

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/265425662>

A mixed primal–dual bases algorithm for regression under inequality constraints. Application to convex regression

ARTICLE *in* SCANDINAVIAN JOURNAL OF STATISTICS · JANUARY 1989

Impact Factor: 0.87

CITATIONS

33

READS

22

2 AUTHORS, INCLUDING:



[Hélène Massam](#)

York University

68 PUBLICATIONS 728 CITATIONS

SEE PROFILE

A Mixed Primal-Dual Bases Algorithm for Regression under Inequality Constraints. Application to Concave regression

D. A. S. FRASER

University of Waterloo and York University

H. MASSAM

York University

ABSTRACT. In some statistical non-parametric models the mean of the random variable y has to satisfy specific constraints. We consider the case where the set defined by the constraints is a closed polyhedral cone K in \mathbb{R}^k . For example, when the mean is required to be concave in x , the set of acceptable means is a closed convex cone defined by $k-2$ linear inequalities in \mathbb{R}^k . The least squares estimate of the mean is then the projection of the data point y on the cone K . In this paper, we present an algorithm to find the least square estimate of the mean in a finite number of steps. Other algorithms to solve this problem have been given before. The successive approximations in such algorithms are usually points on the faces of K . The solution here is reached by following a fixed line joining an arbitrary but suitably chosen initial point y_0 to the data point y . The l -dimensional subspace spanned by the generators of the cone K is divided into 2^l regions which can be described as the set of points with non-negative coordinates in mixed primal-dual bases relative to the cone K . Any point y belongs to one and only one of these regions S_j with corresponding basis \mathcal{B}_j . The projection of y on K is immediately obtained from the expression of y in \mathcal{B}_j by dropping the dual component of y .

Key words: non-linear regression, dual basis, least squares, cone

1. Introduction

1.1 The problem and the algorithm

For some problems in economics, it is known that the marginal productivity is decreasing or that the demand function is concave. It is then appropriate to use a non-parametric statistical model $y = \eta(x) + \varepsilon$ where the mean $\eta(x)$ of the random variable y has the required property of concavity in x (see Hildreth, 1954; Wu, 1982; McFadden, 1985; Amemiya, 1983). The term ε designates the random error. A quantity y_i is measured at some fixed value x_i , $i = 1, \dots, k$, of an independent variable x . The problem is to find the piecewise linear concave function $\eta(x)$ with values $\eta_i = \eta(x_i)$ at x_i which minimizes the weighted sum of squares

$$\sum_{i=1}^k m_i (y_i - \eta_i)^2.$$

We denote the vector $(\eta_1, \dots, \eta_k)'$ by η and the vector $(y_1, \dots, y_k)'$ by y . As we will show in section 2, the problem then becomes

$$\text{Minimize } \|y - \eta\|^2 \quad \text{given } \eta \in K$$

where K is a closed convex cone in \mathbb{R}^k defined by $k-2$ linear inequalities representing the concavity constraints. This concave regression problem is a particular case of a general approximation problem where the mean function or any other given function is known to have a certain underlying property (e.g. concavity). It is then desirable to approximate this given

function by a non-negative linear combination of functions, with the same property, forming a basis in a specified finite dimensional space. The general problem can be formulated as follows.

Let y be a data point in \mathbb{R}^k and let K be a closed convex cone defined in either of two possible ways:

$$(F1) \quad K = \{\eta \in \mathbb{R}^k : \gamma_i' M \eta \leq 0 \text{ for } i = 1, \dots, l\}$$

$$(F2) \quad K = \left\{ \eta \in \mathbb{R}^k : \eta = \sum_{i=1}^l b_i \beta_i + \sum_{i=l+1}^k b_i \beta_i, \quad b_i \geq 0 \text{ for } i = 1, \dots, l \right\}$$

where $l \leq k$. The vectors γ_i , $i = 1, \dots, l$, are assumed to be linearly independent and so are the vectors β_i , $i = 1, \dots, k$. The γ_i 's and β_i 's are linked by the relationship $\gamma_i' M \beta_j = -\delta_{ij} = -1$ if $i = j$, and $= 0$ if $i \neq j$. The matrix M is a given $k \times k$ positive definite matrix. The two formulations will be shown to be equivalent. Since η_i is the mean of the variable y at x_i , the vectors γ_i can be viewed as a set of contrasts which restrict the solution to the regression problem. We wish to find the point $\hat{\eta}$ in K which is closest to y in the metric defined by the matrix M .

It is this general problem

$$\text{Minimize } \|y - \eta\|^2 = (y - \eta)' M (y - \eta) \quad \text{given } \eta \in K$$

where K is defined above, that we consider in this paper. Our aim is to develop an efficient algorithm to find the solution in a finite number of steps.

A basis for \mathbb{R}^k made up of a subset of the primal basis vectors γ_i and a complimentary subset of the dual basis vectors β_i is called a mixed primal-dual basis. The regression of y on the cone K is very simple if y is represented in one of these primal-dual bases. The problem is then to find this particular basis. By starting with a trial basis and by successively interchanging primal and dual basis vectors one by one, and in an orderly way, the solution is found in a finite number of steps. Since only one basis vector is changed at each step, the computations involved are limited. At each step, the calculations are similar to the ones performed in ordinary linear regression when one regressor is interchanged with another (see for example, Draper & Smith (1981) or Seber (1977)).

1.2. Some previous studies

The concave regression problem has been the subject of many studies. Hildreth (1954) was the first one to address it. He gave an algorithm converging in an infinite number of steps, based on the Kuhn-Tucker conditions for optimality. Dent (1973), Dent, Robertson & Johnson (1977) and Holloway (1979) also gave various algorithms.

More recently Wu (1982) and Dykstra (1983) offered simpler solutions either convergent in an infinite number of steps or not necessarily convergent. In the context of approximation theory, Wilhelmsen (1976) gave an algorithm for finding the point in K nearest to y where

$$K = \left\{ \eta \in \mathbb{R}^k : \eta = \sum_{i=1}^N b_i \beta_i, b_i > 0 \right\}$$

where β_i are arbitrary vectors in \mathbb{R}^k and N is any integer which may be larger than k . The algorithm gives successive approximations, on the faces of K , converging to $\hat{\eta}$. Pschenichny & Danilin (1978) give an algorithm using the basis β_i , $i = 1, \dots, l$ only, while K is defined in terms of γ_i , $i = 1, \dots, l$. Their algorithm is similar to the one devised by Wilhelmsen. Both algorithms converge in a finite number of steps.

Our assumptions here are more restrictive than Wilhelmsen's as N is assumed to be less than k and the β_i 's independent. However, the assumption of independence of the β_i 's allows us to

use the concept of a dual basis and proceed to the solution by following a fixed line linking an initial point to the data point. The solution is then obtained by omitting the dual vectors in the final expression.

In section 2, we state the general cone regression problem and the concave regression as a special case. In section 3, we describe the algorithm. We prove that it converges in a finite number of steps, we then describe it geometrically and finally give briefly the step by step procedure. In section 3.3, the algorithm is applied to a numerical example and compared with Wilhelmsen's algorithm applied to the same example.

2. The cone regression problem and the concave regression problem as a special case

2.1. The general problem

Let y be a random variable with mean $\eta(x)$ where $\eta(x)$ is a function of an independent variable x . The variable x does not have to be univariate but for the sake of simplicity, we do not use the vector notation. Let y_i be the measurement at some fixed value x_i of x , for $i = 1, \dots, k$. The point $y = (y_1, \dots, y_k)$ is the data point in \mathbb{R}^k . Let M be a symmetric positive definite matrix that gives the appropriate metric on \mathbb{R}^k and let K be a closed convex cone defined by some constraints on $\eta(x)$ and given by either of formulations (F1) or (F2) above. We want to find the least squares estimate $\hat{\eta}$ of η in K which is the projection of y on K .

Let $\gamma_i, i = 1, \dots, l$, be l given independent vectors in \mathbb{R}^k where $l < k$, provided by the underlying statistical problem. The set $\{\gamma_1, \dots, \gamma_l\}$ can be completed by $k-l$ independent vectors orthogonal to $\gamma_i, i = 1, \dots, l$, and orthonormal to each other (in the M metric of course). The set $\{\gamma_i, i = 1, \dots, k\}$ forms a basis of \mathbb{R}^k .

We can now define the basis $\{\beta_i, i = 1, \dots, k\}$ dual to $\{\gamma_i, i = 1, \dots, k\}$. The relationship between the β_i 's and the γ_i 's is

$$\beta_i M \gamma_j = \begin{cases} -1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases} \quad (2.1)$$

since $\gamma_j, j = l+1, \dots, k$, are orthonormal and orthogonal to $\gamma_j, j = 1, \dots, l$, $\gamma_j = -\beta_j$ for $j = l+1, \dots, k$. Let \tilde{B} and \tilde{C} denote the matrix with column vectors $\beta_i, i = 1, \dots, k$, and $\gamma_i, i = 1, \dots, k$, respectively and let B and C denote the matrices with column vectors $\beta_i, i = 1, \dots, l$, and $\gamma_i, i = 1, \dots, l$, respectively. We have

$$\tilde{B} M \tilde{C} = -I_k \quad \text{and} \quad B M C = -I_l.$$

We now show the expression for the cone K given in terms of the β_i 's is equivalent to the one in terms of the γ_i 's.

Lemma 2.1

Formulations (F1) and (F2) are equivalent.

Proof. Consider the cone K given by (F1). The quantities $\gamma_i' M \eta, i = 1, \dots, l$, are the first l coordinates of $\tilde{C} M \eta$. Since $\tilde{B} M \tilde{C} = -I$, $\tilde{C} M \eta = -\tilde{B}^{-1} \eta$ gives the negative of the coordinates η in the basis $\{\beta_i, i = 1, \dots, k\}$. Therefore points in K have their first l coordinates non-negative and can be written as

$$\eta = \sum_{i=1}^k b_i \beta_i$$

where $b_i \geq 0$ for $i = 1, \dots, l$. This shows that (F1) and (F2) are equivalent.

Let H be the subspace spanned by β_i or γ_i , $i=l+1, \dots, k$ and let $\mathcal{L}(K)$ be the subspace spanned by β_i , $i=1, \dots, l$, or equivalently γ_i , $i=1, \dots, l$. Since H and $\mathcal{L}(K)$ are orthogonal, y is equal to $y_H + z$ where y_H and z are the projections of y on H and $\mathcal{L}(K)$ respectively. If we denote by X the matrix with column vectors β_i , $i=l+1, \dots, k$,

$$y_H = X(X'M^{-1}X)^{-1}X'M^{-1}y \quad \text{and} \quad z = B(B'M^{-1}B)^{-1}B'M^{-1}y. \quad (2.2)$$

The projection of y on K is the sum of the projection of z on K and y_H . The projection y_H can be immediately obtained from (2.2) and therefore in section 3, when determining the projection of y on K , we have only to determine the projection of z on K .

Remark. In some cases, the constraints on η will also comprise equalities (for example, some contrasts would be inequalities $\gamma'_i M \eta \leq 0$, $i=1, \dots, l$, while some others would be equalities $\gamma'_i M \eta = 0$, $i=l+1, \dots, l+r$). The formulation (F1) of K would be

$$K = \{\eta \in \mathbb{R}^k: \gamma'_i M \eta \leq 0, i=1, \dots, l, \quad \gamma'_i M \eta = 0, i=l+1, \dots, l+r \text{ with } l+r \leq k\}.$$

The vectors β_j , $j=1, \dots, k$, can be defined as before with $l+r$ replacing l . By the same argument as before, this is equivalent to the first l coordinates of η being positive and the succeeding r coordinates equal to 0. The cone K is in a $k-r$ dimensional subspace and its formulation (F2) is

$$K = \left\{ \eta \in \mathbb{R}^k: \eta = \sum_{i=1}^l b_i \beta_i + \sum_{i=l+r+1}^k b_i \beta_i, b_i \geq 0, i=1, \dots, l \right\}.$$

In this case, let \bar{H} denote the subspace generated by β_i , $i=l+r+1, \dots, k$, and $\mathcal{L}(K)$ the subspace generated by β_i , $i=1, \dots, l$. The projection of y on K would then be the sum of $y_{\bar{H}}$ and the projection of z on K . The problem would reduce to finding the projection of z on the cone defined by $\{\eta \in \mathbb{R}^k: \gamma'_i M \eta \leq 0, i=1, \dots, l\}$ which is the general problem we consider.

2.2. The concave regression problem

When the mean $\eta(x)$ of y is required to be a concave function of x , the weighted least squares problem is:

$$\text{Minimize } \sum_{i=1}^k m_i (y_i - \eta_i)^2 = (y - \eta)' M (y - \eta)$$

such that

$$\begin{aligned} \Delta x_2 \eta_1 - (\Delta x_1 + \Delta x_2) \eta_2 + \Delta x_1 \eta_3 &\leq 0 \\ &\vdots \\ \Delta x_{k-1} \eta_{k-2} - (\Delta x_{k-2} + \Delta x_{k-1}) \eta_{k-1} + \Delta x_{k-2} \eta_k &\leq 0 \end{aligned} \quad (2.3)$$

where $\Delta x_i = x_{i+1} - x_i$. If we denote by α_i the vector $\alpha_i = (0, \dots, 0, \Delta x_{i+1}, -(\Delta x_i + \Delta x_{i+1}), \Delta x_i, 0, \dots, 0)$ (with $i-1$ zeros on the left and $k-i-2$ on the right) and by γ_i the vector $M^{-1} \alpha_i$, the inequalities (2.2) become $\gamma'_i M \eta \leq 0$ for $i=1, \dots, k-2$. These $k-2$ inequalities define a closed convex cone K . The concave regression problem is therefore a special case of the cone regression problem with M being a diagonal matrix (m_i) .

We now complete the set of vectors $\{\gamma_1, \dots, \gamma_{k-2}\}$ by two vectors γ_{k-1} and γ_k orthogonal to γ_i , $i=1, \dots, k-2$, and orthonormal to each other. We obtain $\gamma_{k-1} = 1/\|1\|$ and $\gamma_k = (x - \bar{x}1)/\|x - \bar{x}1\|$ where 1 is the k -dimensional vector with all components equal to 1,

$$x = (x_1, \dots, x_k)', \bar{x} = \frac{\sum_{i=1}^k m_i x_i}{\sum_{i=1}^k m_i}$$

and $\|\cdot\|$ represents the norm of a vector in the M metric. The vectors γ_{k-1} and γ_k represent a constant and linear function in \mathbb{R}^k while the vector γ_i , $i=1, \dots, k-2$, represents concave two-piece linear functions with a negative second difference at x_{i+1} only. If the data point y is such that $\gamma_i' M y \geq 0$ for $i=1, \dots, k-2$, the graph of $y=(y_1, \dots, y_k)'$ corresponds to a piecewise linear convex function of x . The concave function η which minimizes the distance between y and η can then be seen to be a linear function of x and can therefore be expressed in terms of γ_{k-1} and γ_k only. Wu (1982) had observed that in this case, with convex data, the ordinary linear regression is the solution to the concave regression problem.

If $\gamma_i' M y < 0$ for some $i=1, \dots, k-2$, then generally an iterative algorithm is needed to solve the regression problem.

3. The mixed primal-dual bases algorithm

3.1. The theory

Let K be the non-negative cone given by (F2) or its equivalent formulation (F1) and let z and y_H be the projection of y on $\mathcal{L}(K)$ and H respectively. As noted in section 2, $y = y_H + z$. Of course, z is in $\mathcal{L}(K)$ and our problem is to find the projection of z on the cone

$$K' = K \cap \mathcal{L}(K) = \left\{ \eta \in \mathbb{R}^k : \eta = \sum_{i=1}^l b_i \beta_i, b_i \geq 0 \right\}.$$

We will show that the l -dimensional space spanned by β_1, \dots, β_l can be partitioned into 2^l disjoint regions such that each region can be described as the non-negative orthant in the basis $\mathcal{B}_J = \{\beta_i, i \in J, \gamma_i, i \in L/J\}$ where J is a subset of $L = \{1, \dots, l\}$. Therefore, any z in $\mathcal{L}(K)$ can be expressed as

$$z = \sum_{i \in J} b_i \beta_i + \sum_{i \in L/J} c_i \gamma_i,$$

for some $J \subset L$ where b_i are positive and c_i are non-negative. Since β_i , $i \in J$, are orthogonal to γ_i , $i \in L/J$, the projection of z on K' is just

$$\sum_{i \in J} b_i \beta_i.$$

Let us note, at this point, that any mixed primal-dual set $\{\beta_i, i \in J, \gamma_i, i \in L/J\}$ forms a basis of $\mathcal{L}(K)$ for any subset J of L . Indeed, with respect to the metric M , the β_i 's $i \in J$ are independent orthonormal vectors spanning a subspace of dimension $|J|$, the γ_i 's are also independent, orthonormal and span a subspace of dimension $l - |J|$. Moreover, since γ_i is orthogonal to β_j for any $i \neq j$, the two subspaces are orthogonal and their dimensions add up to l , the dimension of $\mathcal{L}(K)$. Therefore, $\{\beta_i, i \in J, \gamma_i, i \in L/J\}$ forms a mixed primal-dual basis for $\mathcal{L}(K)$.

To identify the partition of $\mathcal{L}(K)$, we need some definitions and preliminary results. A *face* of a closed convex set of K is a convex subset \tilde{K} of K such that every closed line segment in K with a relative interior point in \tilde{K} has both endpoints in \tilde{K} . It follows that if \tilde{K} is the set of points where a linear functional achieves its maximum over K , then \tilde{K} is a face of K (see Rockafellar, 1970). The *support cone* $\mathcal{S}_K(x)$ of a closed convex set K at x is the smallest closed convex cone with vertex at the origin containing $K - x$. For any cone K with vertex at x , the cone $K^\perp = \{\eta^\perp : (\eta^\perp - x)' M (\eta - x) \leq 0, \forall \eta \in K\}$ is called the *dual cone* of K . The following lemma identifies the faces of $K' = K \cap \mathcal{L}(K)$.

Lemma 3.1

Let K be the cone determined by the relations $\gamma'_i M \eta \leq 0$, $i = 1, \dots, l$, and let β_i be the dual basis as defined in (2.1). The faces of $K' = K \cap \mathcal{L}(K)$ are the sets

$$\left\{ \eta \in \mathbb{R}^k : \eta = \sum_{i \in J} b_i \beta_i, b_i \geq 0 \right\}$$

for any subset J of L . Moreover the set of all relatively open faces.

$$F_J = \left\{ \eta \in \mathbb{R}^k : \eta = \sum_{i \in J} b_i \beta_i, b_i > 0 \right\}$$

forms a partition of K' .

Proof. The sets $\{\eta \in K' : \gamma'_j M \eta = 0\}$ are clearly faces of K' . Since

$$\gamma'_j M \eta = \sum_{i=1}^k b_i \gamma'_j M \beta_i = -b_j,$$

these faces are $\{\eta \in K' : b_j = 0\}$. Any intersection of faces is another face and thus any set of the form

$$\bigcap_{j \in L/J} \{\eta \in K' : b_j = 0\} = \left\{ \eta \in K' : \eta = \sum_{i \in J} b_i \beta_i, b_i \geq 0 \right\}$$

is also a face of K' . Since K' is completely determined by the l inequalities $\gamma'_i M \eta \leq 0$, $i = 1, \dots, l$, there are no other faces. The last part of the lemma is a classical result (see Rockafellar, 1970, p. 164).

Let η_0 be any point in K' , then η_0 belongs to a unique relatively open face

$$F_J = \left\{ \eta \in K' : \eta = \sum_{i \in J} b_i \beta_i, b_i > 0 \right\}.$$

This face belongs to the hyperplanes $\gamma'_i M \eta = 0$ for $i \in L/J$. The support cone at η_0 is

$$\mathcal{S}_K(\eta_0) = \bigcap_{i \in L/J} \{\eta : \gamma'_i M \eta \leq 0\}$$

and its dual is

$$\mathcal{S}_K^\perp(\eta_0) = \left\{ \eta : \eta = \sum_{i \in L/J} c_i \gamma_i, c_i \geq 0 \right\}.$$

We will now turn our attention toward the set of points in $\mathcal{L}(K)$ projecting on a given point η_0 of K' : this set is denoted by $P_K^{-1}(\eta_0)$. The following results are intuitively obvious and have been proved in Zarantonello (1971).

Lemma 3.2

Let K' , F_J , β_i and γ_i be as in lemma 3.1. Then

(a) If η is a point of K' , belonging to F_J ,

$$P_K^{-1}(\eta) = \mathcal{S}_K^\perp(\eta) + \eta = \left\{ \eta + \sum_{i \in L/J} c_i \gamma_i, c_i \geq 0 \right\} = \left\{ \sum_{i \in J} b_i \beta_i + \sum_{i \in L/J} c_i \gamma_i, c_i \geq 0 \right\}$$

where

$$\sum_{i \in J} b_i \beta_i = \eta$$

and $b_i > 0$ for $i \in J$.

(b) The sets $P_{K'}^{-1}(\eta)$ are disjoint closed convex cones.

(c) $\bigcup_{\eta \in K'} P_{K'}^{-1}(\eta) = \mathcal{L}(K)$.

This lemma says that any point z in $\mathcal{L}(K)$ projects onto a unique η in K' and belongs to a unique non-negative orthant

$$S_J = \left\{ z : z = \sum_{i \in J} b_i \beta_i + \sum_{i \in L/J} c_i \gamma_i, \text{ with } b_i > 0, c_i \geq 0 \right\}.$$

The following figure (Fig. 1) in $\mathcal{L}(K) = \mathbb{R}^2$ illustrates the result above.

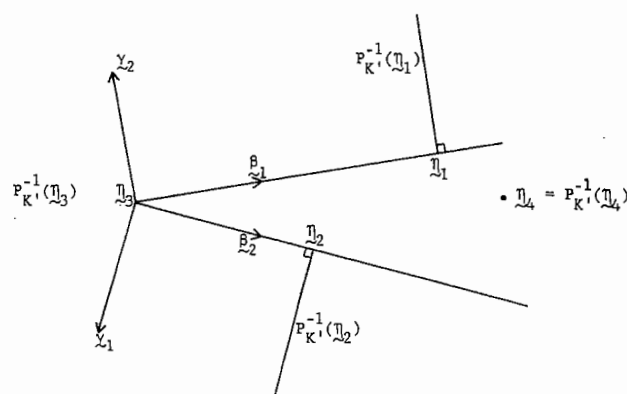


Fig. 1.

In Fig. 1 we see

$$P_{K'}^{-1}(\eta_1) = \{\eta_1 + c_2 \gamma_2, c_2 \geq 0\}$$

$$P_{K'}^{-1}(\eta_2) = \{\eta_2 + c_1 \gamma_1, c_1 \geq 0\}$$

$$P_{K'}^{-1}(\eta_3) = \{c_1 \gamma_1 + c_2 \gamma_2, c_1 \geq 0, c_2 \geq 0\}$$

$$P_{K'}^{-1}(\eta_4) = \{\eta_4\}.$$

The partition of \mathbb{R}^2 is made up of the 2^2 subsets

$$S_{(1)} = \{b_1 \beta_1 + c_2 \gamma_2, b_1 > 0, c_2 \geq 0\}$$

$$S_{(2)} = \{b_2 \beta_2 + c_1 \gamma_1, b_2 > 0, c_1 \geq 0\}$$

$$S_{(0)} = \{c_1 \gamma_1 + c_2 \gamma_2, c_1 \geq 0, c_2 \geq 0\}$$

$$S_{(1,2)} = \{b_1 \beta_1 + b_2 \beta_2, b_1 > 0, b_2 > 0\}.$$

Points in $S_{(1)}$ project on $b_1 \beta_1$. Points in $S_{(2)}$ project on $b_2 \beta_2$. Points in $S_{(0)}$ project on the origin.

Points in $S_{(1,2)}$ are their own projections. In general, points in

$$S_J = \left\{ \sum_{i \in J} b_i \beta_i + \sum_{i \in L/J} c_i \gamma_i, b_i > 0, c_i \geq 0 \right\}$$

project on

$$\sum_{i \in J} b_i \beta_i.$$

To find the projection of any z in $\mathcal{L}(K)$, we must find the subset S_j in which it is located, that is the basis $\mathcal{B}_j = \{\beta_i, i \in J, \gamma_i, i \in L/J\}$ in which z has all its coordinates non-negative. This is the subject of the next section.

3.2. Geometrical description of the algorithm

Let z be the data point in the subspace $\mathcal{L}(K)$ and let z_0 be some initial point. The point z_0 can be chosen on a face of K' reasonably close to z or in the interior of K' , for example $z_0 = \beta_1 + \beta_2 + \dots + \beta_L$. We will choose the latter. This is a reasonable choice since the first iteration will take us to a point on the face of K . Note that the point z_0 has all its coordinates non-negative in the basis $\mathcal{B}_0 = \{\beta_1, \dots, \beta_L\}$.

We compute the coordinates of z in \mathcal{B}_0 . If they are all non-negative, the projection of z is z itself, according to the theory in the previous section, and the problem is solved. If some coordinates are negative, we proceed as follows. Consider points $z_0 + t(z - z_0)$ for t in $(0, 1)$, on the line segment joining z_0 to z . If d_i and d_i^0 are the coordinates of z and z_0 respectively, the coordinates of $z_0 + t(z - z_0)$ are $d_i^0 + t(d_i - d_i^0)$. Since d_i^0 is positive, for t sufficiently small, these coordinates are also positive. However, as t increases, one of the coordinates will first become equal to 0 and then become negative. Let t_1 be the value of t which makes the first one of the coordinates $d_i^0 + t(d_i - d_i^0)$ equal to 0. Let j be the index corresponding to that coordinate, and

$$t_1 = \frac{d_j^0}{d_j^0 - d_j}. \quad (3.1)$$

Let us also assume for now that only one of the coordinates $d_j^0 + t(d_j - d_j^0)$ equals 0, for $t = t_1$.

Consider the point $\eta_t = z_0 + t(z - z_0)$ for t in a small neighbourhood of t_1 . For $t < t_1$, the coordinates of η_t in the basis \mathcal{B}_0 are positive. As t approaches t_1 , the coordinate b_j on β_j becomes 0 while, by continuity with respect to t , the other coordinates remain positive. For $t = t_1 + \varepsilon$, the coordinates on β_i remain positive except for $i = j$. When t increased from $t_1 - \varepsilon$ to $t_1 + \varepsilon$, the point η_t crossed the hyperplane

$$H_j = \left\{ \eta = \sum_{i=1}^k b_i \beta_i \text{ such that } b_j = 0 \right\} = \left\{ \sum_{i=1}^k b_i \beta_i \text{ such that } \gamma_j' M \eta = 0 \right\}.$$

Let us then remove the vector β_j from \mathcal{B}_0 and introduce γ_j . Since γ_j is orthogonal to all β_i for $i \neq j$ and $\gamma_j' M \eta_t > 0$ for $t = t_1 + \varepsilon$, the coordinates of $\eta_{t_1 + \varepsilon}$ in the basis $\mathcal{B}_1 = \{\beta_i, i \neq j, \gamma_j\}$ are all positive. The new region S that the line $z_0 z$ entered when t became slightly larger than t_1 is the section of \mathbb{R}^l corresponding to points with non-negative coordinates in the basis \mathcal{B}_1 . Let $z_1 = z_0 + t_1(z - z_0)$. The coordinates of z_1 in \mathcal{B}_0 are all positive except for its coordinate on β_j which is zero. When we change to \mathcal{B}_1 , the coordinates of z_1 are the same for $\beta_i, i \neq j$. The coordinate of z_1 on γ_j is 0. So the coordinates of z_1 in \mathcal{B}_1 are obtained from the coordinates of z_1 in \mathcal{B}_0 without calculations.

We now compute the coordinates of z in \mathcal{B}_1 . If they are non-negative, our problem is solved. If not, consider points $z_1 + t(z - z_1)$ for $t \in (0, 1)$ on the line segment joining z_1 to z , and proceed from z_1 as we did from z_0 .

Let $\mathcal{B}_s = \{\beta_i, i \in J, \gamma_i, i \in L/J\}$ be the basis obtained at the s th iteration. We keep moving along the line joining z_0 to z and determine the points z_s . The general step is: find the coordinates of z in \mathcal{B}_s . If they are all non-negative, the procedure terminates. Otherwise let

$$t_{s+1} = \min \left\{ \frac{d_i^s}{d_i^s - d_i}, 1 \text{ for } d_i < 0 \right\} = \frac{d_j^s}{d_j^s - d_j}$$

where d_i^s and d_i are the coordinates of z_s and z respectively in \mathcal{B}_s , and j is the index of the coordinate giving the minimum of

$$\frac{d_i^s}{d_i^s - d_i}.$$

Let $z_{s+1} = z_s + t_{s+1}(z - z_s)$, and \mathcal{B}_{s+1} be the basis obtained from \mathcal{B}_s by dropping the vector β_j or γ_j indexed by j and entering the vector γ_j or β_j respectively. If the coordinates of z in \mathcal{B}_{s+1} are all non-negative then the procedure terminates. Otherwise repeat the procedure.

The number of iterations needed is equal to the number of different regions S_j that the line joining z_0 to z has to cross. The number is obviously finite. The algorithm terminates in a finite number of iterations.

Remark. When computing

$$\min_i \left\{ \frac{d_i^s}{d_i^s - d_i}, 1 \right\},$$

it might happen that the minimum value is attained for more than one j . In this case, one should "move" the point z_s slightly while keeping it on the same face of S_j . One could for example increase one or more of the j th coordinate for which the minimum is attained. This would give a unique value of j . The procedure could resume from there.

The actual calculations in the algorithm involve mainly the computation of the coordinates of z in different bases that differ from one another by one vector only and are therefore very simple. A detailed description of the algorithm can be obtained from the authors.

3.3. Example: concave regression

We use the data given on p. 108 of Ratkowsky (1983). The variable x_i represents the age in days of the European rabbit in Australia, while y_i represents the dry weight of the eye lens. There are

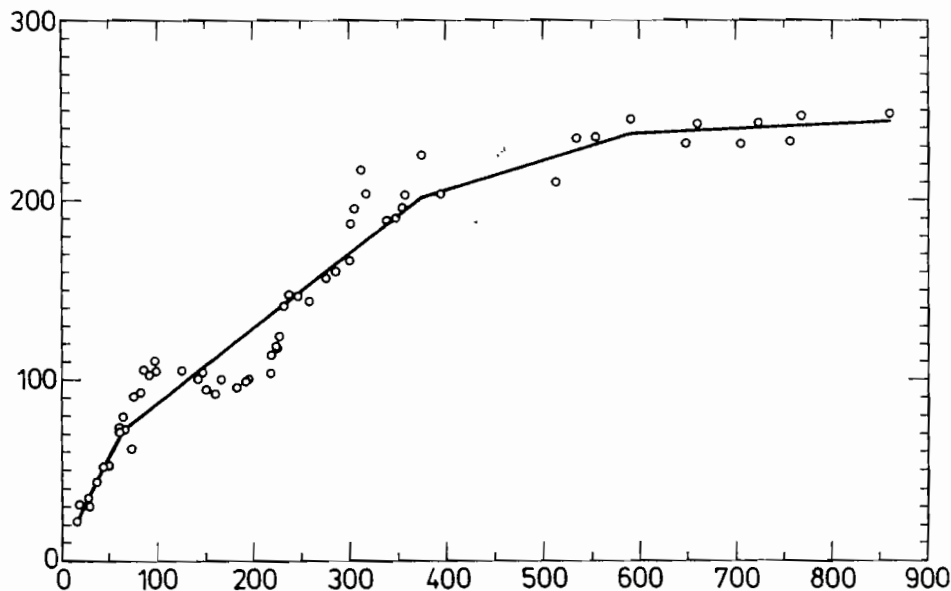


Fig. 2. Observed points and fitted line for the given example. The abscissa is the age in days and the ordinate is the dry weight of the eye lens.

$k = 60$ observations. Some of these actually represent group means and the matrix M is diagonal with diagonal elements giving the group sizes.

Ratkowsky fitted a parametric concave function. We fit a non-parametric concave function following our algorithm. The number of iterations needed to reach the solution is 54 for our algorithm and 75 for Wilhelmsen's. Moreover, in Wilhelmsen's algorithm, we have to compute the inverse of matrices of the type $X'MX$ at each iteration with no guarantee that these matrices will differ by one vector only from iteration to iteration (143 such matrices had to be computed).

The graph of the data and of the fitted function $\hat{\eta}(x)$ is given in Fig. 2.

Acknowledgement

The authors would like to thank the referees for their comments and Professor T. Schweder for many helpful suggestions. This research has been partially funded by NSERC grants A3011 and A8947.

References

- Amemiya, T. (1983). Non-linear regression models. In *Handbook of econometrics* (ed.) Z. Griliches & M. Intriligator, Vol. 1. North-Holland.
- Dent, W. (1973). A note on least squares fitting of function constrained to be either non-negative, nondecreasing or convex. *Management Sci.* **20**, 130–132.
- Dent, W., Robertson, T. & Johnson, M. (1977). Concave regression. Unpublished manuscript, University of Iowa.
- Draper, N. & Smith, H. (1981). *Applied regression analysis*. Wiley.
- Dykstra, R. L. (1983). An algorithm for restricted least squares regression. *J. Amer. Statist. Assoc.* **78**, 837–842.
- Hildreth, C. (1954). Point estimates of ordinates of concave functions. *J. Amer. Statist. Assoc.* **49**, 598–619.
- Holloway, C. A. (1979). On the estimation of convex functions. *Oper. Res.* **27**, 401–407.
- McFadden, D. (1985). Specification of econometric models. Presidential Address, Econometric Society, Fifth World Congress, Cambridge, Massachusetts.
- Pshenichny, B. N. & Danilin, M. Y. (1978). *Numerical methods in extremal problems*. Mir Publishers, Moscow.
- Ratkowsky, D. A. (1983). *Nonlinear regression modeling*, Marcel Dekker, Inc., New York and Basel.
- Rockafellar, R. T. (1970). *Convex analysis*. Princeton University Press, New Jersey.
- Seber, G. A. F. (1977). *Linear regression analysis*. Wiley.
- Wilhelmsen, D. R. (1976). A nearest point algorithm for convex polyhedral cones and applications to positive linear approximation. *Math. Comp.* **30**, 48–57.
- Wu, C. F. (1982). Some algorithms for concave and isotonic regression. *Stud. Management Sci.*, **19**, 105–116.
- Zarantonello, E. H. (1971). Projections on convex sets in Hilbert spaces and spectral theory, Part I: Projections on convex sets. In *Contributions to nonlinear functional analysis*. Academic Press, New York.

Received November 1986, in final form February 1988

H. Massam, Mathematics Department, York University, North York M3J 1P3, Ontario, Canada