector optimization problem, whose set $\mathcal{S}$ of Pareto-optimal solutions has been characterized as:

(i) The set of optimal solutions to problems where the error measure is a strictly increasing function.

(ii) The union of all the elementary convex sets which are bounded.

Furthermore, $\mathcal{S}$ can be obtained by standard vector linear programming techniques. These properties enable us to say, although depending on the error measure used one can obtain different predictions at a point x, the possible predictions are given by the points in a certain interval $\Pi(x)$, which can be explicitly obtained.

An interesting line of future research could be the study of reductions of the set of optimal parameters by reducing the family $\mathcal{F}$ of globalizing functions ϕ allowed (by imposing, say, that ϕ must be a nondecreasing symmetric and convex function, or a norm).

Acknowledgments

Thanks are due to Professors Frank Plastria and Peter Rousseeuw for their valuable comments. We also express our gratitude to an anonymous referee for his suggestions, which have added greatly to the paper.

References

1. T. S. Arthanary and Y. Dodge, "Mathematical Programming in Statistics," Wiley, New York, 1981.
2. A. Brondsted, "An Introduction to Convex Polytopes," Springer-Verlag, New York/Berlin, 1983.

3. V. CHANKONG AND Y. Y. HAIMES, "Multiobjective Decision Making: Theory and Applications," North-Holland, Amsterdam, 1983.
4. Y. DODGE, "Statistical Data Analysis Based on the L_1-Norm and Related Methods," North-Holland, Amsterdam, 1987.
5. R. DURIER, Weighted factor results in vector optimization, *J. Optim. Theory Appl.* **58** (1988), 411–430.
6. R. DURIER, On Pareto-optima, the Fermat–Weber problem, and Polyhedral gauges, *Math. Programming* **47** (1990), 65–79.
7. R. DURIER AND C. MICHELOT, Geometrical properties of the Fermat–Weber problem, *European J. Oper. Res.* **20** (1985), 332–343.
8. R. DURIER AND C. MICHELOT, Sets of efficient points in a normed space, *J. Math. Anal. Appl.* **117** (1986), 506–528.
9. H. EDELSBRUNNER, J. O'ROURKE, AND R. SEIDEL, Constructing arrangements of lines and hyperplanes with applications, *SIAM J. Comput* **15** (1986), 341–363.
10. R. GONIN AND A. H. MONEY, "Nonlinear L_p-Norm Estimation" Dekker, New York, 1989.
11. G. F. MCCORMICK AND V. A. SPOSITO, Using the L_2-estimator in the L_1-estimation, *SIAM J. Numer. Anal.* **13** (1976), 337–343.
12. P. H. NACCACHE, Connectedness of the set of non-dominated outcomes in multicriteria optimisation, *J. Optim. Theor. Appl.* **25** (1978), 459–467.
13. F. PLASTRIA, Localization in single facility location, *European. J. Oper. Res.* **18** (1984), 215–219.
14. R. T. ROCKAFELLAR, "Convex Analysis," Princeton Univ. Press, Princeton, NJ, 1970.
15. P. J. ROUSSEEUW, Least median of squares regression, *J. Amer. Statist. Assoc.* **82** (1987), 794–801.
16. R. E. STEUER, "Multiple Criteria Optimization: Theory, Computation and Application," Wiley, New York, 1986.
17. D. J. WHITE, "Optimality and Efficiency," Wiley, New York, 1982.
18. P. L. YU AND M. ZELENY, The set of all nondominated solutions in linear cases and a multicriteria simplex method, *J. Math. Anal. Appl.* **49** (1975), 430–468.