AD-A102 559 MISSOURI UNIV-COLUMBIA DEPT OF STATISTICS F/G 12/1
AN ALGORITHM FOR ISOTONIC REGRESSION IN THE PLANE.(U)
APR 81 R L DYKSTRA, T ROBERTSON N00014-78-C-0655
UNCLASSIFIED TR-101 NL

1 OF 1
AD A
102559

END
DATE
FILMED
9-81
DTIC

LEVEL II

AD A102559

University of Missouri-Columbia

An Algorithm for Isotonic Regression in the Plane

by

Richard L. Dykstra
Tim Robertson

Technical Report 101
Department of Statistics
April, 1981

Mathematical

Sciences

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 101	2. GOVT ACCESSION NO.: AD-A102559
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): (6) An Algorithm for isotonic regression in the plane (9)	5. TYPE OF REPORT & PERIOD COVERED: Technical Report
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): (10) Richard L. Dykstra, Tim Robertson	8. CONTRACT OR GRANT NUMBER(s): (10) N00014-78-C-0655, N00014-80-C-0321
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Department of Statistics, University of Missouri, Columbia, MO 65211; Department of Statistics, University of Iowa, Iowa City, Iowa 52242	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research, Department of the Navy, Arlington, VA (11)	12. REPORT DATE: Apr[redacted]981
	13. NUMBER OF PAGES: 21
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (12) 23	15. SECURITY CLASS. (of this report):
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: Unclassified

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED.

(14) TR-101

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Isotone regression, ininimal lower sets algorithm, convex cones, dual convex cones, projections

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Algorithms for solving the isotone regression problem in two dimensions are difficult to implement becasue of the large number of lower sets present. ~~Here,~~ a new algorithm for solving this problem, based on a simple iterative technique, is proposed and shown to converge to the correct solution.

AN ALGORITHM FOR ISOTONIC REGRESSION

IN THE PLANE

by

Richard L. Dykstra and Tim Robertson

(University of Missouri, Columbia and University of Iowa)

ABSTRACT

Algorithms for solving the isotonic regression problem in two dimensions are difficult to implement because of the large number of lower sets present. Here a new algorithm for solving this problem based on a simple iterative technique is proposed and shown to converge to the correct solution.

AMS Classification Numbers: Primary 65D15
49D05

Key Words and Phrases: Isotone regression, minimal lower sets algorithm, convex cones, dual convex cones, projections.

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

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 $\binom{a+b}{a}$. 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, etc.

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 to first 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 smoothing 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 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. The data is presented 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 their 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.

Several of the algorithms discussed in Section 2.3 of Barlow et al. are basically methods of finding linear orders which are consistent with a partial order. We have not been able to write a program which implements any of these methods

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

TABLE 2

The probability of making a "B or better" GPA
Kolen and Whitney Method

.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
Least squares isotonic regression (weights = cell frequencies)

.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

for a doubly nondecreasing regression function in a reasonable amount of time.

In this paper we present an algorithm for calculating the least squares isotonic regression function which is increasing in each of two variables. This algorithm uses successive one dimensional smoothings and is very efficient and easy to program. To illustrate, we applied the algorithm to a 20 by 20 table of random numbers. The isotonic regression was obtained, correct to four significant digits, after 400 iterations, required 39 seconds of CPU time at a cost of one dollar and thirty-five cents. This algorithm is described in Section 3. In Section 2 we summarize some well known properties of isotonic regression 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 $i-k \le 0$ and $j-\ell \le 0$. We denote an arbitrary real function whose domain is Ω as a matrix, i.e.,

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

(Note that this is not the usual matrix notation where g_{ij} refers to the entry in the $i^{\underline{th}}$ row and $j^{\underline{th}}$ column.)

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

$$\text{Minimize} \quad \Sigma_{i,j}(g_{ij} - f_{ij})^2 w_{ij}$$

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 (for example, see Theorem 7.8 in Barlow et al.) that the solution to the isotonic regression problem, say G^*, is characterized by the properties, $G^* \in K$,

$$(2.1) \qquad \Sigma_{i,j}(g_{ij} - g^*_{ij})g^*_{ij} w_{ij} = 0, \quad \text{and}$$

$$(2.2) \qquad \Sigma_{i,j}(g_{ij} - g^*_{ij})h_{ij} w_{ij} \le 0,$$

for all functions $H \in K$.

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:

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."

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 $f_{i1} \le f_{i2} \le \cdots \le f_{ib}$ for $i = 1, \cdots, a$. We call $C^{(1)} = \tilde{G}^{(1)} - (G + R^{(1)})$ the first set of "column increments." Note that $\tilde{G}^{(1)} = G + R^{(1)} + C^{(1)}$.

3. Etc. At the beginning of the n^{th} cycle, we obtain $\hat{G}^{(n)}$ by isotonizing $G + C^{(n-1)}$ over rows. The n^{th} 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 n^{th} set of column increments is given by $C^{(n)} = \tilde{G}^{(n)} - (G + R^{(n)})$, or equivalently $\tilde{G}^{(n)} = G + R^{(n)} + C^{(n)}$.

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

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

we first show

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

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

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

and

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

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

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

Similarly,

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

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

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

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 $<<$) such that either $\{r^{(n)}_{i_0,j_0}\}$ or $\{c^{(n)}_{i_0,j_0}\}$ is unbounded. Say there exists a subsequence $\{n_i\}$ such that $r^{(n_i)}_{i_0,j_0} \to -\infty$. (Since $\sum_{i=1}^{i_0} r^{(n)}_{i,j_0} \leq 0$ for all n (see Barlow and Brunk (1972)), $r^{(n_i)}_{i_0,j_0} \to \infty$ 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 $C^{(n_i)}_{i_0,j_0} \to \infty$. This, in turn, contradicts the fact that (i_0,j_0) is minimal since $\sum_{j=1}^{i_0} c^{(n)}_{i_0,j} \leq 0$ for all n.

Projections onto convex sets are distance reducing so that

$$\|C^{(i)}-C^{(i-1)}\|^2 = \|G+C^{(i)}-(G+C^{(i-1)})\|^2$$

$$\geq \|R^{(i+1)}-R^{(i)}\|^2 = \|G+R^{(i+1)}-(G+R^{(i)})\|^2 \tag{3.2}$$

$$\geq \|C^{(i+1)}-C^{(i)}\|^2 \quad \text{for all } i.$$

We now show that

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

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

$$(3.4)\qquad |r^{(i+1)}_{i_0,j_0} - r^{(i)}_{i_0,j_0}| > \epsilon \quad \text{for infinitely many } i.$$

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

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

If we write

$$(3.6)\qquad \|R^{(i+1)} - R^{(i)}\|^2 - \|C^{(i+1)} - C^{(i)}\|^2$$
$$= \|G + R^{(i)} + C^{(i)} - (G+R^{(i+1)}+C^{(i+1)}\|^2$$
$$+ 2(G + R^{(i)} + C^{(i)} - (G + R^{(i+1)} + C^{(i+1)}), C^{(i+1)} - C^{(i)}),$$

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

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

In similar fashion, beginning with

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

we can conclude

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

Subtracting (3.7) from (3.8) yields

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

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

$$(r_{i_0,j_0}^{(N_0+1+i)} - r_{i_0,j_0}^{(N_0+i)}), \quad i = 1,2,\cdots,n_0$$

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 $R^{(n_i)} \longrightarrow R$ and $C^{(n_i)} \longrightarrow C$. 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,

$$(G-G^*,G^*) = (G+R-G^*,G^*) - (R,G^*)$$

$$= \lim_{i\to\infty}(G+R^{(n_i)}-\tilde{G}^{(n_i)},\tilde{G}^{(n_i)}) + \lim_{i\to\infty}(G+C^{(n_i)}-\hat{G}^{(n_i+1)},\hat{G}^{(n_i+1)})$$

$$= 0 + 0.$$

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

$$(G-G^*,V) = (G+R-G^*,V) - (R,V)$$

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

Thus G^* is the desired solution by (2.1) and (2.2). Moreover, since

$$-C \text{ minimizes } \|G + R - F\|^2, \quad F \in K_c^{*w}$$

and

$$-R \text{ minimizes } \|G + C - F\|^2, \quad F \in K_r^{*w},$$

we may use the distance reducing property of projections to say

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

$$= \|G + R^{(n+1)} - (G+R)\|^2 \geq \|C^{(n+1)} - C\|^2 \quad \text{for all } n.$$

Thus

$$R^{(n)} \longrightarrow R \quad \text{and} \quad C^{(n)} \longrightarrow C \quad \text{as} \quad n \longrightarrow \infty$$

which implies that

$$\hat{G}^{(n)} = G + R^{(n)} + C^{(n-1)} \quad \text{and} \quad \tilde{G}^{(n)} = G + R^{(n)} + C^{(n)}$$

both converge to

$$G^* = G + R + C \quad \text{as} \quad n \longrightarrow \infty.$$

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 a column smoothing rather than row smoothing we will obtain different limiting values from 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. The procedure can also be adapted to higher dimensions, although in this case the number of required smoothings quickly becomes large.

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

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

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

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

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 binomial (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 $w_{ij} = n_{ij}$.

Finally, in order to illustrate the algorithm on a larger table, we considered 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.)

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 CPU time and 84 cents. 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.

TABLE 4

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

High School Percentile Rank \ ACTC	1-12	13-15	16-18	19-21	22-24	25=27	28-30	31-33	34-36
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)

TABLE 5

Least squares doubly nondecreasing regression function

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	2.46
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

ACKNOWLEDGEMENT. 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.

REFERENCES

Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T., & Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. Ann. Math. Statist. 26, 641-647.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M., & Brunk, H.D. (1972). Statistical Inference under Order Restriction. John Wiley & Sons, 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., & Utz, W.R. (1957). Minimizing integrals in certain classes of monotone functions. Pacific J. Math. 7, 833-847.

Kolen, Michael J., Smith, William M., & 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. & 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. & Whitney, D.R. (1976). Methods for smoothing expectancy tables applied to the prediction of success in college. _J. Educational Measurement_ 13, 223-231.