The Annals of Statistics
1982, Vol. 10, No. 3, 708–716

AN ALGORITHM FOR ISOTONIC REGRESSION FOR TWO OR MORE INDEPENDENT VARIABLES[1]

BY RICHARD L. DYKSTRA AND TIM ROBERTSON

University of Missouri, Columbia and University of Iowa

Algorithms for solving the isotonic regression problem in more than one dimension are difficult to implement because of the large number of lower sets present or because they involve search techniques which require a significant amount of checking and readjustment. Here a new algorithm for solving this problem based on a simple iterative technique is proposed and shown to converge to the correct solution.

1. Introduction. Algorithms for calculating the least squares isotonic regression function have received a great deal of attention in the literature and six such algorithms are discussed in Section 2.3 of Barlow, Bartholomew, Bremner and Brunk (1972). In situations where there is one independent variable all of the algorithms work very efficiently. Perhaps the most widely used algorithm is the "pool adjacent violators algorithm" which is applicable only in the case of a simple linear ordering or an amalgamation of simple orderings.

In many isotonic regression problems we have more than one independent variable present and are concerned with partial orderings. An important example involves the prediction of success in college. Usually, this prediction is based upon several independent variables such as rank in high school graduating class and score on a standardized examination such as the ACT composite and is measured in terms of a predicted grade point average or predicted probability of obtaining a particular GPA or better. The predicted value is usually obtained by regression methods and is assumed to be nondecreasing in each independent variable. The isotonic regression function has been found to compare very favorably with other techniques with respect to predictive accuracy; cf. Perrin and Whitney (1976) and Kolen and Whitney (1978).

Some of the algorithms described in Barlow et al. are applicable to the case of computing the doubly nondecreasing least squares regression function but the number of computations required can become prohibitive. For example, consider the minimum lower sets algorithm described in Section 2.3 of Barlow et al. Suppose one of our two independent variables has a possible values and the other has b possible values. By counting paths from the upper left hand corner to the lower right hand corner of our $a \times b$ grid, it follows that the number of lower sets is equal to $\binom{a+b}{a}$ (including the empty set). If $a = b$ this number is approximately $(a\pi)^{-1/2} \cdot 4^a$ by Stirling's formula. Thus if $a = b = 20$, and if consideration of each lower set were to require one microsecond of computer time, then finding the first level set would require 2312 minutes or 38.5 hours of CPU time. (One microsecond seems conservative in light of the fact that computation of the average value over that set would take at least two multiplications, two additions and a division and the comparison would require a subtraction. The present standard for making such predictions is four arithmetic operations per microsecond.) Moreover, if the first level set is small (as it would be with good data) the second cycle is nearly as difficult as the first.

The algorithm proposed by Gebhardt (1970), one given by Dykstra (1981), and the minimax order algorithm (cf. Chapter 2 of Barlow et al., op. cit.) all involve search

Received April 1981; revised January 1982.

[1] This work was partially supported by ONR Contracts N00014-78-C-0655 and N00014-80-C-0321.

AMS subject classifications. Primary 65D15, 49D05.

Key words and phrases. Isotone regression, minimal lower sets algorithm, convex cones, dual convex cones, projections.

techniques which can require a significant amount of checking and readjustment. Thus computer programs which implement these algorithms require intricate branching logic and are complicated to program. On the other hand, these algorithms provide exact solutions rather than converge to the correct solution and should be easier to use than the one proposed in this paper for small data sets or if one does not have access to a modern computing facility.

Since the doubly nondecreasing regression function is so difficult to compute, researchers have proposed using ad hoc estimators based upon one dimensional smoothings. (The number of computations required for one dimensional smoothings is essentially linear in the number of entries.) Makowski (1974) studied consistency properties of estimators obtained by successive one dimensional smoothings. Kolen, Smith and Whitney (1977), Perrin and Whitney (1976), and Kolen and Whitney (1978) proposed two different techniques for producing estimates which are nondecreasing in each variable. One of their techniques was first to do one dimensional row smoothings. After all rows had been adjusted, reversals in the columns were adjusted by the same method. They then returned to the original table and did one dimensional column smoothings followed by row smoothings. Neither smoothing necessarily produces a doubly nondecreasing table so they averaged the two results. (The average is not necessarily doubly nondecreasing, but it was for their data.) This method was applied to the problem of estimating the probability of obtaining a "B or better" GPA for entering college students. This data can be found in Table 1. The two entries are the total cell frequencies and the observed relative frequencies. We note that there are a number of "reversals," even with a relatively large sample size. The smoothed estimates, by the above method, are presented in Table 2 and the isotonic regression function with weights equal to frequencies in Table 3. Note that not only the estimates but also the level sets are different. These level sets are very useful for making inferences about equivalent scores within the table.

In this paper we present an algorithm for calculating the least squares isotonic regression function which is increasing in each of two or more variables. This algorithm uses successive one dimensional smoothings and is very efficient and easy to program. This algorithm is described in Section 3 for the case of two variables. The extension to more than two variables is discussed in Section 4. In Section 2 we summarize some well-known properties of isotonic regression for the case of two variables which will be used in the proof that the algorithm yields the desired result.

2. Some preliminaries. We let $\Omega = \{(i, j);\ i = 1, 2, \cdots, a;\ j = 1, 2, \cdots, b\}$ and define the partial order $\ll$ on Ω by $(i, j) \ll (k, \ell)$ if and only if $i - k \leq 0$ and $j - \ell \leq 0$. We denote an arbitrary real function whose domain is Ω as a matrix, i.e.,

$$G = (g_{ij}) = (g((i, j))), \qquad i = 1, 2, \cdots, a; \qquad j = 1, 2, \cdots, b.$$

We say that a function $F: \Omega \to R$ is isotonic or order preserving if $(i, j) \ll (k, \ell)$ implies $f_{ij} \leq f_{k\ell}$. This is equivalent to requiring that F be nondecreasing along both rows and columns. The least squares isotonic regression problem is to

$$\text{minimize } \textstyle\sum_{i,j}(g_{ij} - f_{ij})^2 w_{ij}$$

for F belonging to the class K of isotonic functions, where $w_{ij} > 0$ and G are given.

Since the class of isotonic functions forms a closed convex cone, it is well known (c.f. Theorem 1.4 in Barlow et al., op. cit.) that the solution to the isotonic regression problem, say G^*, is that function $G^* \in K$ satisfying

$$\textstyle\sum_{i,j}(g_{ij} - g^*_{ij})\, g^*_{ij} w_{ij} = 0, \tag{2.1}$$

and

$$\textstyle\sum_{i,j}(g_{ij} - g^*_{ij})\, h_{ij} w_{ij} \leq 0, \tag{2.2}$$

for all functions $H \in K$.

TABLE 1

The probability of making a "B or better" GPA
(top number = total cell frequency; bottom number = relative frequency)

ACT Composite	High School Grade Point Average				
	0-1.55	**1.56-2.25**	**2.26-2.95**	**2.96-3.65**	**3.66-4.00**
28+	0	7	10	47	44
	.0000	.2857	.2000	.5745	.8864
23-27	7	56	88	180	84
	.0000	.1250	.1818	.2833	.5238
18-22	23	166	152	149	33
	.0435	.0301	.0724	.1946	.1212
13-17	27	149	96	61	4
	.0000	.0470	.0313	.0492	.5000
0-12	10	57	33	7	0
	.0000	.0000	.0606	.0000	.0000

TABLE 2

The probability of making a "B or better" GPA, estimated by Kolen and Whitney Method. Same categories as Table 1.

*.0314	.2353	.2353	.5745	.8864
.0314	.1250	.1818	.2833	.5238
.0314	.0375	.0724	.1867	.1934
.0000	.0375	.0402	.0493	.1784
.0000	.0000	.0383	.0421	*.0425

TABLE 3

The probability of making a "B or better" GPA, estimated by least squares isotonic regression (weights = cell frequencies). Same categories as Table 1.

*.0333	.2353	.2353	.5745	.8864
.0333	.1250	.1818	.2833	.5238
.0333	.0377	.0724	.1881	.1881
.0000	.0377	.0377	.0492	.1881
.0000	.0000	.0377	.0377	*.0377

* Any value which satisfies the order restrictions will suffice here since this cell has zero weight.

3. The algorithm. The algorithm which we propose requires only the ability to solve the isotonic regression problem with the usual nondecreasing order (in one dimension) along rows and columns. Our algorithm is given as follows:

Step 1. Let $\hat{G}^{(1)} = (\hat{g}_{ij}^{(1)})$ denote the isotonic regression solution of $G = (g_{ij})$ over rows, i.e., $\hat{G}^{(1)}$ minimizes $\sum_{i=1}^{a} (g_{ij} - f_{ij})^2 w_{ij}$ subject to $f_{1j} \leq f_{2j} \leq \cdots \leq f_{aj}$ for $j = 1, \cdots, b$. We call $R^{(1)} = (r_{ij}^{(1)}) = (\hat{g}_{ij}^{(1)} - g_{ij})$ the first set of "row increments."

Step 2. Let $\tilde{G}^{(1)} = (\tilde{g}_{ij}^{(1)})$ denote the isotonic regression solution over columns from the initial values $G + R^{(1)}$, i.e., $\tilde{G}^{(1)}$ minimizes $\sum_{j=1}^{b} (g_{ij} + r_{ij}^{(1)} - f_{ij})^2 w_{ij}$ subject to $f_{i1} \leq f_{i2} \leq \cdots \leq f_{ib}$ for $i = 1, \cdots, a$. We call $C^{(1)} = \tilde{G}^{(1)} - (G + R^{(1)})$ the first set of "column increments." Note that $\tilde{G}^{(1)} = G + R^{(1)} + C^{(1)}$.

Step 3. At the beginning of the nth cycle, we obtain $\hat{G}^{(n)}$ by isotonizing $G + C^{(n-1)}$ over rows. The nth set of row increments is defined by $R^{(n)} = \hat{G}^{(n)} - (G + C^{(n-1)})$ so that $\hat{G}^{(n)} = G + C^{(n-1)} + R^{(n)}$. We then obtain $\tilde{G}^{(n)}$ by isotonizing $G + R^{(n)}$ over columns. The nth set of column increments is given by $C^{(n)} = \tilde{G}^{(n)} - (G + R^{(n)})$, or equivalently $\tilde{G}^{(n)} = G + R^{(n)} + C^{(n)}$.

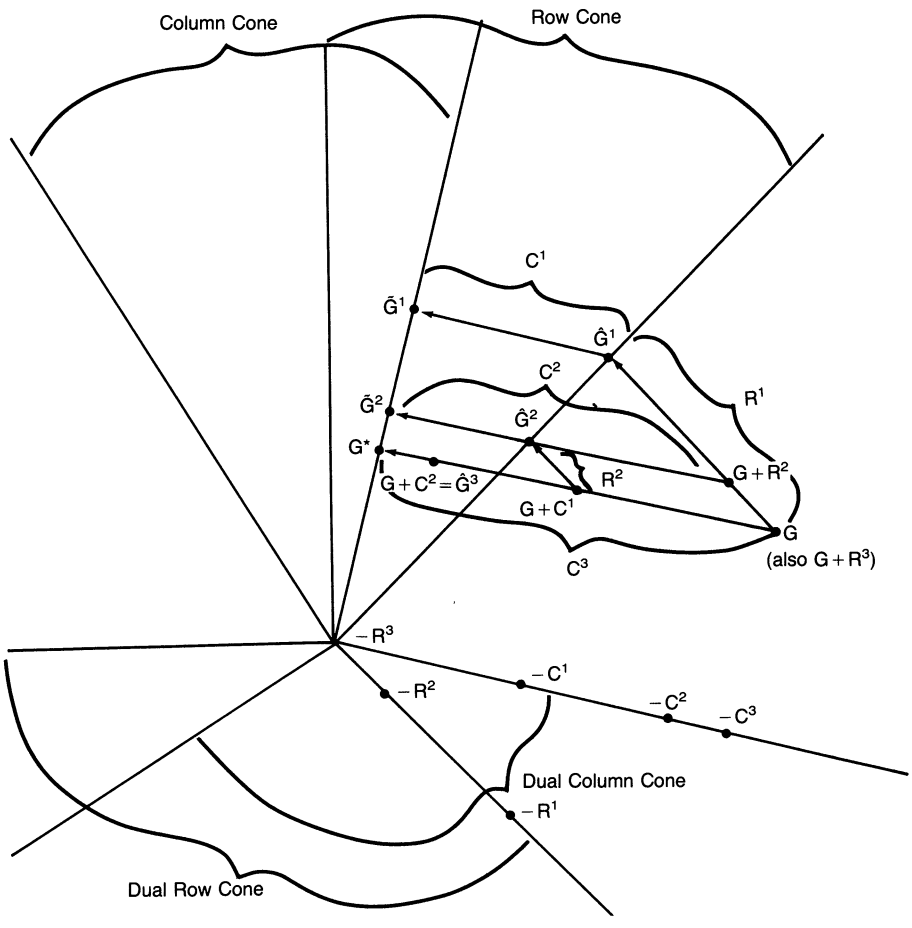

FIG. 1. *Schematic Diagram of Proposed Algorithm for Two Independent Variables.*

The utility of the algorithm lies in the following theorem.

THEOREM 3.1. *Both* $\hat{G}^{(n)}$ *and* $\tilde{G}^{(n)}$ *converge to the true solution* G^* *as* $n \to \infty$.

PROOF. If we denote the inner product norm as

$$\|F\| = (F, F)^{1/2} = \left(\sum_{i=1}^{a} \sum_{j=1}^{b} f_{ij}^2 w_{ij}\right)^{1/2},$$

we first show that

$$\|\hat{G}^{(n)}\|^2 \geq \|\tilde{G}^{(n)}\|^2 \geq \|\hat{G}^{(n+1)}\|^2 \quad \text{for all } n. \tag{3.1}$$

To establish some additional notation, we denote the "row cone" by

$$K_r = \{F; f_{1j} \leq f_{2j} \leq \cdots \leq f_{aj} \quad \text{for} \quad j = 1, \cdots, b\},$$

and the "column cone" by

$$K_c = \{F; f_{i1} \leq f_{i2} \leq \cdots \leq f_{ib} \quad \text{for} \quad i = 1, \cdots, a\}.$$

The respective dual cones, as discussed in Barlow and Brunk (1972), are

$$K_r^* = \{H; \textstyle\sum_{i=1}^{a} h_{ij} f_{ij} w_{ij} \leq 0 \quad \text{for} \quad j = 1, \cdots, b; \quad \text{for every} \quad F \in K_r\}$$

and

$$K_c^* = \{H; \textstyle\sum_{j=1}^{b} h_{ij} f_{ij} w_{ij} \leq 0 \quad \text{for} \quad i = 1, \cdots, a; \quad \text{for every} \quad F \in K_c\}.$$

Since $-R^{(n+1)}$ is the projection of $G + C^{(n)}$ onto K_r^*, the work of Barlow and Brunk guarantees that

$$-R^{(n+1)} \quad \text{minimizes} \quad \| G + C^{(n)} - F \|^2 \quad \text{for} \quad F \in K_r^*.$$

Similarly,

$$-C^{(n)} \quad \text{minimizes} \quad \| G + R^{(n)} - F \|^2 \quad \text{for} \quad F \in K_c^*.$$

Therefore, since $-R^{(n)} \in K_r^*$ and $-C^{(n-1)} \in K_c^*$,

$$\| G + R^{(n)} - (-C^{(n-1)}) \|^2 \geq \| G + R^{(n)} - (-C^{(n)}) \|^2 \geq \| G + C^{(n)} - (-R^{(n+1)}) \|^2$$

which is equivalent to (3.1).

Next we show that $\{C^{(n)}\}$ and $\{R^{(n)}\}$ are bounded. If not, let (i_0, j_0) be a minimal point in Ω (with respect to our partial ordering $\ll$) such that either $\{r_{i_0,j_0}^{(n)}\}$ or $\{c_{i_0,j_0}^{(n)}\}$ is unbounded. Say there exists a subsequence $\{n_i\}$ such that $r_{i_0,j_0}^{(n_i)} \to -\infty$. (Since $\sum_{i=1}^{i_0} r_{i,j_0}^{(n)} w_{i,j_0} \leq 0$ for all n (Barlow and Brunk, 1972), $r_{i_0,j_0}^{(n_i)} \to \infty$ would contradict the fact that (i_0, j_0) is minimal.) But this, together with $\tilde{G}^{(n)} = G + R^{(n)} + C^{(n)}$ and the fact that $\tilde{G}^{(n)}$ is bounded in norm (cf. (3.1)) implies that $c_{i_0,j_0}^{(n_i)} \to \infty$. This, in turn, contradicts the fact that (i_0, j_0) is minimal since $\sum_{j=1}^{j_0} c_{i_0,j}^{(n)} w_{i_0,j} \leq 0$ for all n.

Projections onto convex sets are distance reducing (see Theorem 7.6 of Barlow et al., op. cit.) so that

$$\begin{aligned} \| C^{(i)} - C^{(i-1)} \|^2 &= \| G + C^{(i)} - (G + C^{(i-1)}) \|^2 \geq \| R^{(i+1)} - R^{(i)} \|^2 \\ &= \| G + R^{(i+1)} - (G + R^{(i)}) \|^2 \geq \| C^{(i+1)} - C^{(i)} \|^2 \quad \text{for all } i. \end{aligned} \tag{3.2}$$

We now show that

$$\| R^{(i+1)} - R^{(i)} \|^2 \to 0, \quad \text{hence} \quad \| C^{(i+1)} - C^{(i)} \|^2 \to 0, \text{ as } i \to \infty. \tag{3.3}$$

If (3.3) were not the case, there would exist $(i_0, j_0) \in \Omega$ and $\varepsilon > 0$ such that

$$| r_{i_0,j_0}^{(i+1)} - r_{i_0,j_0}^{(i)} | > \varepsilon \quad \text{for infinitely many } i. \tag{3.4}$$

However, since $\{R^{(n)}\}$ is bounded, there exists a finite M such that

$$| r_{i_0,j_0}^{(i)} - r_{i_0,j_0}^{(j)} | < M \quad \text{for all} \quad i, j. \tag{3.5}$$

If we write

$$\begin{aligned} \| R^{(i+1)} - R^{(i)} \|^2 - \| C^{(i+1)} - C^{(i)} \|^2 &= \| G + R^{(i)} + C^{(i)} - (G + R^{(i+1)} + C^{(i+1)}) \|^2 \\ &\quad + 2(G + R^{(i)} + C^{(i)} - (G + R^{(i+1)} + C^{(i+1)}), C^{(i+1)} - C^{(i)}), \end{aligned} \tag{3.6}$$

the left side of (3.6) converges to 0 since by (3.2) both terms converge to the same quantity. The last term of the right side is nonnegative by (2.1) and (2.2). Thus

$$(R^{(i+1)} - R^{(i)}) + (C^{(i+1)} - C^{(i)}) \to 0 \quad \text{as} \quad i \to \infty. \tag{3.7}$$

In similar fashion, beginning with

$$\| C^{(i+1)} - C^{(i)} \|^2 - \| R^{(i+2)} - R^{(i+1)} \|^2,$$

we can conclude

$$(R^{(i+2)} - R^{(i+1)}) + (C^{(i+1)} - C^{(i)}) \to 0 \quad \text{as} \quad i \to \infty. \tag{3.8}$$

Subtracting (3.7) from (3.8) yields

$$(R^{(i+2)} - R^{(i+1)}) - (R^{(i+1)} - R^{(i)}) \to 0 \quad \text{as} \quad i \to \infty.$$

Thus, for a sufficiently large N_0 and fixed n_0, we can keep

$$(r_{i_0,J_0}^{(N_0+1+i)} - r_{i_0,J_0}^{(N_0+i)}), \qquad i = 1, 2, \cdots, n_0$$

arbitrarily close to

$$(r_{i_0,J_0}^{(N_0+1)} - r_{i_0,J_0}^{(N_0)}).$$

This, however, contradicts (3.4) and (3.5) both being true.

Since $\{R^{(n)}\}$ and $\{C^{(n)}\}$ are bounded, there must exist convergent subsequences. Suppose $R^{(n_i)} \to R$ and $C^{(n_i)} \to C$. Then, in light of (3.3),

$$\tilde{G}^{(n_i)} = G + R^{(n_i)} + C^{(n_i)}$$

and

$$\hat{G}^{(n_i+1)} = G + R^{(n_i+1)} + C^{(n_i)}$$

both converge to $G^* = G + R + C$ (in anticipation of this being the desired solution). Since $\hat{G}^{(n)}$ is an element of K_r and $\tilde{G}^{(n)}$ is an element of K_c for all n, we know that $G^* \in K_r \cap K_c$ (these cones are closed). Furthermore,

$$\begin{aligned}(G - G^*, G^*) &= (G + R - G^*, G^*) - (R, G^*) \\ &= \lim_{i\to\infty}(G + R^{(n_i)} - \tilde{G}^{(n_i)}, \tilde{G}^{(n_i)}) + \lim_{i\to\infty}(G + C^{(n_i)} - \hat{G}^{(n_i+1)}, \hat{G}^{(n_i+1)}) \\ &= 0 + 0.\end{aligned}$$

Similarly, if $V \in K_r \cap K_c$,

$$\begin{aligned}(G - G^*, V) &= (G + R - G^*, V) - (R, V) \\ &= \lim_{i\to\infty}(G + R^{(n_i)} - \tilde{G}^{(n_i)}, V) + \lim_{i\to\infty}(G + C^{(n_i)} - \hat{G}^{(n_i+1)}, V) \le 0 + 0.\end{aligned}$$

Thus G^* is the desired solution by (2.1) and (2.2). Moreover, since $-C$ minimizes $\| G + R - F\|^2$, $F \in K_c^*$, and $-R$ minimizes $\| G + C - F\|^2$, $F \in K_r^*$, we may use the distance reducing property of projections to say

$$\begin{aligned}\| C^{(n)} - C\|^2 &= \| G + C^{(n)} - (G + C)\|^2 \ge \| R^{(n+1)} - R\|^2 \\ &= \| G + R^{(n+1)} - (G + R)\|^2 \ge \| C^{(n+1)} - C\|^2 \quad \text{for all } n.\end{aligned}$$

Thus $R^{(n)} \to R$ and $C^{(n)} \to C$ as $n \to \infty$, which implies that $\hat{G}^{(n)} = G + R^{(n)} + C^{(n-1)}$ and $\tilde{G}^{(n)} = G + R^{(n)} + C^{(n)}$ both converge to $G^* = G + R + C$ as $n \to \infty$.

4. Other points. It is important to note that the solution $G^* = G + R + C$ does not uniquely determine R and C. In fact, if we begin with column smoothing rather than row smoothing we will obtain different limiting values for R and C even though the same limiting G^* is obtained.

As one would expect, this procedure works equally well when the order restrictions are modified to require nonincreasing rows, or nonincreasing columns, or both. One has only to change the one dimensional smoothing to operate in the appropriate direction.

We also wish to point out that G^* itself solves many more minimization (maximization) problems than the least squares problem stated above. For example, from Theorem 1.10 of Barlow et al., if Φ is an appropriate convex function and φ is a subgradient (basically a derivative) of Φ, then G^* solves the problem

$$\text{(4.1)} \qquad \text{maximize}_{F\in K_r\cap K_c} \sum_{i=1}^{a}\sum_{j=1}^{b}\{\Phi(f_{ij}) + (g_{ij} - f_{ij})\varphi(f_{ij})\}w_{ij}.$$

TABLE 4

First year GPA *of students entering The University of Iowa as Freshmen in the Fall of* 1978.

91 ≤ HSR ≤ 99	1.57 (4)	2.11 (5)	2.73 (18)	2.96 (39)	2.97 (126)	3.13 (219)	3.41 (232)	3.45 (47)	3.51 (4)
81 ≤ HSR ≤ 90	1.80 (6)	1.94 (15)	2.52 (30)	2.68 (65)	2.69 (117)	2.82 (143)	2.75 (70)	2.74 (8)	(0)
71 ≤ HSR ≤ 80	1.88 (10)	2.32 (13)	2.32 (51)	2.53 (83)	2.58 (115)	2.55 (107)	2.72 (24)	2.76 (4)	(0)
61 ≤ HSR ≤ 70	2.11 (6)	2.23 (32)	2.29 (59)	2.29 (84)	2.50 (75)	2.42 (44)	2.41 (19)	(0)	(0)
51 ≤ HSR ≤ 60	1.60 (11)	2.06 (16)	2.12 (49)	2.11 (63)	2.31 (57)	2.10 (40)	1.58 (4)	2.13 (1)	(0)
41 ≤ HSR ≤ 50	1.75 (6)	1.98 (12)	2.05 (31)	2.16 (42)	2.35 (34)	2.48 (21)	1.36 (4)	(0)	(0)
31 ≤ HSR ≤ 40	1.92 (7)	1.84 (6)	2.15 (5)	1.95 (27)	2.02 (13)	2.10 (13)	1.49 (2)	(0)	(0)
21 ≤ HSR ≤ 30	1.62 (1)	2.26 (2)	1.91 (5)	1.86 (14)	1.88 (11)	3.78 (1)	1.40 (2)	(0)	(0)
HSR ≤ 20	1.38 (1)	1.57 (2)	2.49 (5)	2.01 (7)	2.07 (7)	(0)	.75 (1)	(0)	(0)
High School Percentile Rank / ACTC	1–12	13–15	16–18	19–21	22–24	25–27	38–30	31–33	34–36

TABLE 5

Least squares doubly-nondecreasing regression function. Same categories as in Table 4.

1.87	2.17	2.73	2.96	2.97	3.13	3.41	3.45	3.51
1.87	2.17	2.52	2.68	2.69	2.79	2.79	2.79	2.79*
1.87	2.17	2.32	2.53	2.57	2.57	2.72	2.76	2.76*
1.87	2.17	2.29	2.29	2.46	2.46	2.46	2.46*	2.45*
1.73	2.06	2.12	2.13	2.25	2.25	2.25	2.25	2.25*
1.73	1.98	2.05	2.13	2.25	2.25	2.25	2.25*	2.25*
1.73	1.94	1.98	1.98	2.02	2.05	2.05	2.05*	2.05*
1.62	1.94	1.96	1.96	1.96	2.05	2.05	2.05*	2.05*
1.38	1.57	1.96	1.96	1.96	1.96	1.96	1.96*	1.96*

* not uniquely determined.

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

$$\text{minimize}_{F\in K_r\cap K_c} \sum_{i=1}^{a}\sum_{j=1}^{b}\{\Phi(f_{ij}) - g_{ij}f_{ij}\}w_{ij} \tag{4.2}$$

is solved by $(\varphi^{-1}(g_{ij}^*))$ where once again Φ is an appropriate convex function and φ is a subgradient of Φ.

Thus G^* solves a much wider range of problems than is readily apparent. For example, suppose one has a multinomial random vector X_{ij}, where the cell probabilities p_{ij} are placed in a rectangular grid and one wishes to find the maximum likelihood estimators for the p_{ij} subject to nondecreasing (nonincreasing) rows and columns. This problem can be phrased in terms of (4.1) from which it follows that the solution is given by G^* where $G = (X_{ij}/n)$ and $w_{ij} \equiv 1$.

Similarly, if the X_{ij} are independent bionomial (n_{ij}, p_{ij}) random variables, one can show that the maximum likelihood estimators for the p_{ij} subject to nondecreasing (nonincreasing) rows and columns is given by G^* where $G = (X_{ij}/n_{ij})$ and $w_{ij} = n_{ij}$.

In order to illustrate the algorithm on a larger table, we consider the data presented in Table 4. The entries are the first year grade point averages of 2397 students who entered the University of Iowa in the Fall of 1978. The independent variables are the composite scores on the ACT Assessment and the student's high school percentile rank. The expected first year grade point average is assumed to be a nondecreasing function of both of these independent variables. (The number in parentheses is the number of students in the category.)

Note that for values in the table having zero weight, the solutions are not uniquely determined. For example the value 2.79 in the second row and ninth column of Table 5 could be replaced by any value in the range [2.79, 3.51].

The least squares solution, correct to four significant digits, was obtained after 500 iterations (250 row smoothings and 250 column smoothings) at a cost of 9 seconds of CPU time. These results are given in Table 5 with the level sets indicated. Since the cost of our algorithm is essentially linear in the number of points in the grid, even very large arrays can be isotonized at a reasonable cost.

The algorithm extends in a natural fashion when one has more then two independent variables, and the proof of convergence follows similar lines to that given in Section 3. Thus suppose one has three independent variables and wants his function to be nondecreasing in rows, columns and layers. The algorithm would proceed as follows:

Step 1. First smooth $G = (g_{ijk})$ over all rows. Let $\hat{G}^{(1)}$ denote the row-smoothed values and $R^{(1)} = (r_{ijk}^{(1)}) = (\hat{g}_{ijk} - g_{ijk})$ the first set of "row increments".

Step 2. Smooth $G + R^{(1)}$ over all columns. Let $\tilde{G}^{(1)}$ denote the column-smoothed values and $C^{(1)} = (c_{ijk}^{(1)}) = \{\tilde{g}_{ijk} - (g_{ijk} + r_{ijk}^{(1)})\}$ the first set of "column increments."

Step 3. Smooth $G + R^{(1)} + C^{(1)}$ over all layers. Let $\check{G}^{(1)}$ denote the layer-smoothed values and $L^{(1)} = (l_{ijk}^{(1)}) = \{\check{g}_{ijk}^{(1)} - (g_{ijk} + r_{ijk}^{(1)} + c_{ijk}^{(1)})\}$ the first set of "layer increments".

Step 4. Now smooth $G + C^{(1)} + L^{(1)}$ over rows again to obtain the second set of row smoothed values $\hat{G}^{(2)}$. The row increments are updated and become $R^{(2)} = (r_{ijk}^{(2)}) = \{\hat{g}_{ijk}^{(2)} - (g_{ijk} + c_{ijk}^{(1)} + l_{ijk}^{(1)})\}$.

Step 5. Continue. Sequentially smooth over columns, layers and rows each time updating the increments.

Additional independent variables are handled in a similar fashion.

While we have been concerned with enforcing monotonicity in independent variables, the scheme is also amenable to certain other types of partial orders as well.

Acknowledgment. We thank Gordon Bril for writing a Fortran program to implement this algorithm. This program is available upon request from the authors. We thank Dr. Anne Cleary of the University of Iowa Evaluation and Examination Services for providing the data in Table 4. We also thank the reviewers for providing helpful and knowledgeable comments and for providing the diagram in Figure 1.

REFERENCES

AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T., and SILVERMAN, E. (1955). An empirical distribution function for sampling with complete information. *Ann. Math. Statist.* **26** 641–647.

BARLOW, R. E., BARTHOLOMEW, D. J., BREMNER, J. M., and BRUNK, H. D. (1972). *Statistical Inference under Order Restriction.* Wiley, New York.

BARLOW, R. E. and BRUNK, H. D. (1972). The isotonic regression problem and its dual. *J. Amer. Statist. Assoc.* **67** 140–147.

BRUNK, H. D., EWING, G. M., and UTZ, W. R. (1957). Minimizing integrals in certain classes of monotone functions. *Pacific J. Math* **7** 833–847.

DYKSTRA, RICHARD L. (1981). An isotonic regression algorithm. *J. Statist. Planning and Inference* **5** 355–363.

GEBHARDT, FRIEDRICH (1970). An algorithm for monotone regression with one or more independent variables. *Biometrika* **57** 263–271.

KOLEN, MICHAEL J., SMITH, WILLIAM M., and WHITNEY, DOUGLAS R. (1977). Methods of smoothing double-entry expectancy tables applied to the prediction of success in college. Research Report No. 91. Evaluation and Examination Service, The University of Iowa (Presented at the annual meeting, American Educational Research Association, New York, N.Y. April 1977.)

KOLEN, MICHAEL J. and WHITNEY, DOUGLAS R. (1978). Methods of smoothing double-entry expectancy tables applied to the prediction of success in college. *J. Educational Measurement* **15** 201–211.

MAKOWSKI, GARY G. (1974). Consistency on logarithmic convergence for iterated nondecreasing regression estimators. Unpublished manuscript.

PERRIN, D. W. and WHITNEY, D. R. (1976). Methods for smoothing expectancy tables applied to the prediction of success in college. *J. Educational Measurement* **13** 223–231.

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242

DEPARTMENT OF STATISTICS
AND ACTUARIAL SCIENCE
UNIVERSITY OF IOWA
IOWA CITY, IOWA 52242