Journal of Statistical Planning and Inference
Volume 81, Issue 1, 1 October 1999, Pages 13-31

An extension of the mixed primal–dual bases algorithm to the case of more constraints than dimensions

Mary C. Meyer

https://doi.org/10.1016/S0378-3758(99)00025-7

Abstract

The problem of minimizing a quadratic function with linear inequality constraints is considered with applications to nonparametric regression with shape assumptions. For many problems, the set defined by the constraints is a closed convex cone. The mixed primal–dual bases algorithm (Fraser and Massam, 1989, Scand. J. Statist. 16, 65–75) for regression under inequality constraints finds a least-squares regression estimate over such a cone in a finite number of steps, with the restriction that the number of constraints does not exceed the number of dimensions in the space. Some applications, however, require more constraints than dimensions, and the main purpose of this paper is to extend the algorithm to this more general case. Properties of the constraint cone and its polar cone are presented in the generality necessary in this situation. One surprising result is that the number of generators of the constraint cone can be much larger than the number of generators for the polar cone.

Applications are presented for least-squares regression, and regression in which a roughness penalty term is included in the function to minimize.

Previous Next

Keywords

Regression; Nonparametric; Least-squares; Cone; Dual basis; Smoothing

1. Introduction

The problem of minimizing a quadratic function with linear inequality constraints has many applications, including shape-restricted nonparametric regression. Given $a<t_1<t_2<\cdots<t_n<b$ and data y_i, $i=1,\ldots,n$, consider the model

$$y_i=\mu(t_i)+\varepsilon_i, \tag{1}$$

where the ε_i are mean zero random variables. The object is to estimate the function μ with assumptions about its shape.

The estimator $\boldsymbol{\theta}$ of μ considered here is a vector which minimizes a weighted sum of squares

$$\sum_{i=1}^{n} w_i(y_i-\theta_i)^2 \tag{2}$$

subject to the shape assumptions, such as monotonicity, convexity, or having one point of inflection, or having other more complicated shape constraints. These can be written as linear inequality constraints and expressed in the form $\mathbf{A}\boldsymbol{\theta}\geqslant 0$ where $\mathbf{A}$ is an $m\times n$ matrix which may depend on the t_i. The constraint space $\boldsymbol{\Omega}=\{\boldsymbol{\theta}:\mathbf{A}\boldsymbol{\theta}\geqslant 0\}$ is a closed convex cone, and the vector $\boldsymbol{\theta}$ which minimizes (2) over $\boldsymbol{\Omega}$ is called the *projection* of $\mathbf{y}$ onto the space $\boldsymbol{\Omega}$, written $\Pi_w(\mathbf{y}|\boldsymbol{\Omega})$.

Fraser and Massam (1989) presented an efficient algorithm for projections which makes use of the properties of a convex cone. Unlike steepest descent or interior point methods for quadratic programming problems, their algorithm converges to the exact solution in a finite number of iterations. Their algorithm, however, is restricted to the case in which the matrix $\mathbf{A}$ is full row-rank, that is, the number of constraints does not exceed the number of dimensions

n. This case covers problems such as isotonic and convex regression but does not cover other important constraint sets such as those shown in the following examples.

Example 1 Positive increasing convex–concave function

Growth curves in biology or economics can often be assumed to be positive increasing convex–concave in shape. The number of constraints in this case is one more than the number of dimensions. If the point of inflection is between t_i and t_{i+1}, the nonzero entries of the constraint matrix are

$$A_{j,j}=t_{j+2}-t_{j+1}, A_{j,j+1}=t_j-t_{j+2}, \text{and } A_{j,j+2}=t_{j+1}-t_j$$

for j=1,…,i–1;

$$A_{j,j}=-t_{j+2}+t_{j+1}, A_{j,j+1}=-t_j+t_{j+2}, \text{and } A_{j,j+2}=-t_{j+1}+t_j$$

for j=i,…,n–2; and

$$A_{n-1,1}=1, A_{n,1}=-1, A_{n,2}=1, A_{n+1,n-1}=-1, \text{and } A_{n+1,n}=1.$$

For the simple case of $t=(1 2 3 4)'$, with the inflection point between 2 and 3, the constraint matrix is

$$A=1\,0\,0\,0\,-1\,1\,0\,0\,1\,-2\,1\,0\,0\,-1\,2\,-1\,0\,0\,-1\,1. \tag{3}$$

Example 2 Bell-shaped function

Bell-shaped constraints are useful for density estimation. We want the function to be constrained to be positive, increasing at the left, convex–concave–convex, then decreasing at the right. For arbitrary n, with $t=(t_1,t_2,\ldots,t_n)'$, if the points of inflection are at the t-coordinates indexed by i_1 and i_2, that is, the first and last indices at which the function is concave, then the nonzero entries of the constraint matrix $A_{(n+2\times n)}$ can be written as

$$A_{j,j}=t_{j+2}-t_{j+1}, A_{j,j+1}=t_j-t_{j+2}, \text{and } A_{j,j+2}=t_{j+1}-t_j$$

for j=1,…,i_1–2 and j=i_2–1,…,n–2; and

$$A_{j,j}=t_{j+1}-t_{j+2}, A_{j,j+1}=t_{j+2}-t_j, \text{and } A_{j,j+2}=t_j-t_{j+1}$$

for j=i_1–1,…,i_2–2; with $A_{n-1,1}$=–1.0, and $A_{n-1,2}$=1.0; $A_{n,1}=1.0, A_{n+1,n-1}=-1.0$, and $A_{n+1,1}$=1.0, and

$A_{n+2,n}$=1.0.

For $\mathrm{t}=(1\,2\,3\,4\,5\,6)'$, and points of inflection between 2 and 3, and 4 and 5, the constraint matrix is

$$\mathrm{A}=\begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

The matrix A is *irreducible* if no row of A is a positive linear combination of other rows, and if the origin is also not a positive linear combination of rows of A. As shown later, if a row is a positive linear combination of other rows, it can be removed without affecting the problem. If the origin can be written as a positive linear combination of rows, then there is an implicit equality constraint in the matrix A, which can be dealt with using other methods.

To see that the solution exists, let $\mathrm{f}(\theta)=\sum \mathrm{w_i}(\mathrm{y_i}-\theta_\mathrm{i})^2$. The subset $\Omega'=\Omega\cap\{\theta\colon \mathrm{f}(\theta)\leqslant\sum_{\mathrm{i}=1}^{\mathrm{n}}\mathrm{w_i}\mathrm{y_i}^2\}$ is closed and bounded in R^n. Since $0\in\Omega$ and the function $f(\theta)$ is continuous in R^n, $\mathrm{f}(\theta)$ must achieve its minimum value in Ω. The following important theorem characterizes projections onto convex cones. For further details, see Robertson et al. (1988, p. 17).

Theorem 1

The vector θ minimizes $\|\mathrm{y}-\theta\|^2$ in Ω if and only if

$$\langle \mathrm{y}-\boldsymbol{\theta},\boldsymbol{\theta}\rangle=0, \tag{5}$$

$$\langle \mathrm{y}-\boldsymbol{\theta},\theta\rangle\leqslant 0, \forall\theta\in\Omega. \tag{6}$$

The inner product is defined as $\langle \mathrm{x},\mathrm{y}\rangle=\sum \mathrm{w_i}\mathrm{x_i}\mathrm{y_i}$ and its norm as $\|\mathrm{x}\|^2=\sum \mathrm{w_i}\mathrm{x_i}^2$.

The inner product should be written as $\langle \mathrm{x},\mathrm{y}\rangle_\mathrm{w}$ and the norm as $\|\mathrm{x}\|^2{}_\mathrm{w}$, but without loss of generality, the weights can be taken to be the 1-vector, since a linear change of variable can be made to the unweighted case. Therefore, in the following, the notation for the weighting is dropped.

In the next section, the constraint set is discussed in the generality necessary for the $m>n$ problem. The algorithm obtains the least-squares solution by iterative projecting onto subspaces defined by edges of the constraint cone and its polar cone. When A is full row-rank, there is a one-to-one correspondence between edges of the cones, and they are both easily obtained from the constraint matrix. Further, the edges define a partition of R^n, called sectors, so that the problem can be reduced to determining the sector which contains y. When A is not full row-rank, the edges of the constraint cone are not so easily obtained from the constraint matrix. There are typically more edges of the constraint cone, so there is no longer the convenient one-to-one correspondence of edges, and finally, the sectors may overlap. These complications are what necessitates an extension of the original algorithm.

In Section 3, projections onto the constraint cone are discussed using the geometry of the constraint space. A review of Fraser's and Massam's algorithm is briefly given in Section 4, and the extension is presented. Finally, some examples are shown in Section 5.

2. The constraint and polar cones

The constraint space is a convex polyhedral cone Ω. Let $[\gamma^1 \dots \gamma^m] = -A'$. Then the constraint space may be written as

$$\Omega = \{\theta : \langle \gamma^i, \theta \rangle \leqslant 0 \text{ for } i = 1, \dots, m\}. \tag{7}$$

The polar cone (see Rockafellar, 1970) is defined as

$$\Omega^o = \{\rho : \langle \theta, \rho \rangle \leqslant 0, \forall \theta \in \Omega\}.$$

Clearly, $\gamma^1 \dots \gamma^m \in \Omega^o$. These vectors are *generators* of the polar cone, i.e., each $\rho \in \Omega^o$ can be written as a nonnegative linear combination of the γ^j. To see this, let R be the cone generated by the γ^i, that is,

$$R = \sum i=1 m a_i \gamma^i : a_i \geqslant 0. \tag{8}$$

Then the constraint space Ω is polar to the cone R, since for any $\theta \in \Omega$,

$$\langle \theta, \rho \rangle = \sum i=1 m a_i \langle \theta, \gamma^i \rangle \leqslant 0$$

for all $\rho = \sum a_j \gamma^j \in R$. This shows that $\Omega \subseteq R^0$. For any $\theta \in R^0$, we have $\langle \theta, \gamma^i \rangle \leqslant 0, \forall i = 1, \dots, m$, so that $R^0 \subseteq \Omega$, and, therefore, $\Omega = R^0$. This implies that $\Omega^0 = R^{00} = R$. See Rockafellar (1970, p. 121).

Note that the requirement that A be irreducible ensures that the dimension of the constraint cone Ω is *n*, where the dimension of a cone is taken to be the dimension of the smallest linear space which contains the cone. For, if the dimension of Ω were less than *n*, then a vector v, say, could be found such that $\langle v, \theta \rangle = 0$ for all θ in Ω, so that both v and $-v$ were in the polar cone. This would contradict the assumption that the origin is not a positive linear combination of rows of A.

Let

$$L=\operatorname{span}\{\gamma^1,\ldots,\gamma^m\}. \tag{9}$$

There is no loss of generality in the assumption that $\dim L=n$, since otherwise, let $\Omega_L=L\cap\Omega$ and $y_L=\Pi(y|L)$. It is easily seen that $\Omega=\Omega_L+L^\perp$, where $L^\perp$ denotes the orthogonal complement of L, and $\Pi(y|\Omega)=\Pi(y_L|\Omega_L)+\Pi(y|L^\perp)$. The problem is reduced to finding $\Pi(y_L|\Omega_L)$, where the data to be projected is in the space spanned by the constraints. Therefore in the following, $L=R^n$ and $m\geqslant n$.

The generators γ^i are actually *edges* of the polar cone if A is irreducible. An "edge" of a convex polyhedral cone, which is sometimes called an "extreme edge", is a vector in the cone which cannot be written as the sum of two linearly independent vectors in the cone. To see this, suppose γ^j is not an edge of the polar cone. Then there exist linearly independent r^1 and r^2 in Ω^0 such that $r^1+r^2=\gamma^j$. Write $r^1=\sum_{i=1}^m c_i^1\gamma^i$ and $r^2=\sum_{i=1}^m c_i^2\gamma^i$, where the coefficients are all nonnegative. Then

$$\gamma^j=(c^1{}_j+c^2{}_j)\gamma^j+\sum_{i\neq j}(c^1{}_i+c^2{}_i)\gamma^i.$$

If $1-(c^1{}_j+c^2{}_j)>0$, then we have written γ^j as a positive linear combination of the other γ^i, so that A is reducible. If $1-(c^1{}_j+c^2{}_j)=0$, we have written the origin as a positive linear combination of the γ^i. If $1-(c^1{}_j+c^2{}_j)<0$, we have

$$-\gamma^j=-\sum_{i\neq j}\frac{c^1{}_i+c^2{}_i}{1-(c^1{}_j+c^2{}_j)}\gamma^i,$$

so that both γ^j and $-\gamma^j$ are in the polar cone. Therefore, if A is irreducible, the negative rows of the constraint matrix are edges of the polar cone.

Note that if a row of the constraint matrix is a positive linear combination of other rows, then that row can be removed without changing the constraint cone: Let $H_i=\{\theta:\langle\gamma^i,\theta\rangle\leqslant 0\}$ be the closed half-space perpendicular to γ^i, not containing γ^i, for each $i=1,\ldots,m$. Then the constraint cone can be written as $\bigcap_{i=1}^m H_i$. If $\gamma^j=\sum_{i\neq j}c_i\gamma^i$ where all the $c_i\geqslant 0$, then it is easily seen that $\bigcap_{i=1}^m H_i=\bigcap_{i\neq j}H_i$, so that the jth row of A may be removed without changing the problem.

The following proposition describes the edges of the constraint cone, assuming A is irreducible.

Proposition 1

Let $J\subseteq\{1,\ldots,m\}$. If $\dim \mathrm{span}\{\gamma^j,j\in J\}=n-1$, then $\mathrm{span}\{\gamma^j,j\in J\}^\perp$ is a line through the origin containing the vectors δ and $-\delta$, say, where $\delta\perp\gamma^j,\forall j\in J$. If $\langle\delta,\gamma^i\rangle\leqslant 0,\forall i\notin J$, then δ is an edge of Ω. Conversely, all edges are of this form.

Proof.

Let $\delta\in\mathrm{span}\{\gamma^j,j\in J\}^\perp$ where $\dim\mathrm{span}\{\gamma^j,j\in J\}=n-1$. If $\delta=\theta_1+\theta_2$ for θ_1 and $\theta_2\in\Omega$, then $\langle\delta,\gamma^i\rangle=\langle\theta_1,\gamma^i\rangle+\langle\theta_2,\gamma^i\rangle=0$ for all $i\in J$. Therefore $\langle\theta_1,\gamma^i\rangle=\langle\theta_2,\gamma^i\rangle=0$ for all $i\in J$, so that $\theta_1,\theta_2\in\mathrm{span}\{\gamma^j,j\in J\}^\perp$. But this is one-dimensional so θ_1,θ_2 are multiples of δ, and so δ is an edge.

Conversely, suppose δ is an edge, and let $J=\{j:\langle\delta,\gamma^j\rangle=0\}$, and $L_J=\mathrm{span}\{\gamma^j,j\in J\}$. If $\dim L_J<n-1$, then there exists an $\eta\in L_J{}^\perp$, linearly independent to δ. For $j\notin J$,

$$\langle\delta\pm\varepsilon\eta,\gamma^j\rangle<0$$

for some $\varepsilon>0$, which contradicts the assumption that δ is an edge; therefore, $\dim L_J=n-1$ and the proposition is proved. □

Let $\Delta=[\delta^1,\ldots,\delta^M]'$ be the matrix whose rows are the edges of the constraint cone. The following proposition demonstrates properties of the duality of the constraint and polar cones.

Proposition 2

The matrix Δ of edges of the constraint cone is irreducible and is the negative of the constraint matrix for the polar cone, i.e., $\rho\in\Omega^0$ if and only if

$$-\Delta'\rho\geqslant 0. \tag{10}$$

Proof.

If there is a set $I_0\subseteq\{1,\ldots,M\}$ such that $\sum_{i\in I_0}a_i\delta^i=0$, with $a_i>0,\forall i\in I_0$, then $\#I_0\geqslant 2$ and for any $j\in I_0$, we can write

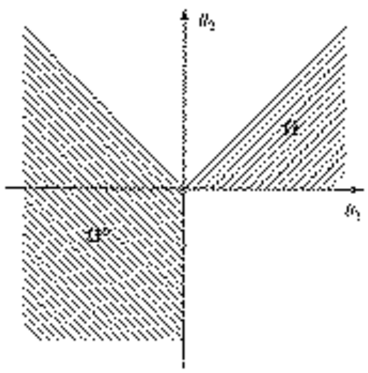
$$a_j\delta^j+\sum_{i\in I_0,i\neq j}a_i\delta^i=0.$$

Since the second term is in Ω, we have that $-\delta^j\in\Omega$, which implies that δ^j is perpendicular to all the $\gamma^i,i=1,\ldots,m$, which contradicts the assumption that the γ^i span R^n. The fact that each row of Δ is an edge ensures that none can be written as a positive linear combination of other rows, so that Δ is irreducible. Now, there is a cone C such that $C=\{\rho:-\Delta'\rho\geqslant 0\}$ and the vectors δ^i are both the edges and the generators of its polar C^0. Clearly, $C^0=\Omega$ and $C=\Omega^0$. □

If $\mathbf{A}$ is of full row-rank, that is, the γ^j are linearly independent and $m=n$, the situation is greatly simplified. It is easily seen that the edges of the constraint cone are

$$[\delta^1,\ldots,\delta^n]=\Delta=-(\Gamma')^{-1}, \tag{11}$$

where $\Gamma=[\gamma^1,\ldots,\gamma^n]$. Clearly $\langle\gamma^i,\delta^j\rangle=-1$ or 0 accordingly as $i=j$ or $i\neq j$. An illustration of the cone and polar cone for $n=2$ and $\gamma^1=(0,-1)',\gamma^2=(-1,1)',\delta^1=(1,1)'$, and $\delta^2=(1,0)'$ is shown in Fig. 1.



Download : Download full-size image

Fig. 1. The constraint and polar cones.

If $m>n$, then the number of generators M of the constraint cone is often greater than m. When n is small, the δ's can be found by making an exhaustive search of all possible combinations of hyperplanes, checking to see if the intersection is in the constraint cone. In Example 1, the matrix of edges of the constraint cone is

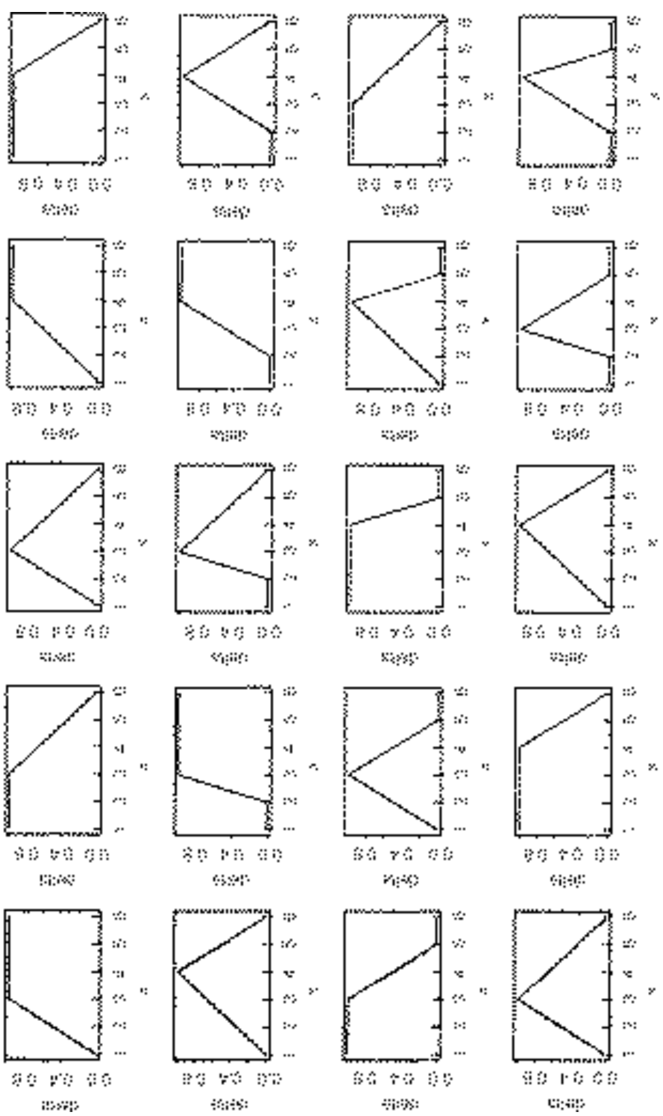
$$\Delta=00110012012201231111. \tag{12}$$

With some experimentation, it can be seen that all piecewise linear functions on (1,2,3,4) with positive increasing convex–concave constraints can be written as nonnegative linear combinations of the five rows of this matrix.

When n is large, the number of possible combinations becomes prohibitive. Patterns in the $\mathbf{A}$ matrix can be used to figure out patterns for the edges. The number of edges of the constraint cone for the bell-shaped constraints can be much larger than the number of constraints for large n. The matrix of 21 edges of the constraint cone for Example 2 where the matrix $\mathbf{A}$ is as in (4) is

$$\begin{gathered}\Delta=0.0000\;0.5000\;1.0000\;1.0000\;1.0000\;1.00001.0000\;1.0000\;1.0000\;0.6667\;0.3333\\ 0.00000.0000\;0.5000\;1.0000\;0.6667\;0.3333\;0.00000.0000\;0.3333\;0.6667\;1.0000\;1.0000\\ 1.00001.0000\;1.0000\;1.0000\;1.0000\;0.5000\;0.00000.0000\;0.3333\;0.6667\;1.0000\;0.5000\\ 0.00000.0000\;0.0000\;1.0000\;1.0000\;1.0000\;1.00000.0000\;0.0000\;1.0000\;0.6667\;0.3333\\ 0.00000.0000\;0.0000\;0.5000\;1.0000\;1.0000\;1.00000.0000\;0.0000\;0.5000\;1.0000\;0.5000\\ 0.00001.0000\;1.0000\;1.0000\;0.5000\;0.0000\;0.00000.0000\;0.5000\;1.0000\;0.5000\;0.0000\\ 0.00001.0000\;1.0000\;1.0000\;1.0000\;0.0000\;0.00000.0000\;0.3333\;0.6667\;1.0000\;0.0000\\ 0.00001.0000\;1.0000\;1.0000\;0.6667\;0.3333\;0.00000.0000\;0.5000\;1.0000\;0.6667\;0.3333\\ 0.00001.0000\;1.0000\;1.0000\;1.0000\;0.5000\;0.00000.0000\;0.3333\;0.6667\;1.0000\;0.5000\\ 0.00000.0000\;0.0000\;1.0000\;0.5000\;0.0000\;0.00000.0000\;0.0000\;0.5000\;1.0000\;0.0000\\ 0.00001.0000\;1.0000\;1.0000\;1.0000\;1.0000\;1.0000.\end{gathered}\tag{13}$$

The first 20 of the $M=21\,\delta$ vectors are shown in Fig. 2. The *y*-coordinates of any bell-shaped function on the *t*-coordinates can be written as a positive linear combination of these vectors.



Download : Download full-size image

Fig. 2. Edges of the constraint cone for Example 2.

If n=50, with points of inflection at i_1=16 and i_2=35, then the number of edges of the constraint cone is found to be M=5141.

3. Projections onto the cones

In the following sections, $\mathbf{A}$ is assumed to be an $m\times n$ irreducible matrix, where $m\geqslant n$. In practice, when $m<n$, the data can be projected onto the linear subspace spanned by the rows of the constraint matrix, so no generality is lost with this assumption. For $\mathbf{y}\in\mathbb{R}^n$, let $\boldsymbol{\theta}=\Pi(\mathbf{y}\,|\,\Omega)$ and $\hat{\rho}=\Pi(\mathbf{y}\,|\,\Omega^0)$. Theorem 1 can be stated in terms of both $\boldsymbol{\theta}$ and $\hat{\rho}$. The vector $\boldsymbol{\theta}$ minimizes $||\mathbf{y}-\theta||^2$ in Ω if and only if

$$\langle \mathbf{y}-\boldsymbol{\theta},\boldsymbol{\theta}\rangle=0$$

and

$$\langle \mathbf{y}-\boldsymbol{\theta},\delta^i\rangle\leqslant 0\ \text{for}\ i=1,\ldots,M.$$

Theorem 1 may also be applied to the polar cone and yields the following characterization of $\hat{\rho}$: The vector $\hat{\rho}$ minimizes $||\mathbf{y}-\rho||^2$ in Ω^0 if and only if $\langle \mathbf{y}-\hat{\rho},\hat{\rho}\rangle=0$, and $\langle \mathbf{y}-\hat{\rho},\gamma^i\rangle\leqslant 0$, for i=1,…,m.

Since $\boldsymbol{\theta}\in\Omega$, there are a set $I\subseteq\{1,\ldots,M\}$ and coefficients $b_i>0, \forall i\in I$, for which

$$\boldsymbol{\theta}=\sum_{i\in I} b_i\delta^i. \tag{14}$$

Similarly, there is a set $J\subseteq\{1,\ldots,m\}$ and coefficients $a_j>0,\forall j\in J$ so that

$$\hat{\rho}=\sum_{j\in J} a_j\gamma^j. \tag{15}$$

In the following proposition, it is shown that the residual of $\boldsymbol{\theta}$ is $\hat{\rho}$, and vice versa.

Proposition 3

The data vector is the sum of the two projections, that is, $\mathbf{y}=\boldsymbol{\theta}+\hat{\rho}$.

Proof:

Let $r=y-\theta$. Then $r\in\Omega^o$ since $\langle r,\delta^i\rangle\leqslant 0$ for all $i=1,\ldots,M$. Also, $\langle r,y-r\rangle=\langle y-\theta,\theta\rangle=0$, and

$\langle y-r,\gamma^j\rangle=\langle\theta,\gamma^j\rangle\leqslant 0,\forall j=1,...,m,$

since $\theta\in\Omega$. So $r=\Pi(y|\Omega^0)=\hat{\rho}$. □

Note that $\langle\delta^i,\gamma^j\rangle=0$ for $i\in I$ and $j\in J$, since $0=\langle\hat{\rho},\theta\rangle=\sum_{i\in I}\sum_{j\in J}a_j b_i\langle\delta^i,\gamma^j\rangle$, and the a_j's and b_i's are strictly positive.

Proposition 4

Let I *be as in* (14). *Then the projection* θ *of* y *on* Ω *is the projection of* y *on the linear space spanned by* $\{\delta^i;i\in I\}$, *i.e.*, $\theta=\Pi(y|\delta^i;i\in I)$. *Similarly, the projection* $\hat{\rho}$ *of* y *on* Ω^0 *is* $\Pi(y|\gamma^j;j\in J)$.

Proof.

By (14), θ is in $\text{span}\{\delta^i:i\in I\}$, and

$\langle y-\theta,\theta\rangle=\sum i\in Ib_i\langle y-\theta,\delta^i\rangle$

with the $b_i>0$. By Proposition 3, this is zero, so $\langle y-\theta,\delta^i\rangle=0$, for all $i\in I$, that is, $y-\theta$ is orthogonal to the space spanned by the $\delta^i,i\in I$. □

Note that for $m>n$, the linear combinations in , are not necessarily unique. In either case, each set $J\subseteq\{1,...,m\}$ determines a sector

$\Omega_J=\sum j\in Ja_j\gamma^j+\sum i\in Ib_j\delta^j:a_j>0,j\in J,\text{and}\,b_j\geqslant 0,i\in I,$

where $I=I(J)=\{i:\langle\gamma^j,\delta^i\rangle\}=0,\forall j\in J$, and these sectors cover R^n. Note that if $J=\{1,...,m\}$, then Ω_J is the interior of Ω^0, and if $J=\varphi$, then $\Omega_J=\Omega$. The following proposition determines how much the sectors may overlap.

Proposition 5

If $m=n$, *these sectors are disjoint; for* $m>n$, *they can overlap. However, if* J_1 *and* J_2 *are such that both* $\{\gamma^j,j\in J_1\}$ *and* $\{\gamma^j,j\in J_2\}$ *are linearly independent sets with* $\text{span}\{\gamma^j,j\in J_1\}\neq\text{span}\{\gamma^j,j\in J_2\}$, *the sectors* Ω_{J_1} *and* Ω_{J_2} *can overlap only in a set of Lebesgue measure zero.*

Proof.

Suppose $x \in R^n$ can be written as $x = \sum_{j \in J_1} a_j^1 \gamma^j + \sum_{i \in I_1} b_i^1 \delta^i$ and also as $\sum_{j \in J_2} a_j^2 \gamma^j + \sum_{i \in I_2} b_i^2 \delta^i$ with $a_j^k > 0, b_i^k \geqslant 0$, and $I_k = I_k(J_k) = \{i: \langle \gamma^j, \delta^i \rangle\} = 0, \forall j \in J_k\}, k = 1,2$. Then the first term in either sum is the unique point $\Pi(x|\Omega^0)$, so that if $m=n$ (the γ^j are linearly independent), we must have $J_1 = J_2$. If $m>n$, then $\Pi(x|\gamma^j, j \in J_1) = \Pi(x|\gamma^j, j \in J_2)$, so that $\Omega_{J_1} \cap \Omega_{J_2} = \{x: (P_1 - P_2)x = 0\}$. Here P_k is the matrix which projects a point in R^n onto $\{\gamma^j, j \in J_k\}$, for $k=1,2$. If $P_1 \neq P_2$, then $\Omega_{J_1} \cap \Omega_{J_2}$ has measure zero. □

The problem of projecting the data y onto the constraint cone is reduced to the problem of finding a set J so that $\hat{\rho} = \Pi(y|\gamma^j, j \in J)$ is in the polar cone and $\theta = y - \hat{\rho}$ is in the constraint cone. Alternatively, we can find an I so that $\theta = \Pi(y|\delta^i, i \in I)$ is in the constraint cone and $\hat{\rho} = y - \theta$ is in the polar cone.

4. The mixed primal–dual bases algorithm

Fraser and Massam call the γ^j's *primal* vectors and the δ^j's *dual* vectors. Their *mixed primal–dual bases algorithm* finds the correct set J by moving along a line segment connecting a starting point y^0 with the data y. A point y^k is reached at step k of the algorithm, in such a manner that the distance between y^k and y is strictly decreasing in k. A set J_k is determined so that y^k is on the face of the sector Ω_{J_k}, and in fact, the point y^{k+1} is on both the face of $\Omega_{J_{k+1}}$ and Ω_{J_k}. The algorithm stops when $J_k = J$, i.e., the point y^k is on the face of the sector containing the data y, which is sure to happen in a finite number of steps, since there are only a finite number of sectors. In this way, J_k is determined, and $\theta = y - \Pi(y|\gamma^j, j \in J)$.

The case $m=n$ (the original version) is reviewed in the next section, and the extension to $m>n$ is presented in Section 4.2.

4.1. The Case $m=n$

If $m=n$, any subset J of $\{1,\ldots,n\}$ determines a basis for R^n, denoted by $B_J = \{\gamma^j, j \in J; \delta^j, j \notin J\}$. The sectors Ω_J with $I(J) = J^C$ partition R^n. For any J and any $y \in R^n$, there exist coefficients b_j such that

$$y = \sum j \in J b_j \gamma^j + \sum j \notin J b_j \delta^j = \rho + \theta,$$

where the $\delta^j, j = 1,\ldots,n$, are given in (11). If $b_j \geqslant 0$ for $j=1,\ldots,n$, then ρ is the projection of y on Ω^0 and θ is the projection of y on Ω. There is a unique set J for which $b_j > 0, \forall j \in J$ and $b_j \geqslant 0, \forall j \notin J$. If this set is known, then θ can be found either by projecting the data y onto the subspace spanned by the $\{\delta^j, j\notin J\}$ or by projecting y onto the subspace spanned by the $\{\gamma^j, j\in J\}$ and subtracting this projection from y. The following is a brief description of the mixed primal–dual bases algorithm for $m=n$. For more details, see Fraser and Massam (1989).

The starting point y^0 for the algorithm may be chosen in the interior of any sector, and the coordinates written in the basis corresponding to that sector, so that the coefficients are strictly positive. For example, choose $J_0=\emptyset$, so that the basis B_0 is $\{\delta^j, j=1,\ldots,n\}$ and

$$y^0=\sum_{j=1}^{n}\delta^j.$$

The data can be written in coordinates corresponding to this basis

$$y=\sum_{j=1}^{n}b_j^0\delta^j$$

and the coefficients are unique. If the $b_j^0\geqslant 0$, then the algorithm stops since $y\in\Omega$. Otherwise, a point y^1 on the line segment between y^0 and y is computed. It is the point on the line segment closest to y and still in (the closure of) the sector Ω. In this way, the set J_1 is determined so that the neighboring sector is Ω_{J_1}, i.e., the point y^1 is on the boundary between Ω and Ω_{J_1}. It can be seen that the set J_1 is a singleton, say $J_1=\{j_0\}$. Both y^1 and y can be written in the coordinates corresponding to J_1:

$$y^1=\sum_{j\notin J_1}a_j\delta^j+a_{j_0}\gamma^{j_0}$$

and

$$y=\sum_{j\notin J_1}b_j\delta^j+b_{j_0}\gamma^{j_0}.$$

Note that all the a_j are nonnegative and a_{j_0} is strictly positive. If the b_j are nonnegative, then $J=J_1$ and the algorithm stops. Otherwise, the point y^2 is computed to be the point on the line segment closest to y and still in (the closure of) the sector Ω_{J_1}. The algorithm continues until the point y^k is on the face of the sector containing y, and thus J is determined.

At each step, a dual vector is exchanged for its primal vector in the basis set, or vice versa. There is a one-to-one correspondence of these edges, so that the new set is guaranteed to span R^n.

The projection θ can also be found by starting in the polar cone with

$$y^0=\sum_{j=1}^{n}\gamma^j,$$

or in any other sector, as long as $\mathbf{y}^0$ is a point in the interior of that sector. The points $\mathbf{y}^0$ and $\mathbf{y}$ are written in coordinates for that sector, and the algorithm moves toward $\mathbf{y}$ in the same manner.

4.2. The Case *m*>*n*

When the number of constraints is greater than the number of dimensions, the expression for $\boldsymbol{\theta}$ as a nonnegative linear combination of constraint cone edges is no longer unique. A given sector may be defined by more than one distinct set of edges, so that when the line segment described above passes from one sector to another, it is not clear how to exchange basis vectors.

In this section, the mixed primal–dual bases algorithm is extended to the case in which $m>n$, with the steps written out in detail. The idea is the same in that $\mathbf{y}^k$ moves from a starting point towards $\mathbf{y}$, but the algorithm has to be structured differently. The coefficients for $\mathbf{y}$ and $\mathbf{y}^k$ are not necessarily unique for a given $\mathbf{J}$, that is, the set $\{\delta^i;i\in \mathbf{I}=\mathbf{I}(\mathbf{J})\}$ is not necessarily linearly independent. Furthermore, there is no longer a one-to-one correspondence between primal and dual vectors.

Therefore, for this case, the goal is to find $\mathbf{J}$ such that

$$\mathbf{y}=\sum_{j\in \mathbf{J}} b_j\gamma^j+\boldsymbol{\theta},$$

where $b_j>0$ and $\boldsymbol{\theta}\in\Omega$. There exists a (not necessarily unique) set $\mathbf{I}$ such that $\boldsymbol{\theta}=\sum_{i\in \mathbf{I}} b_i\delta^i$, with $b_i>0$, but the $\mathbf{I}$ is not found and the coefficients for $\boldsymbol{\theta}$ are not computed. As in the case for $m=n$, at each iteration, a set $\mathbf{J}_k$ and a $\mathbf{y}^k$ are computed. Each $\mathbf{y}^k$ is written as $\mathbf{r}^k+\mathbf{t}^k$, where $\mathbf{r}^k=\Pi(\mathbf{y}^k\,|\,\Omega^0)$ and $\mathbf{t}^k=\Pi(\mathbf{y}^k\,|\,\Omega)$. The data $\mathbf{y}$ is written as $\rho^k+\theta^k$ by projecting $\mathbf{y}$ onto the $\gamma^j, j\in \mathbf{J}_k$, so that $\rho^k=\Pi(\mathbf{y}\,|\,\gamma^j, j\in \mathbf{J}_k)$ and $\theta^k=\mathbf{y}-\rho^k$. The algorithm stops when $\rho^k\in\Omega^0$ and $\theta^k\in\Omega$.

Again, any starting set $\mathbf{J}_0$ can be chosen, such as $\mathbf{J}_0=\emptyset$, and $\mathbf{y}^0$ can be any point in the interior of the sector $\Omega_{\mathbf{J}_0}$, which is the constraint cone if $\mathbf{J}_0=\emptyset$. Suppose the starting point is chosen to be

$$\mathbf{y}^0=\sum_{i=1}^{M}\delta^i.$$

Note that $\langle \mathbf{y}^0,\gamma^j\rangle<0$, for all $j=1,\ldots,m$. If $\mathbf{y}\notin\Omega$, choose

$$\mathbf{y}^1=\mathbf{y}^0+\alpha_1(\mathbf{y}-\mathbf{y}^0)$$

with α_1 as large as possible so that y^1 remains in the closure of the starting sector. This gives

$$\alpha_1=\min\{j:\langle y,\gamma^j\rangle>0\}-\langle y^0,\gamma^j\rangle\langle y,\gamma^j\rangle-\langle y^0,\gamma^j\rangle.$$

The requirement that y^0 be in the interior of the starting sector ensures that $\alpha_1\in(0,1)$.

If the minimum occurs at $j=j_0$ with no ties (the *no ties* criterion will be discussed in more detail later), then $J_1=\{j_0\}$. The point y^1 is in both the constraint cone and the closure of the sector $\Omega_{J_1}=\{a_{j_0}\gamma^{j_0}+\sum_{i\in I_1}a_i\delta^i:a_{j_0}>0,a_i\geqslant0,\forall i\in I_1\}$ where $I_1=\{i:\langle\gamma^{j_0},\delta^i\rangle=0\}$. Then y is written in the basis corresponding to J_1, that is, $y=\rho^1+\theta^1$, where $\rho^1=\Pi(y|\gamma^{j_0})=b^1{}_{j_0}\gamma^{j_0}$. The coefficient $b^1{}_{j_0}=\langle y,\gamma^{j_0}\rangle/||\gamma^{j_0}||^2$ is computed, but there is no need to compute the coefficients for θ^1. The point ρ^1 has to be in the polar cone because j_0 was chosen so that $b^1{}_{j_0}>0$. If $\theta^1=y-\rho^1$ is in the constraint cone, the algorithm stops. Otherwise, it goes on to find y^2. For this first step, $y^1=r^1+t^1$, where $r^1=0$ and $t^1=y^1$.

At the kth step, $y^k=r^k+t^k$, where r^k and t^k are the projections of y^k on Ω^0 and Ω, respectively. The set $J_k\subset\{1,\ldots,m\}$ is such that $\{\gamma^j,j\in J_k\}$ are linearly independent, and

$$r^k=\sum j\in J_k a_j{}^k\gamma^j$$

can be written uniquely. Also, $a_j^k>0$ for $j\in J_k$ and $\langle t^k,\gamma^j\rangle<0$ for $j\notin J_k$, *except* for the previous minimizing index j_{k-1}. If $j_{k-1}\in J_k$ we have $a_{j_{k-1}}{}^k=0$, otherwise, $\langle t^k,\gamma^{j_{k-1}}\rangle=0$. The data y can be written as $\rho^k+\theta^k$, where

$$\rho^k=\sum j\in J_k b_j{}^k\gamma^j$$

is the projection of y onto $\mathrm{span}\{\gamma^j,j\in J_k\}$, and the b_j^k are unique. If $\theta^k=y-\rho^k$ is in the constraint cone and ρ^k is in the polar cone, the algorithm stops. Otherwise, it goes on to find

$$y^{k+1}=y^k+\alpha_{k+1}(y-y^k)=[(1-\alpha_{k+1})r^k+\alpha_{k+1}\rho^k)]+[(1-\alpha_{k+1})t^k+\alpha_{k+1}\theta^k)]=r^{k+1}+t^{k+1}$$

with α_{k+1} as large as possible so that $r^{k+1}\in\Omega^0$ and $t^{k+1}\in\Omega$. The latter conditions are equivalent to

$$\alpha_{k+1}\leqslant a_j{}^k a_j{}^k-b_j{}^k,\forall j\in J_k\ni b_j{}^k<0 \tag{16}$$

and

$$\alpha_{k+1}\leqslant-\langle t^k,\gamma^j\rangle\langle\theta^k,\gamma^j\rangle-\langle t^k,\gamma^j\rangle,\forall j\notin J_k\ni\langle\theta^k,\gamma^j\rangle>0. \tag{17}$$

The minimizing index j_k can either be in J_k or not. If $j_k\in J_k$, then $J_{k+1}=J_k-\{j_k\}$ and necessarily

$a_{j_k}{}^k a_{j_k}{}^k - b_{j_k}{}^k < a_j{}^k a_j{}^k - b_j{}^k, \forall j \in J_k, j \neq j_k, \ni b_j{}^k < 0$

and

$a_{j_k}{}^k a_{j_k}{}^k - b_{j_k}{}^k < -\langle t^k, \gamma^j \rangle \langle \theta^k, \gamma^j \rangle - \langle t^k, \gamma^j \rangle, \forall j \notin J_k, \ni \langle \gamma^j, \theta^k \rangle > 0.$

If $j_k \notin J_k$, then $J_{k+1} = J_k \cup \{j_k\}$ and the conditions for j_k are

$-\langle t^k, \gamma^{j_k} \rangle \langle \theta^k, \gamma^{j_k} \rangle - \langle t^k, \gamma^{j_k} \rangle < a_j{}^k a_j{}^k - b_j{}^k, \forall j \in J_k \ni b_j{}^k < 0$

and

$-\langle t^k, \gamma^{j_k} \rangle \langle \theta^k, \gamma^{j_k} \rangle - \langle t^k, \gamma^{j_k} \rangle < -\langle t^k, \gamma^j \rangle \langle \theta^k, \gamma^j \rangle - \langle t^k, \gamma^j \rangle, \forall j \notin J_k, j \neq j_k \ni \langle \gamma^j, \theta^k \rangle > 0.$

The new set $\{\gamma^j, j \in J_{k+1}\}$ is linearly independent since for any γ^l with $\gamma^l \cup \{\gamma^j, j \in J_k\}$ linearly dependent, $\langle \theta^k, \gamma^l \rangle = 0$ so the index l could not have been chosen. The choice of α_{k+1} ensures that $t^{k+1} \in \Omega$ and $r^{k+1} \in \Omega^0$.

It is necessary to show that $\alpha_{k+1} > 0$, that is, the point progresses toward y. If j_{k-1} was moved into J_k so that $J_k = J_{k-1} \cup \{j_{k-1}\}$, then

$b^k{}_{j_{k-1}} = \langle y, \tilde{\gamma}^{j_{k-1}} \rangle || \tilde{\gamma}^{j_{k-1}} ||^2,$

where $\tilde{\gamma}^{j_{k-1}} = \gamma^{j_{k-1}} - \Pi(\gamma^{j_{k-1}} \mid \gamma^j, j \in J_{k-1})$. Now,

$\langle y, \tilde{\gamma}^{j_{k-1}} \rangle = \langle y - \rho^{k-1}, \tilde{\gamma}^{j_{k-1}} \rangle = \langle y - \rho^{k-1}, \gamma^{j_{k-1}} \rangle = \langle \theta^{k-1}, \gamma^{j_{k-1}} \rangle > 0,$

since j_{k-1} was the minimizing index, so that $b_{j_{k-1}}{}^k > 0$.

If j_{k-1} was removed from J_k so that $J_k = J_{k-1} - \{j_{k-1}\}$, then we had

$0 > b_{j_{k-1}}{}^{k-1} = \langle y, \tilde{\gamma}^{j_{k-1}} \rangle \| \tilde{\gamma}^{j_{k-1}} \|^2,$

where $\tilde{\gamma}^{j_{k-1}} = \gamma^{j_{k-1}} - \Pi(\gamma^{j_{k-1}} \mid \gamma^j, j \in J_k)$. Now,

$\langle \theta^k, \gamma^{j_{k-1}} \rangle = \langle y - \rho^k, \tilde{\gamma}^{j_{k-1}} \rangle = \langle y - \rho^k, \tilde{\gamma}^{j_{k-1}} \rangle = \langle y, \tilde{\gamma}^{j_{k-1}} \rangle$

so that $\langle \theta^k, \gamma^{j_{k-1}} \rangle < 0.$

The vectors r^{k+1} and t^{k+1} are projections of y^{k+1} on Ω^0 and Ω, respectively, since

$\langle r^{k+1}, t^{k+1} \rangle = (1 - \alpha_{k+1})^2 \langle r^k, t^k \rangle + (1 - \alpha_{k+1}) \alpha_{k+1} \langle r^k, \theta^k \rangle + \alpha_{k+1} (1 - \alpha_{k+1}) \langle \rho^k, t^k \rangle + \alpha_{k+1}{}^2 \langle \rho^k, \theta^k \rangle$

and all four of the inner products are zero.

The point $\mathbf{y}^{\mathbf{k}}$ is on a face of the sector Ω_{J_k} since $(1-\varepsilon)\mathbf{y}^{\mathbf{k}}+\varepsilon\mathbf{y}$ is not in Ω_{J_k} for any ε in (0,1). The algorithm stops since there are only a finite number of sectors for the line segment to pass through.

A “tie” occurs in the minimization step if two choices for j_k are possible. This happens when a point on the line segment is on the boundary of three sectors at once: the sector Ω_{J_k}, and two new sectors. The two new sectors are both generated by edges indexed by set J_k, with the addition or subtraction of one of the indices. Either the two new sectors coincide, or the overlap between them has measure zero. If the former, either index can be chosen to form J_{k+1}. The probability of a tie in the latter case is zero.

In the case $m=n$, the event of a tie occurring also has measure zero in theory, although a tie occurring within roundoff error of the computer has a very small but finite probability. Fraser and Massam suggest the following remedy for a tie: move along the face of the previously crossed sector, so that the line segment is moved over slightly and the tie will not occur. Alternatively, the starting point can be moved a bit, and the algorithm restarted. This remedy can also be applied to the $m>n$ case for the second type of tie. It happens that in the author's many simulations, a tie does not occur.

Note that the δ^j are not used in the algorithm except to calculate $\mathbf{y}_0$. Another point in the interior of the constraint cone may be used to avoid computing the δ^j. Alternatively, if the δ^j are known, they may be used as the basis vectors, so that the coefficients for θ^k are computed at each step rather than those for ρ^k.

5. Applications and simulations

The shape constrained regression function is often not very smooth, but it is not difficult to produce a smoother regression function by adding a term to penalize roughness. For example, the criterion function

$$\sum_{i=1}^{n} w_i (y_i - f(t_i))^2 + \lambda \int_a^b f''(t)^2\, dt, \tag{18}$$

where $a=t_0<t_1<\cdots<t_n<t_{n+1}=b$ produces a smoother fit. The scalar λ is called the “smoothing parameter”, since larger values of λ result in smoother fitted functions. The unconstrained minimizer of (18) is the well-known *natural cubic smoothing spline*. Some good references for

smoothing splines are Wahba (1990), Silverman (1985), and Wegman and Wright (1983). Expression (18) can easily be minimized under the shape constraints $A\theta \geqslant 0$ if $\theta = (f(t_1)\ldots f(t_n))'$ is restricted to the class of natural cubic splines. The function f is a natural cubic spline if $f''(t)$ is piecewise linear and continuous on $[a,b]$, and $f''=0$ on $[t_0,t_1]$ and $[t_n,t_{n+1}]$. If $f'''(t)=\alpha_i$ on (t_{i-1},t_i), the integral term of the criterion function can be written as

$$\int_a^b f''(t)^2\,dt = f''(t)f'(t)\big|_a^b - \int_a^b f'''(t)f'(t)\,dt = -\sum_{i=2}^n \int_{t_{i-1}}^{t_i} f'(t)\,\alpha_i\,dt = -\sum_{i=2}^n \alpha_i (f(t_i)-f(t_{i-1})) = \sum_{i=1}^n f(t_i)(\alpha_{i+1}-\alpha_i) = \sum_{i=1}^n f(t_i)\beta_i,$$

where $\beta_i = \alpha_{i+1}-\alpha_i, i=1,\ldots,n$. Now the criterion function can be written as a constant plus:

$$\sum_{i=1}^n w_i f(t_i)^2 - 2\sum_{i=1}^n w_i f(t_i) y_i + \alpha \sum_{i=1}^n f(t_i)\beta_i,$$

or, in vector form,

$$\theta' D\theta - 2\theta' Dy + \lambda\theta'\beta,$$

where $\theta_i=f(t_i)$, for $i=1,\ldots,n$, and D is a diagonal matrix with $D_{i,i}=w_i$. With some rather tedious algebra, it can be shown that $\beta = P\theta$, where P is an $n\times n$ matrix depending only on the t_i's. Now the criterion function becomes

$$\theta' D\theta - 2\theta' Dy + \lambda\theta' P\theta = \theta'(D+\lambda P)\theta - 2\theta' Dy.$$

Adding the shape constraints, this becomes a quadratic programming problem:

$$\text{minimize}\ \tfrac{1}{2}\theta' Q\theta - y_w'\theta\ \text{subject to}\ A\theta \geqslant 0,$$

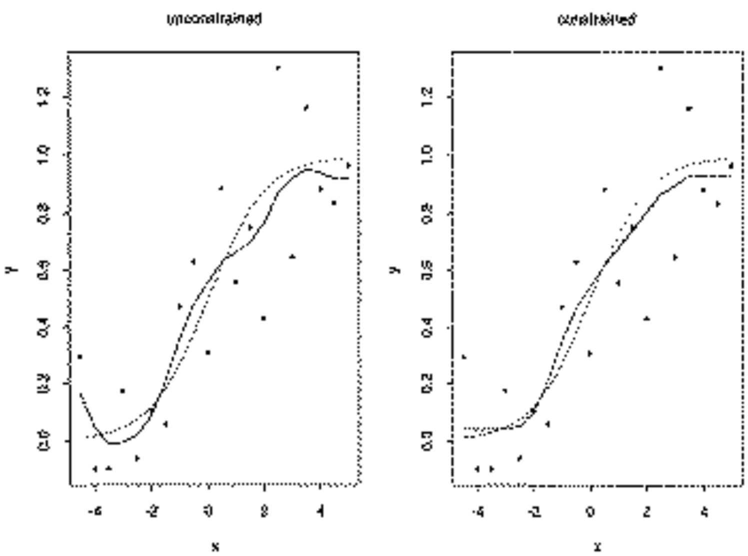
where $y_w = Dy$. With a change of variable, this version of a quadratic programming problem can be written as a sum of squares: If LL' is the Cholesky decomposition of Q, let $x = L'\theta, B = A(L')^{-1}$, and $c = (L')^{-1} y_w$. Then the problem is to find x to

$$\text{minimize}\ \|x-c\|^2\ \text{subject to}\ Bx \geqslant 0.$$

Taking the γ^j are the rows of the negative of the new constraint matrix B, these vectors are used to project the data c onto the polar cone of the constraint space, to get the projection $\hat{r}$. Then the solution is

$$\hat{\theta} = (L')^{-1}(c-\hat{r}).$$

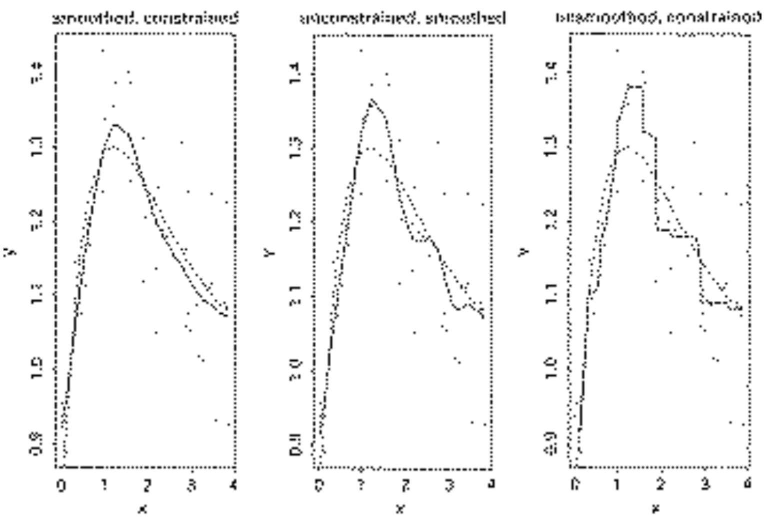
Simulated data were generated for the sets of constraints in both Example 1, Example 2. A comparison of the smooth fits to data generated from a logistic function is shown in Fig. 3. The constrained fit seems to be closer to the underlying function than the unconstrained fit, with the same smoothing parameter. Measurements of squared error loss (SEL) confirm this. SEL is the sum of the squared distances (at the data) from the estimated function to the underlying function from which the data were generated. In 100 such repetitions of simulated data, the average ratio of unconstrained SEL to constrained SEL was 1.55.



Download : Download full-size image

Fig. 3. Smooth fits to simulated data from positive increasing convex–concave function.

An example of fits to data generated using a bell-shaped function are shown in Fig. 4. Here the unconstrained smooth fit is compared with the smoothed unconstrained fit and the fit that is both smooth and constrained. Repeated simulations show that the average SEL for the smooth and constrained fit is smaller than the other two.



Download : Download full-size image

Fig. 4. Comparison of fits to simulated data from bell-shaped function.

Recommended articles   Citing articles (37)

References

Fraser and Massam (1989) D.A.S. Fraser, H. Massam
A mixed primal–dual bases algorithm for regression under inequality constraints. Application to convex regression
Scand. J. Statist, 16 (1989), pp. 65-74
View Record in Scopus Google Scholar

Robertson et al. (1988) Robertson, T., Wright, F.T., Dykstra, R.L., 1988. Order Restricted Statistical Inference. Wiley, New York.
Google Scholar

Rockafellar (1970) Rockafellar, R.T., 1970. Convex Analysis. Princeton University Press, New Jersey.
Google Scholar

Silverman (1985) B.W. Silverman

Some aspects of the spline smoothing approach to nonparametric regression curve fitting
J. Roy. Statist. Soc. B., 47 (1) (1985), pp. 1-52
View Record in Scopus Google Scholar

Wahba (1990) Wahba, G., 1990. Spline Models for Observational Data, CMBS-NSF Regional Conference Series in Applied Mathematics.
Google Scholar

Wegman and Wright (1983) E.J. Wegman, I.W. Wright
Splines in statistics
J. Amer. Statist. Assoc., 78 (382) (1983), pp. 351-365
CrossRef View Record in Scopus Google Scholar

View Abstract