An Algorithm for Restricted Least Squares Regression

Richard L. Dykstra

Journal of the American Statistical Association, Volume 78, Issue 384 (Dec., 1983), 837-842.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28198312%2978%3A384%3C837%3AAAFRLS%3E2.0.CO%3B2-D

An Algorithm for Restricted Least Squares Regression

RICHARD L. DYKSTRA*

A commonly occurring problem in statistics is that of minimizing a least squares expression subject to side constraints. Here a simple iterative algorithm is presented and shown to converge to the desired solution. Several examples are presented, including finding the closest concave (convex) function to a set of points and other general quadratic programming problems. The dual problem to the basic problem is also discussed and a solution for it is given in terms of the algorithm. Finally, extensions to expressions other than least squares are given.

KEY WORDS: Regression; Projections; Convex cones; Dual convex cones; Least squares; Concave (convex) functions; Mahalanobis distance; Linear constraints; Restricted maximum likelihood.

1. INTRODUCTION

Many problems involve minimizing a sum of squares expression subject to the restriction that the solution must satisfy certain side conditions. We refer the reader to Barlow et al. (1972) for numerous examples and an overview of a particular subset of this type of problem.

Some problems of this type have elegant closed-form solutions as in Barlow and Brunk (1972), Shaked (1979), and Dykstra and Robertson (1982), while others require extensive numerical work to obtain solutions.

We propose here a simple iterative technique that will apply to a wide subset of this type of problem. This method is not based on search techniques and complicated branching logic and hence is generally easy to program and use on modern high-speed computers. We will state our results in terms of closed convex cones and projections and then give several examples to illustrate the application of the proposed technique.

2. NOTATION AND BACKGROUND MATERIAL

Let R^n denote $R \times R \times \cdots \times R$ (n copies of the real line) and let g and w ($w_i > 0$) be fixed points in R^n. We denote the inner product of x and y (with respect to w) as

$$(x, y) = \sum_{i=1}^{n} x_i y_i w_i.$$

* Richard L. Dykstra is Professor of Statistics, Department of Statistics and Actuarial Science, University of Iowa, Iowa City, IA 52242. This work was supported partially by ONR Contract N00014-78-C-0655. The author would like to thank Tim Robertson for helpful discussions on this topic, and the referee for his help.

The inner product norm of x is

$$\| x \| = (x, x)^{1/2} = \left(\sum_1^n x_i^2 w_i \right)^{1/2}. \qquad (2.1)$$

Of course, $d(x, y) = \| x - y \|$ defines a metric on R^n. We assume that K is a closed (in the metric) convex cone in R^n. That is, $x, y \in K$; $a, b \geq 0$ implies $ax + by \in K$.

The dual cone of K is given by

$$K^* = \left\{ y; (y, x) = \sum_1^n y_i x_i w_i \leq 0 \quad \text{for all} \quad x \in K \right\}. \qquad (2.2)$$

Of course K^* is also a closed convex cone with the property that $K^{**} = K$. A commonly occurring problem is

$$\underset{x \in K}{\text{Minimize}} \| g - x \|. \qquad (2.3)$$

A vector $g^* \in K$ achieves the minimal value in (2.3) if

$$\text{(i)} \ (g - g^*, g^*) = \sum_1^n (g_i - g_i^*) g_i^* w_i = 0, \quad \text{and}$$

$$\text{(ii)} \ (g - g^*, f) = \sum_1^n (g_i - g_i^*) f_i w_i \leq 0 \quad \text{for all} \quad f \in K. \qquad (2.4)$$

(See Theorem 7.8 of Barlow et al. 1972.)

Barlow and Brunk (1972) point out that if g^* solves (2.3), $g - g^*$ solves the dual problem

$$\underset{x \in K^*}{\text{Minimize}} \| g - x \|. \qquad (2.5)$$

3. THE GENERAL ALGORITHM

Many important problems are of the form

$$\underset{x \in \cap_1^r K_i}{\text{Minimize}} \| g - x \|, \qquad (3.1)$$

where $K_1, K_2, \ldots, K_r$ are closed convex cones. We assume that we can solve the problem (find the vector in K_i that will)

$$\underset{x \in K_i}{\text{Minimize}} \| f - x \| \qquad (3.2)$$

for any f and any i, and we wish to use this knowledge to solve the more involved problem (3.1). We now give a simple, iterative procedure based only on (3.2), which will enable us to solve (3.1). We place no restrictions on the K_i other than that they be closed, convex cones, and that r be finite.

We note that since $K_1, \ldots, K_r$ are closed convex cones, so is $K_1 \cap K_2 \cap \cdots \cap K_r$, and hence the solution to (3.1) can be characterized by (2.4). If the K_i are sufficiently nice (say finitely generated), the direct sum

$$K_1 + K_2 + \cdots + K_r = \{x_1 + x_2 + \cdots + x_r; x_i \in K_i, i = 1, \ldots, r\}$$

is also a closed, convex cone. However, in general the closure property is not guaranteed (see Hestenes 1975, pp. 196–198). Nevertheless, intersections and direct sums of closed, convex cones are closely related because

$$(K_1 + \cdots + K_r)^* = K_1^* \cap K_2^* \cap \cdots \cap K_r^* \quad (3.3)$$

is always true, and

$$(K_1 \cap K_2 \cap \cdots \cap K_r)^* = K_1^* + \cdots + K_r^* \quad (3.4)$$

if the latter cone is closed. This is guaranteed if the relative interiors of the K_i have a point in common (see Rockafeller 1970, p. 146) or, as we said, if the K_i^* are finitely generated.

Our procedure can be concisely expressed with the aid of the following notation: (a) Let $P(f \mid K_i)$ (the projection of f onto K_i) denote the vector that solves (3.2). (b) For any positive integer n, we define $n \bmod r = i$ if $n = kr + i$ for integers k and i where $1 \le i \le r$. (c) Initially, set $g_0 = g$, $I_i \equiv 0$, $i = 1, \ldots, r$, and $n = 1$. The procedure is to

(i) Set $g_n = P(g_{n-1} - I_{n \bmod r} \mid K_{n \bmod r})$, and then update $I_{n \bmod r}$ by letting it now be $g_n - (g_{n-1} - I_{n \bmod r})$.

(ii) Replace n by $n + 1$ and go to (i). (3.5)

We note that if $K_1, \ldots, K_r$ are actually subspaces, then $P(\cdot \mid K_i)$ is a linear operator. Moreover, the updated $-I_{n \bmod r}$ is the projection of $g_{n-1} - I_{n \bmod r}$ onto $K_{n \bmod r}^*$, and hence

$$P(I_{n \bmod r} \mid K_{n \bmod r}) \equiv 0. \quad (3.6)$$

Thus, for subspaces, our procedure reduces to exactly the cyclic, iterated projections first discussed by von Neumann (1950). Von Neumann established the convergence of this procedure in the more general setting of subspaces of Hilbert space. This result was also proven independently by Wiener (1955), who used it in the area of prediction theory for multivariate stochastic processes (see Masani and Wiener 1958).

However, if the K_i are not subspaces, $P(\cdot \mid K_i)$ is not a linear operator, nor does (3.6) necessarily hold.

For the purpose of clarification, and to introduce notation needed for the proof of Theorem 3.1, we now restate the algorithm given in (3.5) in a step-by-step fashion emphasizing the role played by the increments.

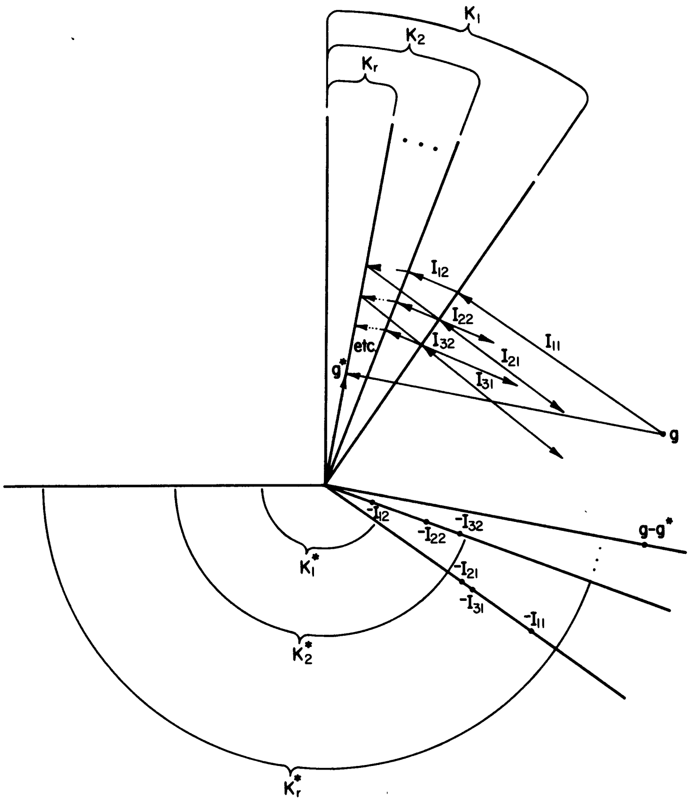

Figure 1. Schematic Drawing of Suggested Algorithm.

1. Let $g_{1,1}$ denote the projection of g onto the cone K_1. We let $I_{1,1} = g_{1,1} - g$ denote the incremental change incurred by the projection, so that $g_{1,1} = g + I_{1,1}$.
2. Let $g_{1,2}$ denote the projection of $g + I_{1,1}$ onto K_2. The incremental change is $I_{1,2} = g_{1,2} - (g + I_{1,1})$ so that $g_{1,2} = g + I_{1,1} + I_{1,2}$.
3. Let $g_{1,3}$ be the projection of $g + I_{1,1} + I_{1,2}$ onto K_3. The incremental change is $I_{1,3} = g_{1,3} - (g + I_{1,1} + I_{1,2})$ so that $g_{1,3} = g + I_{1,1} + I_{1,2} + I_{1,3}$.
4. Continue. After $g_{1,r}$ and $I_{1,r} = g_{1,r} - (g + I_{1,1} + \cdots + I_{1,r-1})$ are found, we let $g_{2,1}$ denote the projection of $g + I_{1,2} + \cdots + I_{1,r}$ onto K_1. Note that we have removed the increment $I_{1,1}$ before the projection. Our new increment is $I_{2,1} = g_{2,1} - (g + I_{1,2} + \cdots + I_{1,r})$, so that $g_{2,1} = g + I_{2,1} + I_{1,2} + \cdots + I_{1,r}$.
5. Continue.

Note that as we cyclically project onto the cones, $g_{n,i}$ is the projection onto the ith cone during the nth cycle. Note that the last increment for that cone is removed prior to that projection and that a new increment is always formed (see Figure 1). Thus, in general, $g_{n,i}$ is the projection of

$$g + I_{n,1} + \cdots + I_{n,i-1} + I_{n-1,i+1} + \cdots + I_{n-1,r}$$

onto K_i and

$$I_{n,i} = g_{n,i} - (g + I_{n,1} + \cdots + I_{n,i-1} + I_{n-1,i+1} + \cdots + I_{n-1,r}).$$

The utility of the algorithm is based on the following theorem.

Theorem 3.1. The vectors $g_{n,j}$ converge to the true solution of (3.1), say g^* as $n \to \infty$ for $j = 1, \ldots, r$. (Equivalently, $g_n \to g^*$ as $n \to \infty$ for the g_n defined in (3.5).) The proof is deferred to the Appendix.

By Barlow and Brunk's (1972) duality results, we can also use our procedure to solve the dual problem

$$\underset{x \in K_1 + K_2 + \cdots + K_r}{\text{Minimize}} \ \| g - x \|, \tag{3.7}$$

where $K_1, K_2, \ldots, K_r$ are closed convex cones and $K_1 + \cdots + K_r$ is also closed. We do this by first using the algorithm to solve the dual problem (see (3.3))

$$\underset{x \in \cap_1^r K_i^*}{\text{Minimize}} \ \| g - x \|. \tag{3.8}$$

Note that if we can project onto K_i, we can also project onto K_i^* since $P(f \mid K_i^*) = f - P(f \mid K_i)$. Thus if $\tilde{g}$ denotes the solution to (3.8), the solution to (3.7) is given by $g - \tilde{g}$.

The usefulness of the algorithm is greatly enhanced by the fact that solutions to least squares problems solve many other types of optimization problems.

For example, if $\Phi: R \to R$ is an appropriate convex function with derivative ϕ such that $x \in \cap_1^r K_i$ implies

$$(\phi(x_1), \ldots, \phi(x_n)) \in \bigcap_1^r K_i,$$

then g^* solves the problem

$$\underset{x \in \cap_1^r K_i}{\text{Maximize}} \sum_{i=1}^n \{\Phi(x_i) + (g_i - x_i)\phi(x_i)\} w_i \tag{3.9}$$

(see Theorem 1.10 of Barlow et al.).

Along somewhat similar lines, Theorem 3.1 of Barlow and Brunk (1972) guarantees that the problem

$$\underset{x \in \cap_1^r K_i}{\text{Minimize}} \sum_{i=1}^n (\Phi(x_i) - g_i x_i) w_i \tag{3.10}$$

is solved by $(\phi^{-1}(g_i^*))$ providing Φ is an appropriate convex function with derivative ϕ and $(\phi^{-1}(g_1^*), \ldots, \phi^{-1}(g_n^*)) \in \cap_1^r K_i$.

Many constrained maximum likelihood problems can be handled by appropriate use of (3.9) and (3.10).

4. EXAMPLES AND APPLICATIONS

4.1 Concave Restrictions

A problem that has received considerable attention (see Hildreth 1954, Hanson and Pledger 1976, and Wu 1982) is that of finding the closest (in the sense of least squares) concave (convex) function to a set n dependent variable values $g_1, \ldots, g_n$ corresponding to the specified independent variable values $y_1 < y_2 < \cdots < y_n$. To my knowledge, no closed-form solution to this problem exists.

Thus we want to minimize $\sum_1^n (g_i - x_i)^2 w_i$ subject to the concavity restrictions; that is,

$$\frac{x_{i+2} - x_{i+1}}{y_{i+2} - y_{i+1}} \le \frac{x_{i+1} - x_i}{y_{i+1} - y_i}, \quad i = 1, 2, \ldots, n - 2. \tag{4.1}$$

If we define K_i to be those vectors that satisfy (4.1) for a fixed i, ($i = 1, \ldots, n - 2$), then K_i is a closed convex cone in R^n, and our goal is to

$$\underset{x \in \cap_1^{n-2} K_i}{\text{Minimize}} \sum_1^n (g_i - x_i)^2 w_i.$$

The key point is that projection of any g onto K_i is very easy since it only involves three points, while projection onto $\cap_1^{n-2} K_i$ is very difficult. It is easily shown that if g satisfies (4.1) for the fixed i, the projection of g onto K_i is just g. If g does not satisfy (4.1) for this i, the projection onto K_i is the appropriate straight line fitting the points g_i, g_{i+1}, and g_{i+2}, and just g_j elsewhere. Explicitly, in this case

$$\begin{aligned} P(g \mid K_i)_j &= \bar{g}_i + \hat{\beta}_i(y_j - \bar{y}_i), & j &= i, i + 1, i + 2 \\ &= g_j, & j &\neq i, i + 1, i + 2, \end{aligned}$$

where

$$\bar{g}_i = \sum_{j=i}^{i+2} g_j w_j \Big/ \sum_{j=i}^{i+2} w_j,$$

$$\bar{y}_i = \sum_{j=i}^{i+2} y_j w_j \Big/ \sum_{j=i}^{i+2} w_j,$$

and

$$\hat{\beta}_i = \frac{\sum_{j=i}^{i+2} (g_j - \bar{g}_i) y_j w_j}{\sum_{j=i}^{i+2} (y_j - \bar{y}_i)^2 w_j}.$$

This projection onto K_i is of course very easy to program on a computer. We then need only proceed according to the algorithm. The convergence to the solution is quite rapid using the computer since each projection is so simple. Even for large values of n, the solution is obtained very quickly.

4.2 Linear Restrictions

A more general quadratic programming problem than that considered in Example 4.1) can be handled in essentially the same way. Suppose we wish to

$$\text{Minimize} \sum_1^n (g_i - x_i)^2 w_i$$

subject to

$$\sum_{j=1}^n a_{ij} x_j \le 0 \quad \text{for} \quad i = 1, \ldots, r.$$

If we let $K_i = \{x; \sum_{j=1}^n a_{ij} x_j \le 0\}$, the problem is precisely of the form discussed in Section 3. The projection of g

onto K_i is very easy since it involves only the ith constraint. A closed form for the projection that can easily be programmed is

$$P(g \mid K_i) = g, \quad \text{if } \sum_{j=1}^{n} a_{ij} g_j \le 0$$

$$= (g_1', \ldots, g_n'), \quad \text{if } \sum_{j=1}^{n} a_{ij} g_j > 0,$$

where

$$g_j' = g_j - \left(\sum_{l=1}^{n} g_l a_{il}\right) a_{ij} w_j^{-1} \Big/ \sum_{l=1}^{n} a_{il}^2 w_l^{-1}.$$

We then proceed according to our algorithm. Note that no derivatives need be found and that the only checking procedure we need do is ascertain whether $\sum_{j=1}^{n} a_{ij} h_j \le 0$ for appropriate h.

4.3 Mahalanobis Distance

The problem of isotonic regression has received much attention over the years (see Brunk 1955,1965, and Barlow et al. 1972). Basically the problem is to

$$\underset{x \in K}{\text{Minimize}} \left\{ \| g - x \|^2 = \sum_{1}^{n} (g_i - x_i)^2 w_i \right\},$$

where K consists of those vectors that are isotonic (order preserving) with respect to a particular partial ordering. Elegant closed-form solutions have been found for this problem (see Brunk 1955 and Barlow et al. 1972). This is of course a special case of the problem considered in 4.2. A natural extension of the isotonic regression problem would be to use a Mahalanobis distance concept in our objective function rather than a weighted sum of squares. Thus our new problem would be expressible as

$$\text{Minimize } (g - x)' F (g - x)$$

subject to constraints of the form $\sum_1^n a_{ij} x_j = a_i' x \le 0$, $i = 1, \ldots, r$, where F is a given, symmetric, positive definitive matrix.

However, in this case there exists an orthogonal matrix P such that $F = P' \Lambda P$, where Λ is the diagonal matrix consisting of the eigenvalues of F. In this case, the problem can be rephrased as

$$\text{Minimize} \sum_{1}^{n} ((Pg)_i - y_i)^2 \lambda_i$$

subject to

$$((Pa_i)' y) \le 0, \quad i = 1, \ldots, r,$$

where $y = Px$. We can now use the algorithm to solve for y, and then obtain the desired solution as $x = P'y$.

4.4 Binomial Parameters

As a final example, let us consider the case of finding MLE's for the parameters p_{ij}, $i = 1, \ldots, a; j = 1, \ldots, b$ subject to

$$p_{ij} \le p_{i,j+1} \quad \text{for all } i, j$$

and

$$p_{ij} \le p_{i+1,j} \quad \text{for all } i, j, \tag{4.2}$$

where x_{ij} are independent binomial (n_{ij}, p_{ij}) random variables.

The likelihood function to be maximized is of course

$$L(p) = \prod_{j=1}^{b} \prod_{i=1}^{a} \binom{n_{ij}}{x_{ij}} p_{ij}^{x_{ij}} (1 - p_{ij})^{n_{ij} - x_{ij}}, \; 0 \le p_{ij} \le 1.$$

This is equivalent to minimizing

$$\sum_{j=1}^{b} \sum_{i=1}^{a} (x_{ij}/n_{ij} - p_{ij})^2 n_{ij}, \; 0 \le p_{ij} \le 1$$

subject to the constraints (4.2) (see Barlow and Brunk 1972 for a verification of this). If we depict our constraints as requiring that our estimates fall in the following cones:

$$K_1 = \{(p_{ij});\, p_{ij} \le p_{i,j+1}, \quad j = 1, \ldots, b - 1;\, i = 1, \ldots, a\}$$

and

$$K_2 = \{(p_{ij});\, p_{ij} \le p_{i+1,j}, \quad i = 1, \ldots, a - 1;\, j = 1, \ldots, b\},$$

then the problem fits into the framework described for our algorithm. (We note that since $0 \le x_{ij}/n_{ij} \le 1$ for all i, j, the solution obtained by our algorithm will automatically fall between 0 and 1.) Moreover, since projection onto K_1 (or K_2) is just a one-dimensional smoothing, it is well known how to accomplish these projections (see Barlow et al. 1972). This seems to provide an efficient method of obtaining these constrained MLE's even for large values of a and b.

APPENDIX

Lemma A.1. Suppose a sequence of real numbers $\{a_n\}_{n=1}^{\infty}$ is such that $\sum_1^\infty a_n^2 = M < \infty$. Then there exists a subsequent $\{a_{n_j}\}_{j=1}^{\infty}$ such that

$$\sum_{m=1}^{n_j} | a_m | \, | a_{n_j} | \to 0 \quad \text{as} \quad j \to \infty.$$

Proof. The result is clearly true if $\{a_n\}_{n=1}^{\infty}$ contains an infinite number of zeros, so assume otherwise.

Let n_j be chosen such that

$$| a_{n_j} | = \min\{ | a_m |;\, m \le n_j,\, | a_m | > 0\}.$$

Clearly such a sequence exists since $| a_n | \to 0$. Now, for a given $\epsilon > 0$, choose $n_{j'}$ such that

$$\sum_{1}^{n_{j'}} a_n^2 > M - \epsilon/2.$$

Choose $n_{j''} \ge n_{j'}$ such that

$$\frac{|a_{n_{j''}}|}{|a_{n_{j'}}|} \le \epsilon/2M.$$

Then, for $n_j \ge n_{j''}$,

$$\sum_{m=1}^{n_j} |a_m||a_{n_j}| = \sum_{m=1}^{n_{j'}} |a_m| \frac{|a_{n_j}|}{|a_{n_{j'}}|} |a_{n_{j'}}| + \sum_{m=n_{j'}+1}^{n_j} |a_m||a_{n_j}| \le \sum_{m=1}^{n_{j'}} a_m^2 \cdot \frac{|a_{n_j}|}{|a_{n_{j'}}|} + \sum_{m=n_{j'}+1}^{n_j} a_m^2 < M \cdot \frac{\epsilon}{2M} + \frac{\epsilon}{2} = \epsilon,$$

which concludes the proof.

Proof of Theorem 3.1. First note that the key relationships

$$g_{n,i-1} - g_{n,i} = I_{n-1,i} - I_{n,i}, \quad i = 2, \ldots, r,$$

and

$$g_{n-1,r} - g_{n,1} = I_{n-1,1} - I_{n,1} \tag{A.1}$$

hold among the projections and increments. Since $\cap_{i=1}^r K_i$ is a closed convex cone (nonempty since it contains the origin), the unique true projection of g onto $\cap_1^r K_i$, say g^*, exists. Then from (A.1), we may write ($I_{0,i} \equiv 0$)

$$\begin{aligned}\| g_{n,i-1} - g^* \|^2 &= \| (g_{n,i} - g^*) + (I_{n-1,i} - I_{n,i}) \|^2 \\ &= \| g_{n,i} - g^* \|^2 + \| I_{n-1,i} - I_{n,i} \|^2 \\ &\quad + 2(g^*, I_{n,i} - I_{n-1,i}) + 2(g_{n,i}, I_{n-1,i} - I_{n,i})\end{aligned}$$

for $i \ge 2$. Note that the last term is nonnegative since $(g_{n,i}, I_{n,i}) = 0$ by the properties of projections onto convex cones, and $(g_{n,i}, I_{n-1,i}) \ge 0$ since $g_{n,i} \in K_i$ and $-I_{n-1,i} \in K_i^*$.

In similar fashion,

$$\| g_{n-1,r} - g^* \|^2 \ge \| g_{n,1} - g^* \|^2 + \| I_{n-1,1} - I_{n,1} \|^2 + 2(g^*, I_{n,1} - I_{n-1,1}).$$

Noting the "telescoping property" of the term $(g^*, I_{n,i} - I_{n-1,i})$, we may write

$$\| g - g^* \|^2 \ge \| g_{n,r} - g^* \|^2 + \sum_{k=1}^{n} \sum_{l=1}^{r} \| I_{k-1,l} - I_{k,l} \|^2 + 2(g^*, I_{n,1} + \cdots + I_{n,r}) \tag{A.2}$$

for every n. Since $g^* \in \cap_1^r K_i$ and $-I_{n,i} \in K_i^*$, the last term is nonnegative and

$$\sum_{k=1}^{\infty} \sum_{l=1}^{r} \| I_{k-1,l} - I_{k,l} \|^2 < \infty. \tag{A.3}$$

Thus

$$\| I_{n-1,l} - I_{n,l} \| = \| g_{n,l-1} - g_{n,l} \| \; (l \ge 2)$$

and (A.4)

$$\| I_{n-1,1} - I_{n,1} \| = \| g_{n-1,r} - g_{n,1} \| \to 0 \quad \text{as } n \to \infty.$$

Note that (A.2) implies that $g_{n,r}$ and $g - g_{n,r} = I_{n,1} + I_{n,2} + \cdots + I_{n,r}$ are uniformly bounded. However, we cannot guarantee that the $I_{n,i}$ are uniformly bounded and this complicates the proof.

From (A.4), it will suffice to show $g_{n,r} \to g^*$ as $n \to \infty$. First we show that there exists a subsequence such that

(i) $(I_{n_j,1} + \cdots + I_{n_j,r}, g^*) \to 0,$

and (A.5)

(ii) $g_{n_j,r} \to g^*$ as $j \to \infty$.

Now note that

$$\begin{aligned}|(I_{n,1} + \cdots + I_{n,r}, g_{n,1})| &= \left| \sum_{i=1}^{r} (I_{n,i}, g_{n,1} - g_{n,i}) \right| \quad (\text{since } (I_{n,i}, g_{n,i}) = 0) \\ &\le \sum_{i=2}^{r} \| I_{n,i} \| \| g_{n,1} - g_{n,i} \| \\ &\le \sum_{i=2}^{r} \sum_{m=1}^{n} \| I_{m,i} - I_{m-1,i} \| a_n \\ &= \sum_{m=1}^{n} \sum_{i=2}^{r} \| I_{m,i} - I_{m-1,i} \| a_n = \sum_{m=1}^{n} a_m a_n,\end{aligned}$$

where

$$\begin{aligned}a_n &= \| g_{n,1} - g_{n,2} \| + \| g_{n,2} - g_{n,3} \| + \cdots + \| g_{n,r-1} - g_{n,r} \| \\ &= \| I_{n,2} - I_{n-1,2} \| + \| I_{n,3} - I_{n-1,3} \| + \cdots + \| I_{n,r} - I_{n-1,r} \|.\end{aligned}$$

Note that $a_n^2 \le 2^{r-2} \sum_{l=2}^r \| I_{n,l} - I_{n-1,l} \|^2$. Thus from (A.2),

$$\sum_{1}^{\infty} a_n^2 \le 2^{r-2} \| g - g^* \|^2 < \infty.$$

While this does not imply $\sum_1^n a_m a_n \to 0$ unless the a_n are nonincreasing, it does imply (see Lemma A.1) that there exists a subsequence such that

$$\sum_{m=1}^{n_j} a_m a_{n_j} \to 0 \quad \text{as } j \to \infty. \tag{A.6}$$

Moreover, since the $g_{n,r}$ are uniformly bounded, we may assume that we have chosen a subsequence such that (A.6) holds and $g_{n_j,r}$ converges, say to h. Of course, by (A.4), $g_{n_j,1}$ also converges to h, and hence we may write

$$(g - h, h) = \lim_{j \to \infty} - (I_{n_j,1} + I_{n_j,2} + \cdots + I_{n_j,r}, g_{n_j,1}) = 0. \tag{A.7}$$

Note also that (A.4) implies that $h \in \cap_1^r K_i$ since $g_{n_j,i}$, which is in K_i, becomes arbitrarily chose to h and K_i is closed. Also, for $f \in \cap_1^r K_i$,

$$(g - h, f) = \lim_{j\to\infty} - (I_{n_j,1} + \cdots + I_{n_j,r}, f) \le 0$$

since $I_{n_j,i} \in -K_i^*$. Thus $h = g^*$ by (2.4).

Finally, noting that $I_{n_j,1} + \cdots + I_{n_j,r}$ is uniformly bounded and that $g_{n_j,1} \to g^*$, (A.7) will imply (A.5 (i)).

Now we need only show that $g_{n,r} \to g^*$. However, for $n > n_j$, we have (in a manner similar to (A.2)),

$$\| g_{n_j,r} - g^* \| \ge \| g_{n,r} - g^* \|^2 + \sum_{m=n_j+1}^{n} \sum_{l=1}^{r} \| I_{m,l} - I_{m-1,l} \|^2 + 2(g^*, I_{n,1} + \cdots + I_{n,r}) - 2(g^*, I_{n_j,1} + \cdots + I_{n_j,r}).$$

Then, since the last term can be made arbitrarily small by (A.5 (i)), the next to the last term is nonnegative, and the left side goes to zero as $j \to \infty$, it must follow that

$$\lim_{n\to\infty} \| g_{n,r} - g^* \| = 0.$$

[*Received April 1982. Revised February 1983.*]

REFERENCES

BARLOW, R.E., and BRUNK, H.D. (1972), "The Isotonic Regression Problem and Its Dual," *Journal of the American Statistical Association*, 67, 140–147.

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M., and BRUNK, H.D. (1972), *Statistical Inference Under Order Restrictions*, New York: John Wiley.

BRUNK, H.D. (1955), "Maximum Likelihood Estimates of Monotone Parameters," *Annals of Mathematical Statistics*, 26, 607–616.

——— (1965), "Conditional Expectation Given a σ-Lattice and Applications," *Annals of Mathematical Statistics*, 36, 1339–1350.

DYKSTRA, RICHARD L., and ROBERTSON, TIM (1982), "Order Restricted Statistical Tests on Multinomial and Poisson Parameters: The Starshaped Restriction," *Annals of Statistics*, 10, 1246–1252.

HANSON, D.L., and PLEDGER, G. (1976), "Consistency in Concave Regression," *Annals of Statistics*, 4, 1038–1050.

HESTENES, MAGNUS R. (1975), *Optimization Theory*, New York: John Wiley.

HILDRETH, C. (1954), "Point Estimates of Ordinates of Concave Functions," *Journal of the American Statistical Association*, 49, 598–619.

MASANI, P., and WIENER, N. (1958), "The Prediction Theory of Multivariate Stochastic Processes II," *Acta Mathematica*, 99, 93–137.

ROCKAFELLER, R. TYRRELL (1970), *Convex Analysis*, Princeton, N.J.: Princeton University Press.

SHAKED, MOSHE (1979), "Estimation of Starshaped Sequences of Poisson and Normal Means," *Annals of Statistics*, 7, 729–741.

VON NEUMANN, JOHN (1950), *Functional Operators* (Vol. II), Princeton, N.J.: Princeton University Press.

WIENER, N. (1955), "On Factorization of Matrices," *Comment Mathematici Helvetici*, 29, 97–111.

WU, CHIEN-FU (1982), "Some Algorithms for Concave Regression and Isotonic Regression," *Management Science* (forthcoming).