AN ABSTRACT OF THE THESIS OF

 CHU-IN CHARLES LEE (Name)
 for the degree DOCTOR OF PHILOSOPHY (Degree)

 in STATISTICS (Major Department)
 presented on June 9, 1975 (Date)

 Title:
 CONDITIONAL EXPECTATION GIVEN A σ-LATTICE.

 ISOTONIC REGRESSION, UNIVARIATE AND MULTIVARIATE.

 Redacted for Privacy

 Abstract approved:

Hugh D. Brunk

Statistical inference sometimes involves order restrictions which are usually due to prior knowledge. Such order restrictions whenever they occur, will be a major factor in performing a good and reliable statistical analysis. How to make use of these order restrictions is one of the most interesting and the most facinating subjects in statistics nowadays.

Isotonic regression has been formulated and studied for the past 25 years by many statisticians. Their researches are very successful and very fruitful. The theory of such a statistical analysis is called the conditional expectation given a σ -lattice which is an extension of the conditional expectation. Conditional expectation given a σ -lattice has been analyzed in the same direction as that of the conditional expectation. The understanding of the concept of the latter will be very helpful for the study of the former. When several measurements have been made at each given sample and each measurement has its own restriction, the point estimation of this type is called the multivariate isotonic regression.

The structures, the properties and the algorithms of isotonic regression and of multivariate isotonic regression are the major research in this thesis. Conditional expectation given a σ -lattice and isotonic regression are presented in separate chapters. They shall be considered as a single unit. The former emphasizes properties and the latter emphasizes algorithms. Multivariate isotonic regression is treated in the simplest case. Only bivariate isotonic regression with linear ordering in each variate will be considered.

The fundamental concept is the generalized projection. Some necessary and sufficient conditions have been presented. Isotonic regression and multivariate isotonic regression are discussed in the finite case. In such a situation, they are generalized projections to finitely generated cones. However, in such general structures, the monotonicity and the averaging property will not be preserved. Although the algorithms are different from each other, they are presented in the same pattern, i.e., as successive projections to linear spaces.

Conditional Expectation Given a σ-Lattice. Isotonic Regression, Univariate and Multivariate.

by

Chu-In Charles Lee

A THESIS

submitted to

Oregon State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Completed June 1975

Commencement June 1976

APPROVED:

Redacted for Privacy

Professor of Statistics

in charge of major

Redacted for Privacy

Head of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented _____ June 9, 1975

Typed by Clover Redfern for _____ Chu-In Charles Lee

ACKNOWLEDGMENT

The author wishes to express his deepest appreciation to Professor H.D. Brunk who initially suggested the problem and whose guidance has been very precious all throughout this research. The author considers himself very fortunate to have the opportunity to work with such a person. To Professor Brunk the author expresses his wholehearted gratitude.

The author wishes to express his appreciation to Dr. David Birkes for many rewarding conversations during the time this thesis was being written.

The author wishes to thank the many members of the Department of Statistics and of the Department of Mathematics with whom he has had valuable course work and discussions, especially to Dr. Thomas Chow, Dr. Mark Lembersky, Dr. Justus Seely and Dr. David Thomas.

The author also acknowledges the financial assistance of the Department of Statistics under the chairmanship of Dr. Lyle Calvin, the Department of Mathematics and the Computer Center of the University, which allowed him to complete his educational program.

Final thanks go to the author's wife, Lisa, for her encouragement, understanding, patience, typing and moral support.

TABLE OF CONTENTS

Chapter	Page
 INTRODUCTION I.1 Forword I.2 Statistical Problems Under Order Restrictions I.3 Organization 	1 1 2 8
 II. GENERALIZED PROJECTION II.1 On a Closed Convex Set II.2 On a Closed Convex Cone II.3 On a Finitely Generated Cone 	12 12 19 25
 III. ISOTONIC REGRESSION III. 1 Preliminaries III. 2 Upper Sets and Indicators III. 3 On a Linearly Ordered Set III. 4 On a Tree III. 5 On a Partially Ordered Set III. 6 Some Related Problems 	33 33 41 48 60 66 76
 IV. CONDITIONAL EXPECTATION GIVEN A σ-LATTICE IV.1 σ-Lattice and Σ-Measurable Random Variables IV.2 Conditional Expectation as a Generalized Projection IV.3 Conditional Expectation Given a σ-Lattice IV.4 Properties of Conditional Expectation IV.5 On a Linearly Ordered σ-Lattice 	85 85 95 100 106 115
 V. MULTIVARIATE ISOTONIC REGRESSION V.1 Introduction V.2 Bivariate Isotonic Regression V.3 Simplified Projection V.4 Approximation BIBLIOGRAPHY 	120 120 124 136 144
 APPENDICES Appendix I: (Daily Maximum) Temperatures Measured at Oregon State University in 1974 Appendix II: Combined Explosive Rates Appendix III: MLE of Variance Components Appendix IV: Bounded Isotonic Regression Appendix V: Simplified Projection Algorithm Index 	152 153 154 157 160 164

CONDITIONAL EXPECTATION GIVEN A σ-LATTICE. ISOTONIC REGRESSION, UNIVARIATE AND MULTIVARIATE.

I. INTRODUCTION

I.1 Forward

Over the past 25 years, statisticians have formulated and studied problems of statistical inference in the presence of order conditions arising in various contexts. For the most part, these problems may be interpreted in terms of isotonic regression over a quasi-ordered set. Some examples will be presented in the next section. Because of its theoretical interest and broad applications, the author devoted himself to research in this field.

The regression function of one random variable X on another, Y, is the conditional expectation $\mu(y) = E(X | Y = y)$, and furnishes the best fit to the distribution of X by a function of Y in the sense of least-squares. Isotonic regression is introduced in Chapter III by means of least-squares, as a generalization of the regression function. As a generalization of conditional expectation, this concept is called conditional expectation given a σ -lattice. Since it involves least-squares, isotonic regression is an instance of quadratic programming. Examining the objective function and the feasible region of an isotonic regression problem, we shall quickly discover that the arithmetical manipulations involved are mainly the routine calculation of means.

Following the isotonic regression is the multivariate isotonic regression. The latter is quite complicated as we shall soon find out in Chapter V. But its nature is essentially the same as that of the former. Thus its research may take similar directions to that of the isotonic regression. Generalized projection, isotonic regression, conditional expectation given a σ -lattice, and multivariate isotonic regression are studied respectively from Chapter II to Chapter V. The author hopes that these results will provide a general outlook for the problems of least-squares under order restrictions.

I.2 Statistical Problems Under Order Restrictions

The (daily maximum) temperatures measured (in Fahrenheit) at Oregon State University during the two periods, March 16th, 17th, 18th and April 16th, 17th, 18th, 19th, 20th of 1974, were 68°, 58°, 56° and 56°, 59°, 63°, 58°, 55° respectively. The three days' average temperature in March is $x = 60.7^{\circ}$ and the five days' average temperature in April is $y = 58.2^{\circ}$. One wishes to estimate long run average (daily maximum) temperature in March μ and that in April ν with these eight observations. If we believe that $\mu \leq \nu$, then the estimates $\hat{\mu}$ and $\hat{\nu}$ for μ and ν respectively should therefore satisfy the constraing $\hat{\mu} < \hat{\nu}$. The statistical hypothesis testing for $\mu = \nu$ against its alternative $\mu < \nu$ based on the eight observations will favor the former. Therefore, $\hat{\mu} = \hat{\nu}$ and $\hat{\mu} = \hat{\nu} = 59.1^{\circ}$, the average of the above eight observations, is the best fit to x and y in the sense that $3(x-\hat{\mu})^2 + 5(y-\hat{\nu})^2$ has the smallest value among $3(x-\mu)^2 + 5(y-\nu)^2$ with $\mu \leq \nu$. (2.4) shows that $3(x-\mu)^2 + 5(y-\nu)^2 \geq [3(x-\hat{\mu})^2+5(y-\hat{\nu})^2] + [3(\hat{\mu}-\mu)^2+5(\hat{\nu}-\nu)^2]$. It follows that whenever x and y are consistent estimators for μ and ν then so are $\hat{\mu}$ and $\hat{\nu}$.

The average (daily maximum) temperature measured at Oregon State University in March 1974 was 54° and that in April 1974 was 57.5° (cf. Appendix I). If we compare a three consecutive days' average temperature in March x with a five consecutive days' average in April y of that year, we shall find out that there are 284 pairs such that x > y, 2 pairs such that x = y and 468 pairs such that x < y. This shows that $\mu = 54°$ is less than $\nu = 57.5°$ by as much as 3.5°, but there is about 3/8 chance that one may observe x > y.

Experimental investigations sometimes deal with continuous variables which can not be measured in practice. For example, Dixon and Massey (1957) describe a procedure for testing the sensitivity of explosives to shock. A weight is dropped on specimens of the same explosive mixture from various heights. Suppose that a given specimen, if it is dropped at some chosen height, will explode; so will it if dropped from any greater height. On the other hand, if it does not explode at that height, neither will any lesser height cause it to explode. Therefore, we may assume that there is a critical height associated with each specimen. The investigator's interest is in the rates of explosions from a population of such specimens dropped from various heights.

In such an experiment it is impossible to make more than one observation on a given specimen. Once a test has been made, and the specimen does not explode, one may suspect that its critical height will be altered. Thus a valid result can not be obtained from a second test. An experimental designer usually divides the sample of specimens into several groups and tests one group at one height, a second group at another height, etc. The data gathered are the numbers of those exploded and of those not exploded at each height.

Let X be the random variable, critical height, with distribution function $F(h) = P_r \{X \le h\}$. Suppose 50 experiments have been made with ten tests at each height h_i , i = 1, 2, ..., 5; for convenience suppose the heights $h_1, h_2, ..., h_5$ are arranged in increasing order. The 50 tests may be regarded as a set of 50 independent trials of events having probabilities $p_i = F(h_i)$, i = 1, 2, ..., 5, of success if success means that the given specimen will explode at the height. If a large number of trials is made at each height h_i , i = 1, 2, ..., 5, the ratios r_i , number of success divided by number of trials, each determined for a particular height, will with high probability be in non-decreasing order. The best estimates of the probability are then these ratios. Suppose that a small number of trials will be allowed; these ratios may not be in monotone order. In such a case, the maximum likelihood estimates of the probabilities p_1, p_2, \ldots, p_5 are determined under order restrictions. This is a typical example of an isotonic regression problem over a linearly ordered finite set.

Ayer and coworkers (1955) showed that the maximum likelihood estimates of p_1, p_2, \dots, p_n subject to $p_1 \leq p_2 \leq \dots \leq p_n$ are the same as the least-squares estimates of p_1, p_2, \ldots, p_n subject to the same constraints. Suppose that $r_1 = 0.3$, $r_2 = 0.2$, $r_3 = 0.7$, $r_4 = 0.8$ and $r_5 = 0.5$. Then the optimal solution is that $\hat{p}_1 = \hat{p}_2 = 0.25$ and $\hat{p}_3 = \hat{p}_4 = \hat{p}_5 = 0.67$ (cf. Example 3.1). There are ten tests at the fifth height with five specimens exploded, 20 tests at the fifth height and the fourth height with 13 specimens exploded, 30 tests at the fifth, fourth and third heights with 20 specimens exploded, etc. The explosive ratios in such combinations of heights are respectively .5, .65, .67, .55 and .5 with .67 the largest value. If we drop the 30 specimens at the fifth height which were dropped at the third, fourth and fifth heights, then the explosive ratio will be at least . 67. On the other hand, if we drop these 30 specimens at the third height, then the explosive ratio will be no more than .67.

Similarly for the situation at the first and second heights (cf. Appendix II).

Let $q_i = (p_{i+1} - p_i)/(h_{i+1} - h_i)$, i = 1, ..., 4. The ordering $p_1 \leq p_2 \leq ... \leq p_5$ is equivalent to that $q_i \geq 0$, i = 1, 2, 3, 4. In some situations, we require that $q_i \geq a_i$, i = 1, 2, 3, 4 for a set of non-negative real numbers a_1 , a_2 . a_3 and a_4 . Such a consideration was introduced by Reid (1968). The maximum likelihood estimates of $p_1, p_2, ..., p_5$ subject to $p_1 + 0.05 \leq p_2$, $p_2 + 0.05 \leq p_3$, $p_3 + 0.05 \leq p_4$, $p_4 + 0.05 \leq p_5$ are $p_1 = 0.225$, $p_2 = 0.275$, $p_3 = 0.617$, $p_4 = 0.667$, $p_5 = 0.717$.

Estimation of variance components in random models is also an example with order restrictions. Consider the two-way random model

$$y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk}$$

i = 1,...,r; j = 1,...,s; k = 1,...,t where $\{a_i\}, \{b_j\}, \{c_{ij}\}$ and $\{e_{ijk}\}$ are sets of mutually independent normal variates with means zero and variance $\sigma_A^2, \sigma_B^2, \sigma_{AB}^2$ and σ^2 respectively. Let

$$SS_{A} = st \Sigma_{i=1}^{r} (y_{i..} - y_{...})^{2}$$

$$SS_{B} = rt \Sigma_{j=1}^{s} (y_{.j.} - y_{...})^{2}$$

$$SS_{AB} = t \Sigma_{i=1}^{r} \Sigma_{j=1}^{s} (y_{ij.} - y_{i..} - y_{.j.} + y_{...})^{2}$$

and

$$SS_{e} = \Sigma_{i=1}^{r} \Sigma_{j=1}^{s} \Sigma_{k=1}^{t} (y_{ijk} - y_{ij.})^{2}$$

and let $MS_A = SS_A/(r-1)$, $MS_B = SS_B/(s-1)$, $MS_{AB} = SS_{AB}/(r-1)(s-1)$ and $MS_e = SS_e/rs(t-1)$. Then the expectations of MS_A , MS_B , MS_{AB} and MS_e are $\sigma^2 + t\sigma^2_{AB} + st\sigma^2_A$, $\sigma^2 + t\sigma^2_{AB} + rt\sigma^2_B$, $\sigma^2 + t\sigma^2_{AB}$ and σ^2 respectively. Since variances are non-negative, $E(MS_A) \ge E(MS_{AB}) \ge E(MS_e)$ and $E(MS_B) \ge E(MS_{AB}) \ge E(MS_e)$. However, the sample estimates MS_A , MS_B , MS_{AB} and MS_e may not be in such order. Estimating these variance components subject to order constraints is a problem of an isotonic regression over a tree (cf. Appendix III).

Let (X(t), Y(t)) be a random vector having a bivariate normal distribution with mean vector $(\mu(t), \nu(t))$ for each t and with a known constant covariance matrix. It has been indicated that the means $\mu(t)$ and $\nu(t)$ are respectively a monotone increasing function of t and a monotone decreasing function of t. Suppose a random sample of size N has been given with N_i observations at each parameter value t_i , i = 1, ..., n. Multivariate isotonic regression will yield the maximum likelihood estimates to the sample by a monotone vector-valued function.

For instance, Bhattacharyya and Kotz (1966) use freezing dates and thawing dates for Lake Mendota for a period of 111 years to test against warming trend. Suppose there is a warming trend in the Lake Mendota during that period and suppose the covariance matrix is constant for each year and it is given. Then the means $\mu(t)$ and $\nu(t)$ can be obtained by using the Simplified Projection algorithm.

I.3 Organization

As the title implies, the core of the study is the conditional expectation given a σ -lattice and its applications to the isotonic regression and to the multivariate isotonic regression. The application to the latter is indirect.

The algorithms that appear in this paper, Projection of Minimum Violators, Projection of Violators, Pool-Adjacent-Violators, Minimum Violators, Minimum Upper Sets, Maximum Lower Sets, and Simplified Projection have the similar structures. Making use of the smoothing property, we obtain the projections of a given vector X to a strictly decreasing sequence of linear spaces successively until we have the desired solution. Therefore, the justification for one algorithm can be applied to that of others. The only difference is the way we identify those pivotal elements and the way we obtain the projection to a linear space. Since the Projection of Minimum Violators algorithm comes first, its justification is given in more detail. Conditional expectation of a square-integrable random variable given a σ -lattice, isotonic regression and multivariate isotonic regression are generalized projections to closed convex cones. The generalized projection presented in Chapter II furnishes the background for these topics. Let H be a Hilbert space, let C be a closed convex cone in H and let S be a linear space in H such that $S \supset C$. The smoothing property shows that P(X|C) = P(P(X|S)|C) for every $X \in H$. This is the property that we shall use constantly for various algorithms in this thesis. The chapter is by itself an extension of results in Brunk (1965).

The aim of Chapter III, isotonic regression, is to develop some efficient algorithms for various types of isotonic regression problems through the use of indicators. We introduce a novel approach, according to which the isotonic regression can be seen as an orthogonal projection to a linear space. But which linear space is a proper one is uncertain. Therefore, we have to use the smoothing property and pivotal elements successively to obtain a linear space we want. Once a proper linear space is found, the isotonic regression for a given function X is the orthogonal projection of X to that linear space. Similarly for the multivariate isotonic regression. For the problems of isotonic regression over a linearly ordered set and isotonic regression over a tree, pivotal elements can be identified easily. For the isotonic regression over a partially ordered set, Alexander (1970) introduced a method which rewrites a partial ordering into a consistent linear ordering.

The properties of isotonic regression are presented in Chapter IV, conditional expectation given a σ -lattice. Observing the similarities between the isotonic regression and the regression, one can generate a parallel concept, conditional expectation given a σ -lattice, by that of conditional expectation. Johansen (1967), and Brunk and Johansen (1970) introduced that conditional expectation given a σ -lattice as the Lebesgue-Radon-Nikodym derivative given a σ -lattice. Conditional expectation appears as a special case of conditional expectation given a σ -lattice. A special interest attaches to the case when the σ -lattice is linearly ordered. In such a situation, there are martingale and submartingale structures.

Although multivariate isotonic regression is the title of Chapter V, all the work has been done on the case of bivariate isotonic regression over a linearly ordered finite set. The author hopes that these results can serve as the "frontier" for the general case. The averaging property and the monotonicity in the bivariate problem have quite different significance as compared with the univariate case, which makes the problem a lot more complicated.

The monograph of Barlow and coworkers (1972), which contains a bibliography listing of 247 published works referred to this subject, is a complete reference for this thesis, especially its Chapters 1, 2 and 7.

The sequences of theorems, lemmas, and examples will be numbered for each chapter individually. Each corollary will be numbered according to that of the theorem it follows. Theorems and formulas which are quoted will be presented without proof.

Terminologies and notations are standard. Terms will be underlined when they are introduced for the first time in the thesis. Capital letters X, Y and Z will be used to denote functions, vectors, matrices or random variables. Small letters α , β , γ , x, y, z will be used to denote real numbers. The inequality \leq between a pair of functions stands for the same inequality between corresponding components. The same notation \leq between a pair of Greek letters μ and ν stands for a quasi-ordering. The meet $\alpha \wedge \beta$ is the smaller number of α and β . The meet $X \wedge Y$ between a pair of functions is a function such that $(X \wedge Y)(\omega) = X(\omega) \wedge Y(\omega)$ at each argument ω .

11

II. GENERALIZED PROJECTION

II.1 On a Closed Convex Set

Let H be a <u>Hilbert space</u>: a linear space with an inner product (\cdot, \cdot) such that every Cauchy sequence converges in the space, where the <u>norm</u> of the element X is defined by $||X|| := (X, X)^{1/2}$. Let C be a subset of H. An element Y is said to be a <u>boundary point</u> of C if for every $\delta > 0$, there exist $Z \in C$ and $X \notin C$ such that $||Z-Y|| < \delta$ and $||X-Y|| < \delta$. C is said to be <u>closed</u> if it contains all its boundary points. It therefore follows that the limit of a convergent sequence from a closed set is in the set. C is said to be <u>convex</u> if Y, $Z \in C$ implies $\lambda Y + (1-\lambda)Z \in C$ for each real number λ between zero and one. Throughout the rest of the section, the letter C will denot a nonempty closed convex set in the Hilbert space H. The following result is well known.

<u>Theorem 2.1</u>. Let C be a nonempty closed convex set in the Hilbert space H. For every $X \in H$, there exists a unique $X^* \in C$ which minimizes ||X-Z|| among all $Z \in C$.

The closedness of C yields the existence of X^* and the convexity of C yields the uniqueness of X^* . Such an element X^* , denoted by P(X|C), is called a projection of X on C and the

operator $P(\cdot | C)$ is called a <u>generalized projection</u>. The generalized projection $P(\cdot | C)$ depends on C and also depends on the inner product (\cdot, \cdot) associated to the Hilbert space H. It is trivial that

(2.1)
$$P(Y+X|Y+C) = Y + P(X|C)$$
 for each $Y \in H$

and

(2.2) P(aX | aC) = aP(X | C) for each real a,

where $Y + C := \{Y+Z: Z \in C\}$ and $aC := \{aZ: Z \in C\}$. The sets Y + C and aC are closed and convex.

If $X \in C$, then P(X | C) = X. A well known result which plays an important role in the theory of projection is the following theorem.

<u>Theorem 2.2.</u> Let C be a nonempty closed convex set in the Hilbert space H. If $X \notin C$, then $P(X \mid C)$ is a boundary point of C.

Therefore, if $X \notin C$, then the candidates for P(X|C) are boundary points of C. Let $X \notin C$, $Z \notin C$ and let $B(X,Z) := \{Y: || Y-(X+Z)/2 || \le || X-Z ||/2 \}$. From a geometric point of view, if $Y \notin B(X,Z)$ and $Y \neq Z$ then $|| Y-X || \le || Z-X ||$. It follows that if X^* is P(X|C) then $B(X,X^*) \cap C$ is a singleton $\{X^*\}$. Since both C and $B(X,X^*)$ are convex, there is a separating hyperplane which separates C from $B(X,X^*)$. The hyperplane supports $B(X, X^*)$ at X^* , so it can be characterized as $\{Y:(X-X^*, Y) = (X-X^*, X^*)\}$. From the separation by the hyperplane, we have $(X-X^*, X^*-Z) \ge 0$ for every $Z \in C$. On the other hand, the inequality $||Y-(X+Z)/2|| \le ||X-Z||/2$ is equivalent to $(X-Y, Y-Z) \ge 0$. If we replace Y by X*, then $||X-X^*|| \le ||X-Z||$. We have thus proved the following theorem.

<u>Theorem 2.3.</u> Let C be a closed convex set in the Hilbert space H and let $X \in H$. An element $X^* \in C$ is the projection of X on C if and only if

$$(2.3) \qquad (X-X^*, X^*-Z) \ge 0 \quad \text{for every} \quad Z \in C.$$

Analytic proofs of the theorem have been given by Brunk (1965) and by Barlow and coworkers (1972). By the identity $\|X-Z\|^2 = \|X-X^*\|^2 + \|X^*-Z\|^2 + 2(X-X^*,X^*-Z)$, we have the following result.

<u>Corollary 2.3.1</u>. An element $X^* \in C$ is the projection of X on C if and only if

(2.4)
$$||X-Z||^2 \ge ||X-X^*||^2 + ||X^*-Z||^2$$
 for every $Z \in C$.

It follows that

(2.5)
$$\| P(X|C) - Z \| \leq \| X - Z \|$$
 for every $Z \in C$,

and if the equality holds for an element $Z \in C$, then P(X | C) = X.

After a simple manipulation, an immediate result from (2.3) is that

(2.6)
$$\| \mathbf{P}(\mathbf{X}_1 | \mathbf{C}) - \mathbf{P}(\mathbf{X}_2 | \mathbf{C}) \| \le \| \mathbf{X}_1 - \mathbf{X}_2 \|$$

for each pair $X_1, X_2 \in H$. This yields the following corollary.

<u>Corollary 2.3.2.</u> The generalized projection $P(\cdot | C)$ reduces distance. Therefore, it is a continuous mapping on H.

Let C_1 and C_2 be closed convex sets in H with $C_1 \subset C_2$. For every X \in H, (2.4) showed that

(2.7)
$$\| P(X | C_2) - P(X | C_1) \|^2 \le \| X - P(X | C_1) \|^2$$

- $\| X - P(X | C_2) \|^2$

Inequalities (2.3), (2.4), (2.6) and (2.7) are given by Brunk (1965). They are fundamental to the concept of the generalized projection $P(\cdot | C)$.

Let S be a closed linear space. The closed convex set Y + Sis called an <u>affine space</u>. Let A = Y + S. Then $A = Y_1 + S$ for any $Y_1 \in A$. Write A = P(X|A) + S. By (2 3), we have that $X^* = P(X|A)$ if and only if $X^* \in A$ and $(X-X^*, Z) = 0$ for every $Z \in S$ or equivalently, for any $Y_1 \in A$, $(X-X^*, Y_1-Z) = 0$ for every $Z \in A$. If A is also a closed linear space, i.e., $0 \in A$, then $X^* = P(X|A)$ if and only if $(X-X^*, Z) = 0$ for every $Z \in A$. It is obvious that $P(P(X|S_2)|S_1) = P(X|S_1)$ if S_1 and S_2 are closed linear spaces and $S_1 \subset S_2$. A similar result under a weaker hypothesis is given below.

<u>Theorem 2.4.</u> Let C_1 and C_2 be two closed convex sets in H. For every $X \in H$, if either C_2 is an affine space and $P(P(X | C_2) | C_1) \in C_2$, or C_1 is an affine space and $C_1 \subset C_2$, then $P(P(X | C_2) | C_1) = P(X | C_1 \cap C_2)$.

<u>Proof</u>. The intersection of closed convex sets is also closed and convex. Therefore, the operator $P(\cdot | C_1 \cap C_2)$ is well defined. Let $Y_2 = P(X | C_2)$ and let $Y_1 = P(Y_2 | C_1)$. We are going to show that $Y_1 = P(X | C_1 \cap C_2)$ if either one of the above two hypotheses is true.

Suppose C_2 is an affine space and $Y_1 \in C_2$. Then $(X-Y_2, Y_1-Z) = 0$ for every $Z \in C_2$ by the fact that $Y_2 = P(X | C_2)$, and $(Y_2-Y_1, Y_1-Z) \ge 0$ for every $Z \in C_1$ by the fact that $Y_1 = P(Y_2 | C_1)$. Hence, $(X-Y_1, Y_1-Z) \ge 0$ for every $Z \in C_1 \cap C_2$. By Theorem 2.3, $Y_1 = P(X | C_1 \cap C_2)$.

Suppose C_1 is an affine space and $C_1 \subset C_2$. Then there exists a closed linear space S in H such that $C_1 = Y_1 + S$. If we can show that $(X - Y_1, Z) = 0$ for every $Z \in S$, then $Y_1 = P(X | C_1) = P(X | C_1 \cap C_2)$. We claim that $Y_2 + S$ is a subset of C₂. By the convexity of C₂, $\lambda Y_2 + (1-\lambda)(Y_1+\alpha Z) \in C_2$ for each $\lambda \in (0,1)$, for each real α and for each $Z \in S$. Set $\alpha = (1-\lambda)^{-1}$. By letting $\lambda \rightarrow 1$ and using the fact that C₂ is closed, we find that $Y_2 + Z \in C_2$.

Since $Y_2 = P(X | C_2)$, $(X - Y_2, Y_2 - Y) \ge 0$ for each $Y \in C_2$. It follows that $(X - Y_2, Z) = 0$ for each $Z \in S$. Since $Y_1 = P(Y_2 | C_1)$ and C_1 is affine, $(Y_2 - Y_1, Z) = 0$ for all $Z \in S$. The last two equalities yield that $(X - Y_1, Z) = 0$ for each $Z \in S$. This completes the proof. []

<u>Corollary 2.4.1</u>. Let C_1 and C_2 be two closed convex sets with $C_1 \subset C_2$. If either C_1 or C_2 is an affine space, then (2.8) $P(P(\mathbf{X} | C_2) | C_1) = P(\mathbf{X} | C_1).$

The identity (2.8) is called the <u>smoothing property</u>. Most of the algorithms developed in Chapter III and Chapter V make use of this identity.

The important convergence theorems, Theorem 2.5, Corollary 2.5.1 and Theorem 2.6, are introduced by Brunk (1965) except that Theorem 2.5 is given here under a weaker hypothesis. Theorem 2.5 and Corollary 2.5.1 given below show that P(X|C) can be obtained as the limit of $\{P(X|C_n)\}$ or $\{P(Y_n|C_n)\}$ as the monotone sequence $\{C_n\}$ converges to C. <u>Theorem 2.5.</u> Let $\{C_n\}$ be a monotone sequence of closed convex sets in H, let $X \in H$ and let $X_n = P(X | C_n)$. If the sequence is monotone increasing, then $\lim_{n \to \infty} X_n$ exists and the limit is $P(X | C_\infty)$ where C_∞ is the closure of $\bigcup_{n=1}^{\infty} C_n$. If the sequence is monotone decreasing and $\bigcap_{n=1}^{\infty} C_n$ is nonempty, then $\lim_{n \to \infty} X_n$ exists and the limit is $P(X | \bigcap_{n=1}^{\infty} C_n)$.

A countable union of a monotone increasing sequence of convex sets is convex and the closure of a convex set is convex. A countable intersection of closed convex sets is closed and convex. Therefore, C_{∞} and $\bigcap_{n=1}^{\infty} C_n$ are closed and convex under their corresponding assumptions. If $\{C_n\}$ is monotone increasing then by (2.7), the sequence $\{\|X-X_n\|\}$ is monotone decreasing and $\{X_n\}$ is Cauchy. If $\{C_n\}$ is monotone decreasing and $\bigcap_{n=1}^{\infty} C_n \neq \phi$, then by (2.7), the sequence $\{\|X-X_n\|\}$ is monotone increasing, it is bounded from above by $\|X-Z\|$ for any $Z \in \bigcap_{n=1}^{\infty} C_n$ and $\{X_n\}$ is Cauchy.

<u>Corollary 2.5.1</u>. Additional to the assumptions in Theorem 2.5, let $\{Y_n\}$ be a convergent sequence with X as its limit. Then the sequence $\{P(Y_n | C_n)\}$ converges to $P(X | C_{\infty})$ or $P(X | \bigcap_{n=1}^{\infty} C_n)$ according as $\{C_n\}$ is monotone increasing or monotone decreasing.

A sequence $\{X_n, C_n\}$ is called a <u>martingale</u> if for each n,

 C_n is a closed convex set in H, $X_n \in H$, $C_n \subset C_{n+1}$ and $X_n = P(X_{n+k} | C_n)$ for any non-negative integer k. If for each n, C_n is a closed linear space in H and $X_n = P(X | C_n)$ for an element $X \in H$, then $\{X_n, C_n\}$ is a martingale provided that $\{C_n\}$ is a monotone increasing sequence. An example of a martingale which is not composed of closed linear spaces will be given in Corollary 4.18.1.

<u>Theorem 2.6.</u> Let $\{X_n, C_n\}$ be a martingale in H with the sequence $\{X_n\}$ being bounded. Then $\lim_{n \to \infty} X_n = X_{\infty}$ exists and $X_n = P(X_{\infty} | C_n)$ for each n. Consequently, every bounded martingale is of the form $\{P(X | C_n), C_n\}$ for some $X \in C_{\infty}$, where C_{∞} is the closure of $\bigcup_{n=1}^{\infty} C_n$.

A necessary and sufficient condition for the bounded sequence $\{X_n, C_n\}$ to be a martingale when the C_n 's are cones is given by Corollary 2.11.1.

II.2 On a Closed Convex Cone

A subset C of the Hilbert space H is said to be a <u>cone</u> if $X \in C$ implies $\delta X \in C$ for each $\delta \ge 0$. The family of isotonic functions described in Chapter III, the family of isotonic functions described in Chapter V, and the family of square-integrable Σ -measurable random variables described in Chapter IV are closed convex cones with respect to their corresponding Hilbert spaces. Throughout this section we shall let C denote a non-empty closed convex cone in the Hilbert space H. From (2.2) we have

(2.9)
$$P(\delta X | C) = \delta P(X | C)$$
 for each $\delta \ge 0$.

An immediate result from Theorem 2.3 is the following theorem.

<u>Theorem 2.7</u>. Let C be a closed convex cone in H and let $X \in H$. An element $X^* \in C$ is the projection of X on C if and only if it satisfies

$$(2.10) \qquad (X - X^*, X^*) = 0$$

and

$$(2.11) \qquad (X-X^*,Z) \leq 0 \quad \text{for each} \quad Z \in C.$$

The generalized projection $P(\cdot | C)$ is positive homogeneous in the sense of (2.9) and orthogonal in the sense of (2.10) provided that C is a closed convex cone.

Corollary 2.7.1. Let X^* be the projection of X on C. Then

(2.12)
$$||\mathbf{X}*|| = (\mathbf{X}, \mathbf{X}*) \le ||\mathbf{X}||^2$$

and if $\|\mathbf{X}^*\| = \|\mathbf{X}\|$, then $\mathbf{X}^* = \mathbf{X}$.

A geometric interpretation of (2.10) can be given as below. Let $B_X := \{Y \in H: (X-Y, Y) \ge 0\}$. Consider a simple case first. Let $H = R^2$ and let the inner product (\cdot, \cdot) be defined by $(X, Y) = X^t V Y$ for a given positive definite matrix V. Then B_X is the region bounded by the ellipse $Y^t V(X-Y) = 0$ which passes through X and the origin. Let C be a closed convex cone in R^2 and set $C_X := C \cap B_X$. Then for each $X \in C_X$, we have $\|Z\| \le \|P(X|C)\|$, $0 \le (X, Z) \le (X, P(X|C))$ and $(P(X|C), Z) \ge 0$. The above results also hold in general; the proof follows.

<u>Theorem 2.8.</u> Let C be a closed convex cone in H, let X ϵ H and let C_X := {Z ϵ C:(X-Z,Z) \geq 0}. Then C_X is closed and convex, and for each Z ϵ C_x we have

(2.13)
$$||Z||^2 \le (X,Z) \le (P(X|C),Z) \le (P(X|C),X) = ||P(X|C)||^2$$

and if the last inequality becomes an equality, then Z = P(X | C).

<u>Proof.</u> The set $B_X = \{Y \in H: (X - Y, Y) \ge 0\}$ is a closed ball with center at X/2 and with radius ||X||/2. Therefore, $C_X = C \cap B_X$ is closed and convex. For each $Z \in C_X$, the first inequality holds by the definition of Z, the second inequality is from (2.11) and the identity is from (2.12). Since $||Z||^2 \le (P(X|C), Z)$, by the Holder's inequality $(P(X|C), Z) \le ||P(X|C)|| ||Z||$, we have $||Z|| \le |P(X|C)||$ and hence the last inequality follows.

If any one of the identities $||Z||^2 = (P(X|C), X),$ (X, Z) = (P(X|C), X) and (P(X|C), Z) = (P(X|C), X) holds, then $(P(X|C), Z) = ||P(X|C)|| ||Z|| = ||P(X|C)||^2$ and hence Z = P(X|C). []

The theorem shows that if $Z \in C$ satisfies (2 10), then $||Z|| \leq ||P(X|C)||$; and the equality holds if and only if Z is the projection of X on C. This conclusion can also be made by observing that $||X||^2 = ||X-Z||^2 + ||Z||^2$.

Let C be a closed convex cone in H and let $C* := \{Y:(Y, Z) \leq 0 \text{ for each } Z \in C\}$. The set C* is a closed convex cone and $C* \cap C = \{0\}$. Such a set C* is called the <u>dual cone</u> of C. A simple and obvious result about the projection of X to C* is given below.

<u>Theorem 2.9.</u> Let C be a closed convex cone in H and let C* be the dual cone of C. For each $X \in H$, we have

(2.14)
$$P(X | C) + P(X | C^*) = X$$

Consequently, P(X | C) = 0 if and only if $(X, Z) \leq 0$ for each $Z \in C$.

If Y satisfies (2.11), i.e., $(X-Y,Z) \leq 0$ for each $Z \in C$, then X - Y $\in C^*$. Therefore, the collection $\{Y \in H: (X-Y,Z) \leq 0\}$ is simply $X - C^*$.

<u>Theorem 2.10</u>. Let C be a closed convex cone in H and let C* be the dual cone of C. For every $X \in H$, we have

(2.15)
$$\| P(X|C) \|^2 = (P(X|C), X) \le (P(X|C), Y) \le \| Y \|^2$$

for every $Y \in X - C^*$ and if $(P(X | C), Y) = ||Y||^2$, then Y = P(X | C). Further, if $Y \in C \cap (X - C^*)$, then $(X - Y, Y) \leq 0$.

<u>Proof.</u> If $Y \in X - C^*$, then $(X - Y, P(X | C)) \leq 0$ and hence the first inequality in (2.15) follows. By the Holder's inequality $(P(X | C), Y) \leq ||P(X | C)|| ||Y||$, we have $||P(X | C)|| \leq ||Y||$ and the second inequality follows. If $(P(X | C), Y) = ||Y||^2$, then $||Y||^2 = (P(X | C), Y) \leq ||P(X | C)|| ||Y||$. But we have just shown that $||P(X | C)|| \leq ||Y||$, so (P(X | C), Y) = ||P(X | C)|| ||Y|| and hence Y = P(X | C). If $Y \in C \cap (X - C^*)$ then $(X - Y, Y) \leq 0$ since $X - Y \in C^*$ and $Y \in C$. []

By Theorem 2.8, the closed convex set C_X is contained in the closed ball with center at origin and with radius ||P(X|C)||, while from Theorem 2.10 it follows that the closed convex set $X - C^*$ is contained in the half-space $\{Y:(P(X|C), Y) \ge ||P(X|C)||^2\}$. The intersection of C_X and $X - C^*$ is a singleton, $\{P(X|C)\}$.

<u>Theorem 2.11.</u> Let C_1 and C_2 be closed convex cones in H with $C_1 \subseteq C_2$ and let $\mathbf{X} \in \mathbf{H}$. Set $Y_i = P(\mathbf{X} | C_i)$, i = 1, 2 and set $Y_0 = P(Y_2 | C_1)$. Then

$$(2.16) ||Y_1|| \le ||Y_0|| \le ||Y_2|| \le ||X||,$$

and if any equality holds, then the corresponding two elements in H are identical.

<u>Proof.</u> Since $Y_2 = P(X | C_2)$ and $Y_0 = P(Y_2 | C_1)$, the results for the sequence $||Y_0|| \le ||Y_2|| \le ||X||$, follow from Corollary 2.7.1. Since $C_1 \subset C_2$, by (2.11), we have $(X - Y_2, Z) \le 0$ and $(Y_2 - Y_0, Z) \le 0$ for each $Z \in C_1$. Therefore, $(X - Y_0, Z) \le 0$ for each $Z \in C_1$ and the results follow immediately from Theorem 2.10. []

<u>Corollary 2.11.1</u>. Under the same assumptions as in Theorem 2.11, a necessary and sufficient condition for $P(P(X | C_2) | C_1) = P(X | C_1)$ is that $(P(X | C_2) - P(X | C_1), Z) \le 0$ for each $Z \in C_1$.

conclude our result by applying the above theorem. [

By the identity (2.12), the identity $\|X-P(X|C)\|^2 = \|X\|^2 - \|P(X|C)\|^2$ and the inequality (2.7), we may obtain that

(2.17)
$$\| P(X | C_2) - P(X | C_1) \|^2 \le \| P(X | C_2) \|^2 - \| P(X | C_1) \|^2$$

= $(P(X | C_2) - P(X | C_1), X)$

for each pair of closed convex cones C_1 and C_2 with $C_1 \subset C_2$. Theorem 2.7, Corollary 2.7.1, the identity (2.9) and the inequality (2.17) were introduced by Brunk (1965).

II.3 On a Finitely Generated Cone

Let Z_1, Z_2, \ldots, Z_m be a finite sequence of elements in the Hilbert space H and let C be the set of all non-negative linear combinations of the sequence. It has been shown that C is a closed convex cone (cf. Rockafellar (1970)). Such a closed convex cone is said to be <u>finitely generated</u>. And we denote it by $C[Z_1, \ldots, Z_m]$. Let $Z \in C[Z_1, \ldots, Z_m]$. Then there exists a set of non-negative real numbers a_1, \ldots, a_m such that $Z = \sum_{i=1}^m a_i Z_i$. The representation need not be unique and we may even have a representation which is composed of some non-negative coefficients and some negative coefficients. An element Y of a closed convex cone C is said to be an <u>extreme vector</u> of C if there do not exist linearly independent elements Y_1 and Y_2 in C such that $Y = Y_1 + Y_2$. If Y is an extreme vector of the cone $C[Z_1, \dots, Z_m]$, then Y must be a scalar multiple of Z_i for an i between 1 and m. If Z_1, \dots, Z_m are linearly independent, then they form a complete set of extreme vectors for $C[Z_1, \dots, Z_m]$. Let S be the smallest linear space in H which contains the finitely generated cone C. Then S is closed in H. By the smoothing property (2.8), P(P(X|S)|C) = P(X|C) for each $X \in H$. Since our interest here is to obtain P(X|C) and the projection P(X|S) can be obtained very easily, without loss of generality we may assume that H is the smallest linear space containing C. The following theorem is an immediate result of Theorem 2.7.

<u>Theorem 2.12.</u> Let C be the cone $C[Z_1, \ldots, Z_m]$ and let X ϵ H. An element X* ϵ C is the projection of X on C if and only if X* satisfies $(X-X^*, X^*) = 0$ and

(2.18) $(X-X*, Z_i) \leq 0$ for i = 1, 2, ..., m.

If $X \notin C[Z_1, \dots, Z_m]$, then by Theorem 2.2, $P(X \mid C[Z_1, \dots, Z_m])$ is a boundary point of $C[Z_1, \dots, Z_m]$. Boundary points of a finitely generated cone may not be easily identified. An element in a cone is said to be an <u>interior point</u> if it is not a boundary point. Let $Y = \sum_{i=1}^{m} a_i Z_i$ with a_1, a_2, \ldots, a_m positive. We claim that Y is an interior point. Without loss of generality, we may assume that Z_1, Z_2, \ldots, Z_n are linearly independent where $n \leq m$ and the dimension of H is n. Let $Z_0 = \sum_{i=n+1}^{m} a_i Z_i$. Then $Z_0 \in C[Z_1, \ldots, Z_m]$. It can be shown that there is an open ball B containing Y such that if $Z \in B$, then Z can be represented by $Z = Z_0 + \sum_{i=1}^{n} \beta_i Z_i$ where β_i, \ldots, β_n are positive. Therefore, the claim is established. Let S be a linear subspace of H and let C be a closed convex cone contained in S. <u>Relative boundary point</u> and <u>relative interior point</u> of C with respect to the linear space S are defined in the same way except that only elements of S are considered.

Let $C = C[Z_1, \ldots, Z_m]$ be a cone in H and let $X \in H$. Then there is a sequence of non-negative real numbers a_1, a_2, \ldots, a_m such that $P(X | C) = \sum_{i=1}^{m} a_i Z_i$. Let us define $\Lambda := \{i: a_i > 0\}$, $C_{\Lambda} := \{\sum_{i \in \Lambda} \beta_i Z_i: \beta_i \ge 0$ for each $i \in \Lambda\}$ and $S_{\Lambda} := \{\sum_{i \in \Lambda} \beta_i Z_i: \beta_i$ is real for each $i \in \Lambda\}$. Since $P(X | C) \in C_{\Lambda}$, we have $P(X | C) = P(X | C_{\Lambda})$. By the smoothing property, $P(X | C_{\Lambda}) = P(P(X | S_{\Lambda}) | C_{\Lambda})$. We claim that $P(X | S_{\Lambda}) \in C_{\Lambda}$ and hence $P(X | S_{\Lambda}) = P(X | C)$. If it were not true, i.e., $P(X | S_{\Lambda}) \notin C_{\Lambda}$, then by Theorem 2.2, $P(X | C_{\Lambda})$ would be a relative boundary point of C_{Λ} with respect to S_{Λ} . But on the other hand, $P(X | C_{\Lambda}) = P(X | C) = \sum_{i \in \Lambda} a_i Z_i$ with $a_i > 0$ for each $i \in \Lambda$, so $P(X | C_{\Lambda})$ must be a relative interior point of C_{Λ} with respect to S_{Λ} . Therefore, we have a contradiction, and thus the following theorem has been proved.

<u>Theorem 2.13.</u> Let C be the cone $C[Z_1, \ldots, Z_m]$ and let $X \in H$. Then there exists a linear space S generated by a subset of $\{Z_1, \ldots, Z_m\}$, such that P(X|C) = P(X|S).

The linear space S satisfying $P(\mathbf{X} | \mathbf{C}) = P(\mathbf{X} | \mathbf{S})$, is not unique. Since the representation of X by Z_1, \ldots, Z_m is not unique, the linear space S_{Λ} described above need not be unique. If we can identify a subset Γ of $\{1, 2, \ldots, m\}$ such that $P(\mathbf{X} | \mathbf{C}) \in C_{\Gamma}$, then $P(\mathbf{X} | \mathbf{S}_{\Gamma})$ is the solution we want, provided that $P(\mathbf{X} | \mathbf{S}_{\Gamma}) \in C_{\Gamma}$. Such an identification can be achieved for the problems of isotonic regression over a linearly ordered set, isotonic regression over a tree, and bivariate isotonic regression with $\rho \geq 0$. However, in these three cases, the closed convex cones are each generated by a set of linearly independent vectors.

<u>Corollary 2.13.1</u>. Let C be the cone $C[Z_1, \ldots, Z_m]$ and let $X \in H$. Let Λ and S_{Λ} be defined as above and let Γ be a subindex set of $\{1, 2, \ldots, m\}$. If $S_{\Gamma} \supset S_{\Lambda}$, then $P(X|S_{\Gamma}) = P(X|C)$ provided that $P(X|S_{\Gamma}) \in C_{\Gamma}$. <u>Proof.</u> If $S_{\Gamma} \supset S_{\Lambda}$, then by (2.17) we have $\|P(X|S_{\Gamma})\| \ge \|P(X|C)\|$. Recall that $(X - P(X|S_{\Gamma}), P(X|S_{\Gamma})) = 0$, so if $P(X|S_{\Gamma}) \in C_{\Gamma}$ then by Theorem 2.8 we have $P(X|S_{\Gamma}) = P(X|C)$. []

It is obvious that if $P(X|S_{\Gamma}) = P(X|C)$, then P(X|S) = P(X|C)for any linear space such that $C_{\Lambda} \subset S \subset S_{\Gamma}$. If $S_{\Gamma} \supset S_{\Lambda}$ and $P(X|S_{\Gamma}) \notin C_{\Gamma}$, then by the smoothing property, we have $P(X|C_{\Lambda}) \subset P(P(X|S_{\Gamma})|C_{\Lambda})$. Therefore, we may suppose that our Hilbert space at this stage is S_{Γ} . As the dimension of S_{Γ} goes down, we shall eventually obtain P(X|C) by successive projections on linear spaces and that is the process employed in Chapter III and Chapter V.

Suppose that P(X|C) is unknown and there is no way to identify a subindex set Γ such that either $P(X|C) \in C_{\Gamma}$ or $P(X|C) \in S_{\Gamma}$. Theorem 2.8 shows that P(X|C) has the largest norm among those $P(X|S_{\Gamma})$'s such that $P(X|S_{\Gamma}) \in C_{\Gamma}$. We state this formally as the following theorem.

<u>Theorem 2.14.</u> Let C be the cone $C[Z_1, \ldots, Z_m]$, let X ϵ H and let $S_X := \{P(X | S_{\Gamma}) : \Gamma \text{ is a subset of } \{1, \ldots, m\}$ and $P(X | S_{\Gamma}) \epsilon C\}$. If $P(X | S_{\Gamma})$ has the largest norm in S_X , then $P(X | S_{\Gamma}) = P(X | C)$. There are $2^{\mathbf{m}}$ possible linear spaces of the form S_{Γ} . $P(\mathbf{X} | \mathbf{C}) = P(\mathbf{X} | \mathbf{H})$ if and only if $\mathbf{X} \in \mathbf{C}$, and $P(\mathbf{X} | \mathbf{C}) = P(\mathbf{X} | \{0\})$ if and only if $(\mathbf{X}, \mathbf{Z}_i) \leq 0$ for $i = 1, 2, ..., \mathbf{m}$. A further reduction of candidates for $P(\mathbf{X} | \mathbf{C})$ of the form $P(\mathbf{X} | \mathbf{S}_{\Gamma})$ can be described as follows. $P(\mathbf{X} | \mathbf{S}_{\Gamma})$ satisfies (2.10). Let (P1) be the property that $P(\mathbf{X} | \mathbf{S}_{\Gamma}) \in \mathbf{C}$. If (P1) is true, then by Theorem 2.8, $\| P(\mathbf{X} | \mathbf{S}_{\Gamma}) \| \leq \| P(\mathbf{X} | \mathbf{C}) \|$. Let (P2) be the property that $(\mathbf{X} - P(\mathbf{X} | \mathbf{S}_{\Gamma}), \mathbf{Z}_i) \leq 0$ for $i = 1, ..., \mathbf{m}$. If (P2) is true, then by Theorem 2.10, $\| P(\mathbf{X} | \mathbf{S}_{\Gamma}) \| \geq \| P(\mathbf{X} | \mathbf{C}) \|$. A program to obtain $P(\mathbf{X} | \mathbf{C})$ can be made through calculation of $P(\mathbf{X} | \mathbf{S}_{\Gamma})$ by the following theorem.

<u>Theorem 2.15</u>. Let C be the cone $C[Z_1, \ldots, Z_m]$, let X \in H and let Γ be a subindex set. Then we have the following statements.

- (1) $P(\mathbf{X}|\mathbf{S}_{\Gamma}) = P(\mathbf{X}|\mathbf{C})$ if and only if $P(\mathbf{X}|\mathbf{S}_{\Gamma})$ satisfies (P1) and (P2).
- (2) If (P1) holds but (P2) fails, then for any Δ such that $\|P(X|S_{\Delta})\| \leq \|P(X|S_{\Gamma})\|$, we have $P(X|S_{\Delta}) \neq P(X|C)$. In particular, $P(X|S_{\Delta}) \neq P(X|C)$ for each $\Delta \subset \Gamma$.
- (3) If (P1) fails but (P2) holds, then for any Δ such that $\|P(X|S_{\Delta})\| \ge \|P(X|S_{\Gamma})\|$, we have $P(X|S_{\Delta}) \ne P(X|C)$. In particular, $P(X|S_{\Lambda}) \ne P(S|C)$ for each $\Delta \supset \Gamma$.
<u>Proof.</u> The projection $P(X|S_{\Gamma})$ satisfies (2.10), i.e., $(X-P(X|S_{\Gamma}), P(X|S_{\Gamma})) = 0$. Statement (1) follows from Theorem 2.12. If (P1) holds but (P2) fails, then $P(X|S_{\Gamma}) \in C_X$ and $P(X|S_{\Gamma}) \neq P(X|C)$. Statement (2) follows from Theorem 2.8. If (P2) holds but (P1) fails, then $P(X|S_{\Gamma}) \in X - C^*$ and $P(X|S_{\Gamma}) \neq P(X|C)$. Statement (3) follows from Theorem 2.10. []

Let C be the cone generated by a linearly independent sequence Z_1, \ldots, Z_n of elements in H and let the dimension of H be n. For each i, there exists an element $Y_i \in H$ such that $(Y_i, Z_i) < 0$ and $(Y_i, Z_j) = 0$ for each $j \neq i$. The sequence Y_1, \ldots, Y_n is linearly independent and it is unique up to scalar multiplication. The cone $C[Y_1, \ldots, Y_n]$ is the dual cone of C.

<u>Theorem 2.16</u>. Let C be the cone $C[Z_1, \ldots, Z_n]$ and let $C^* = C[Y_1, \ldots, Y_n]$ be the dual cone of C. For every $X \in H$, there exists a unique representation $X = \sum_{i \in \Lambda} a_i Z_i + \sum_{j \in \Lambda^C} \beta_j Y_j$ for some index set Λ such that $a_i \ge 0$ for each $i \in \Lambda$ and $\beta_j \ge 0$ for each $j \in \Lambda^C$. Furthermore $P(X | C) = \sum_{i \in \Lambda} a_i Z_i$ and $P(X | C^*) = \sum_{j \in \Lambda^C} \beta_j Y_j$.

<u>Proof</u>. Let P(X|C) be represented by $P(X|C) = \sum_{i=1}^{n} a_i Z_i$, where $a_i \ge 0$ for each i. Let $\Lambda := \{i: a_i > 0\}$. Then $P(X|C) = P(X|S_{\Lambda})$ and hence $(X-P(X|C), Z_i) = 0$ for each $i \in \Lambda$. Let $P(X | C^*)$ be represented by $P(X | C^*) = \sum_{j=1}^{n} \beta_j Y_j$, where $\beta_j \ge 0$ for each j. By Theorem 2.9, $P(X | C^*) = X - P(X | C)$ and hence $(P(X | C^*), Z_i) = \sum_{j=1}^{n} \beta_j (Y_j, Z_i) = \beta_i (Y_i, Z_i) = 0$ for each $i \in \Lambda$. It follows that $\beta_i = 0$ for each $i \in \Lambda$ and therefore, $X = \sum_{i \in \Lambda} a_i Z_i + \sum_{j \in \Lambda^c} \beta_j Y_j$.

If X has a representation $\sum_{i \in \Lambda} a_i Z_i + \sum_{j \in \Lambda^c} \beta_j Y_j$ with a_i 's and β_j 's non-negative, then it is trivial that $P(X \mid C) = \sum_{i \in \Lambda} a_i Z_i$ and $P(X \mid C^*) = \sum_{j \in \Lambda^c} \beta_j Y_j$. Since $\{Z_1, \ldots, Z_n\}$ and $\{Y_1, \ldots, Y_n\}$ are each linearly independent, the representations of $P(X \mid C)$ by Z_1, \ldots, Z_n and of $P(X \mid C^*)$ by Y_1, \ldots, Y_n are unique. For each i, the product of the corresponding coefficients a_i and β_i must be zero. The uniqueness of the above representation for X must therefore be satisfied. []

Let Λ be an index set. Theorem 2.9 shows that $P(X \mid C) = P(X \mid S_{\Lambda})$ if and only if $P(X \mid C^*) = P(X \mid S_{\Lambda}^{\perp})$, where $S_{\Lambda}^{\perp} = \{Y:(Y, Z) = 0 \text{ for each } Z \in S_{\Lambda}\}$ is the orthogonal complement of S_{Λ} . Under the hypothesis of Theorem 2.16, $S_{\Lambda}^{\perp} = \{\Sigma_{j \in \Lambda C} \beta_{j} Y_{j}: \beta_{j} \text{ real}\}$. The property that $a_{i}\beta_{i} = 0$ for each i reminds us that $\beta_{1}, \ldots, \beta_{n}$ are Lagrangian multipliers. The inequalities $(Y_{j}, P(X \mid C)) \leq 0$ for $j = 1, \ldots, n$ are the constraints. Therefore, the theorem is equivalent to the Kuhn-Tucker condition.

III. ISOTONIC REGRESSION

III.1 Preliminaries

Isotonic regression problems discussed in this chapter are those defined on some quasi-ordered finite sets. We shall leave the general case to the next chapter. The problems are standard mathematical programming problems. Their objective functions are the weighted sums of squares described in (3.1). Their feasible regions are the intersections of some closed half-spaces which are each determined by a pair of arguments.

The binary relation \leq defined on a set Ω is said to be a partial ordering if

- (1) it is reflexive; $\omega \leq \omega$ for each ω in Ω
- (2) it is transitive; $\omega, \mu, \nu \in \Omega$, $\omega \leq \mu$ and $\mu \leq \nu$ imply $\omega \leq \nu$

and (3) it is antisymmetric; $\omega, \mu \in \Omega$, $\omega \leq \mu$ and $\mu \leq \omega$ imply $\omega = \mu$.

A <u>quasi-ordering</u> is reflexive and transitive but not necessarily antisymmetric. A pair of elements μ and ν in Ω is said to be <u>comparable</u> if either $\mu \leq \nu$ or $\nu \leq \mu$. A <u>linear ordering</u> is a partial ordering such that each pair of elements is comparable.

Let \leq be a quasi-ordering on Ω and let Γ be a subset of Ω . An element $\omega \in \Gamma$ is <u>maximal</u> in Γ if $\mu \in \Gamma$ and $\omega \leq \mu$ imply $\mu \leq \omega$; $\omega \in \Gamma$ is minimal in Γ if $\mu \in \Gamma$ and $\mu \leq \omega$ imply $\omega \leq \mu$. Γ is said to be <u>bounded from above</u> if there exists an element $\mu \in \Omega$ such that $\omega \leq \mu$ for each ω in Γ ; such an element μ is called an <u>upper bound</u> of Γ . If the set Λ of all upper bounds of Γ is non-empty and it has a unique minimal element, then such an element is said to be the <u>least upper bound</u> of Γ , and is denoted by $\nabla \Gamma$. Similarly for the definitions of "<u>bounded from</u> <u>below</u>", "<u>lower bound</u>" and the "<u>greatest lower bound</u>". The greatest lower bound of Γ is denoted by $\Lambda \Gamma$.

A partial ordering is said to have a <u>tree structure</u> if every non-comparable pair has a greatest lower bound but does not have an upper bound. A set Ω with an ordering \leq is said to be an <u>ordered set</u>; such a pair is denoted by (Ω, \leq) . The pair (Ω, \leq_t) is called a <u>tree</u> if \leq_t is a partial ordering having a tree structure. Some problems involving trees have been given by Thompson (1962) and Hartigan (1967). A finite tree, by its definition, has a unique minimal element. A simple tree is a tree such that every element is either maximal or minimal. A tree with a unique maximal element is a linearly ordered set. From the definitions, a linearly ordered set is a tree, a tree is a partially ordered set and a partially ordered set is a quasi-ordered set.

Let $(\Omega, \leq q)$ be a quasi-ordered finite set. The finiteness of Ω is our assumption throughout this chapter. Let $\Gamma_1, \ldots, \Gamma_k$ be subsets of Γ . They are <u>unrelated</u> if every pair of elements from different subsets is not comparable. Let Λ be a subset of Ω and let $\mu \in \Lambda$. An element $\nu \in \Lambda$ can be <u>reached</u> from μ in Λ if there exist $\omega_1, \omega_2, \ldots, \omega_k$ in Λ such that $\omega_i \leq \omega_{i+1}$ or $\omega_i \geq \omega_{i+1}$ for $i = 0, 1, \ldots, k$ where $\omega_0 = \mu$ and $\omega_{k+1} = \nu$. Let $\Lambda(\mu) := \{\nu \in \Lambda: \nu \text{ can be reached from } \mu \text{ in } \Lambda\}$. It is obvious that if $\nu \in \Lambda(\mu)$, then $\Lambda(\nu) = \Lambda(\mu)$; and if $\nu \in \Lambda$ and $\nu \notin \Lambda(\mu)$, then $\Lambda(\nu)$ and $\Lambda(\mu)$ are unrelated. Λ is said to be <u>connected</u> if $\Lambda(\mu) = \Lambda$ for some $\mu \in \Lambda$. Λ is said to be <u>separable</u> if it is not connected. If Λ is separable, then there exist $\omega_1, \ldots, \omega_k$ in Λ such that $\Lambda = \bigcup_{i=1}^k \Lambda(\omega_i)$ and $\Lambda(\omega_1), \ldots, \Lambda(\omega_k)$ are unrelated. Each subset $\Lambda(\omega_i)$ is called a <u>component</u> of Λ . Linearly ordered sets and trees are connected.

Let \leq_1 and \leq_2 be two quasi-orderings on Ω . \leq_1 is said to be the <u>reversal</u> of \leq_2 provided that for μ, ν in Ω $\mu \leq_1 \nu$ if and only if $\nu \leq_2 \mu$. The reversal of a quasi-ordering \leq is denoted by \leq_r . The quasi-ordered set (Ω, \leq_r) is called the <u>reversal</u> of (Ω, \leq) . When there is no ambiguity, we shall use Ω to denote a quasi-ordered set and Ω_r to denote its reversal. The quasi-ordered set Ω is connected or separable if and only if its reversal is connected or separable. The reversal of a tree is called a reversed tree. A pair of elements μ and ν in Ω is said to be an <u>immediately comparable pair</u> if $\mu \leq \nu$ and there does not exist an element ω other than μ and ν such that $\mu \leq \omega$ and $\omega \leq \nu$. We denote such a pair by $[\mu \leq \nu]$. If $[\mu \leq \nu]$ is an immediately comparable pair, then μ is called an <u>immediate predecessor</u> of ν , and ν is called an <u>immediate successor</u> of μ . For a partially ordered set Ω , the ordering on each component can be described by listing all immediately comparable pairs. Let Γ be a component of Ω with m elements. If Γ is a linearly ordered set or a tree, then there are m-1 immediately comparable pairs in Γ ; and m-1 is the smallest number of immediately comparable pairs that a connected partially ordered set Γ may have.

A real-valued function Z defined on a quasi-ordered set is <u>isotonic</u> if $\mu, \nu \in \Omega$ and $\mu \leq \nu$ imply $Z(\mu) \leq Z(\nu)$. Isotonic functions are constant over each <u>equivalence class</u> $[\omega] := \{\mu; \mu \leq \omega, \omega \leq \mu\}$. Let $M(\Omega)$ be the collection of all isotonic functions defined on Ω . It is obvious that $M(\Omega_r) = -M(\Omega)$. When there is no ambiguity, we shall use M instead of $M(\Omega)$. A <u>weight</u> <u>function</u> W is a non-negative function defined on Ω . The pair (Ω, W) , where Ω is a quasi-ordered set and W is a weight function, will furnish the structure for the isotonic regression problems described below. Let X be a given real-valued function defined on Ω . An <u>isotonic regression</u> of X over (Ω, W) is an element in M which minimizes

(3.1)
$$f(Z) := \sum_{\omega \in \Omega} [X(\omega) - Z(\omega)]^2 W(\omega)$$

among all functions Z in M. If Ω is separable, then the minimization problem can be studied in each component of Ω independently. If in addition, each component is an equivalence class $[\omega]$, then it is a regression problem. Let Γ be a subset of Ω and let $X|\Gamma$ and $W|\Gamma$ be the restrictions of X and of W to Γ . Then Γ is by itself a quasi-ordered set. An isotonic regression of $X|\Gamma$ over $(\Gamma, W|\Gamma)$ is called a <u>restricted isotonic</u> <u>regression</u> of X to Γ .

Let $\Omega_0 = \{\mu \in \Omega : W(\mu) > 0\}$. If Ω_0 is an empty set, the problem is trivial. Suppose Ω_0 is non-empty and X_0^* is a restricted isotonic regression of X to Ω_0 . The function X^* defined by $X^*(\omega) = \min\{X_0^*(\mu): \mu \in \Omega_0\}$ if $\{\mu \in \Omega_0: \omega \ge \mu\} = \phi$ and $X^*(\omega) = \max\{X_0^*(\mu): \mu \le \omega, \mu \in \Omega_0\}$ otherwise, is an isotonic regression of X over (Ω, W) . Since isotonic functions are constant on each equivalence class, the objective function can be written as

$$f(Z) = \sum_{\omega \in \Omega_0} [X(\omega) - Z(\omega)]^2 W(\omega)$$

=
$$\sum_{\omega \in \Omega_0} [\overline{X}(\omega) - Z(\omega)]^2 \overline{W}(\omega) + \sum_{\omega \in \Omega_0} [X(\omega) - \overline{X}(\omega)]^2 W(\omega)$$

where $\overline{\mathbf{X}}(\omega)$ is the weighted average of X over the equivalence class $[\omega]$ in Ω_0 and $\overline{\mathbf{W}}(\omega)$ is the average of W over the same equivalence class $[\omega]$. The second term on the right-hand side of the above equation is independent of Z and $[\overline{\mathbf{X}}(\omega) - \mathbf{Z}(\omega)]^2 \overline{\mathbf{W}}(\omega)$ is constant on each equivalence class. Therefore, without loss of generality, we may assume that Ω is a connected partially ordered set and W is a positive weight function.

Let H be the linear space of all real-valued functions defined on the connected partially ordered set Ω , let W be a positive weight function on Ω and let (\cdot, \cdot) be a bilinear functional defined on H x H by

$$(X, Y) = \sum_{\omega \in \Omega} X(\omega) Y(\omega) W(\omega)$$
 for each pair X, Y \in H.

The bilinear functional (\cdot, \cdot) is an inner product on H, and the linear space H with the inner product described above is a Hilbert space. The objective function f(Z) can be represented as $||X-Z||^2$ where $||Y|| = (Y, Y)^{1/2}$.

For each immediately comparable pair $[\mu \leq \nu]$, let us define a linear functional $g_{\mu,\nu}$ on H by

$$g_{\mu, \nu}(Z) = Z(\mu) - Z(\nu).$$

The set $\{Y:g_{\mu,\nu}(Y) \leq 0\}$ is a closed half-space containing 0 and hence is a closed convex cone. Since an element $Z \in H$ is isotonic if and only if $g_{\mu,\nu}(Z) \leq 0$ for each immediately comparable pair, the family M of all isotonic functions can be characterized as the intersection of all such closed half-spaces, i.e.,

 $M = \{Z \in H: g_{\mu, \nu}(Z) \leq 0 \text{ for each } [\mu \leq \nu]\}$. Therefore, the family M is a closed convex cone and the isotonic regression is the generalized projection to M. The uniqueness and existence of the isotonic regression of a given X follows from Theorem 2.1.

Let Z be a non-constant isotonic function and let c be a constant function. Then c and Z-c are linearly independent and Z = c + Z - c. Since M contains constant functions, Z is not an extreme vector of M and hence the only extreme vectors of M are constant functions, positive or negative. Let M* be the dual cone of M, i.e., $Y \in M^*$ if $(Y,Z) \leq 0$ for each $Z \in M$. For each $[\mu \leq \nu]$, let $Y_{\mu,\nu}$ be the function in H defined by $Y_{\mu,\nu}(\mu) = W(\mu)^{-1}, Y_{\mu,\nu}(\nu) = -W(\nu)^{-1}$ and $Y_{\mu,\nu}(\omega) = 0$ elsewhere. Since $(Y_{\mu,\nu}, Z) = g_{\mu,\nu}(Z), Y_{\mu,\nu} \in M^*$ for each $[\mu \leq \nu]$. The dual cone of the closed half-space $\{Y: g_{\mu, \nu}(Y) \leq 0\}$ is the ray $\{\delta Y_{\mu,\nu}: \delta \ge 0\}$. The <u>convex hull</u> of a given set A is the intersection of all convex sets containing A, and is denoted by conh(A). Let C_1, C_2, \ldots, C_k be a finite sequence of closed convex cones in H. It is known that $(C_1 \cap C_2 \cap \cdots \cap C_k)^* = \operatorname{conh}(C_1^* \cup C_2^* \cup \cdots \cup C_k^*)$ (cf. Rockafeller (1970)). Therefore,

 $M^* = \operatorname{conh} \cup \{\delta Y_{\mu,\nu} : [\mu \leq \nu] \text{ an immediately comparable pair}\}.$ In

other words, M* is the cone generated by the $Y_{\mu,\nu}$'s, $[\mu \leq \nu]$ an immediately comparable pair.

If Y is an extreme vector of M^* , then Y is a scalar multiple of $Y_{\mu,\nu}$ for some $[\mu \leq \nu]$. We shall show that $Y_{\mu,\nu}$ is an extreme vector of M* for each $[\mu \leq \nu]$. If Z \in M*, then $(\mathbf{Z}, \mathbf{1}_{\omega}) \leq 0$ and $(\mathbf{Z}, \mathbf{1}_{\omega}^{0}) \leq 0$ for each $\omega \in \Omega$ because $\mathbf{1}_{\omega}$ and l_{ω}^{0} are in M where $l_{\omega}(\nu) = 1$ if $\omega \leq \nu$ and $l_{\omega}(\nu) = 0$ otherwise; $l_{\omega}^{0}(\nu) = 1$ if $\omega \leq \nu$, $\omega \neq \nu$ and $l_{\omega}^{0}(\nu) = 0$ otherwise. It follows that M* does not contain any line {aZ:a real}. Suppose that there is an immediately comparable pair $[\xi \leq \eta]$ such that $Y_{\xi, \eta}$ is not an extreme vector of M*. Then the cone generated by Y 's is M* because M* does not contain a line and it μ, ν otheris the cone generated by the Y 's (cf. Rockafeller (1970)). By μ, ν 's the identities $(C_1 \cap C_2 \cap \cdots \cap C_{k-1})^* = \operatorname{conh}(C_1^* \cup C_2^* \cup \cdots \cup C_{k-1}^*)$ $M = (M^*)^*$, it will then follow that M can be represented as and an intersection with one less closed half-space $\{Y:g_{\xi,\eta}(Y) \leq 0\}$. Therefore, $g_{\xi,\eta}$ is a redundant constraint. On the other hand, let Z be defined by $Z(\omega) = 1$ if $\omega \ge \xi$, $\omega \ne \eta$ with zeros elsewhere. Then Z satisfied all other constraints but Z is not isotonic. This contradicts that $g_{\xi,\eta}$ is redundant. It follows that $Y_{\mu,\nu}$ is an extreme vector of M^* for each $[\mu \leq \nu]$.

Let S_0 be the linear space $\{Y:(Y,1)=0\}$. Since $(Y_{\mu,\nu},1)=0$ for each $[\mu \leq \nu]$, M* is a subset of S_0 .

III. 2 Upper Sets and Indicators

A subset U of a quasi-orderes set Ω is said to be an upper set if $\mu \in U$ and $\nu \geq \mu$ imply $\nu \in U$. A subset L of Ω is said to be a lower set if $\nu \in L$ and $\mu \leq \nu$ imply $\mu \in L$. It is obvious that if U is an upper set then U^{C} is a lower set and if L is a lower set than L^c is an upper set. A non-empty connected upper set is called a basic set. If an upper set is separable, then each of its components is an upper set and hence a basic set. For each $\omega \in \Omega$, the set $U(\omega) := \{\mu: \mu \ge \omega\}$ is a basic set. Such a basic set $U(\omega)$ is said to be determined by ω . Ω and ϕ are upper sets and lower sets. Ω is separable if and only if there exists a proper subset of Ω which is both an upper set and a lower set. Further discussion of upper sets and lower sets will be given in Section IV. 1. In the remainder of the chapter, we shall let Ω denote a connected partially ordered finite set. The intersection of an upper set and a lower set is called a level set. A subset Γ of Ω is a level set if and only if $\mu, \nu \in \Gamma$ and $\mu \leq \omega \leq \nu$ imply that $\omega \in \Gamma$. A component of a level set is a level set. The intersection of level sets is a level set.

Let Γ be a subset of Ω . The <u>indicator</u> of Γ , denoted by ${}^{1}\Gamma$, is a real-valued function which assumes the value one at each element in Γ and zero elsewhere. Indicators of Ω and ϕ are denoted by 1 and 0 respectively. The indicator ${}^{1}\Gamma$ is isotonic

if and only if Γ is an upper set. The indicator of a basic set is called a <u>basic function</u>. For a basic set $U(\omega)$ determined by ω , its basic function is denoted by 1_{ω} . If Z is an isotonic function, then for each real α , $[Z \ge \alpha] := \{\omega: Z(\omega) \ge \alpha\}$ is an upper set. Suppose Z assumes the values $\alpha_1 \le \alpha_2 \le \cdots \le \alpha_k$. Then $[Z = \alpha_i] := \{\omega: Z(\omega) = \alpha_i\}$ is a level set for each i. Let $\beta_i = \alpha_i - \alpha_{i-1}, \quad i = 1, \dots, k$ where $\alpha_0 = 0$. Then $Z = \sum_{i=1}^k \beta_i 1_i [Z \ge \alpha_i]$. Since an indicator of an upper set is the sum of the indicators of its components, every non-negative isotonic function is a non-negative linear combination of basic functions and every isotonic function is a linear combination of basic functions such that the coefficient associated with each indicator other than 1 is non-negative. But such a representation need not be unique.

Let M_{+} be the family of all non-negative isotonic functions. Then M_{+} is the cone generated by the basic functions. Therefore M_{+} has finitely many extreme vectors. Let U be a basic set and let l_{U} be it indicator. We clain that l_{U} is an extreme vector of M_{+} . Suppose there are Z_{1} and Z_{2} in M_{+} such that $l_{U} = Z_{1} + Z_{2}$. Then $Z_{1}(\omega) = Z_{2}(\omega) = 0$ for all $\omega \notin U$. Let μ and ν be two elements in U with $\mu \leq \nu$. Then $Z_{i}(\mu) \leq Z_{i}(\nu)$, i = 1, 2 and $Z_{1}(\mu) + Z_{2}(\mu) = Z_{1}(\nu) + Z_{2}(\nu) = 1$. Therefore $Z_{i}(\mu) = Z_{i}(\nu)$, i = 1, 2. If $\omega \in U$ can be reached from μ in U, then it follows that $Z_i(\omega) = Z_i(\mu)$, i = 1, 2. Since a basic set is a non-empty connected upper set, every element $\omega \in U$ can be reached from μ and hence $Z_1 = \lambda I_U$ and $Z_2 = (1-\lambda)I_U$ for some λ between zero and one. Our claim therefore has been established and basic functions form a complete set of extreme vectors for M_+ . If Ω is a tree, then every upper set with more than one minimal element is separable. It follows that basic functions must be of the form I_{ω} , and M_+ has exactly the same number of extreme vectors as the number of elements in Ω . Thus we have proved the following theorem.

<u>Theorem 3.1.</u> Let Ω be a connected partially ordered finite set, let M be the family of isotonic functions and let M_+ be the family of non-negative isotonic functions. Then M_+ is the cone generated by the basic functions, and basic functions form a complete set of extreme vectors for M_+ . The family M is the cone generated by basic functions and -1.

Following the same procedure, one can show that the above theorem also holds for general quasi-ordered finite sets connected or separable.

The isotonic regression of X over Ω is the projection P(X | M) of X on the closed convex cone M. In the finite case, L₂(Σ), introduced in Chapter IV is the same as M and the conditional expectation $E(X \mid \Sigma)$, also introduced in Chapter IV, is the same as $P(X \mid M)$. Therefore, properties of isotonic regression can be found in Chapter II in the form of $P(X \mid M)$ and also can be found in Chapter IV in the form of either $P(X \mid L_2(\Sigma))$ or $E(X \mid \Sigma)$. The main interest in the remaining part of this chapter is in algorithms for various isotonic regression problems and some results which lead to these algorithms.

<u>Theorem 3.2.</u> Let $X^* = \Sigma_{U \text{ basic}} a_U^1 U$ with $a_U \ge 0$ for each basic set U different from Ω , and let $\Lambda = \{U: U \text{ is basic}, a_U > 0\} \cup \{\Omega\}$. Then X^* is the isotonic regression of X if and only if $(X-X^*, 1_U) = 0$ for each $U \in \Lambda$ and $(X-X^*, 1_U) \le 0$ for each basic set U.

<u>Proof.</u> The function X^* given above is obviously an isotonic function. Let $S_{\Lambda} = \{Z = \Sigma_{U \in \Lambda} \beta_U l_U : \beta_U \text{ real}\}$. The condition $(X - X^*, l_U) = 0$ for each $U \in \Lambda$ is equivalent to $X^* = P(X | S_{\Lambda})$. Therefore, the result follows from Theorem 2.12 and Theorem 2.13. The only difference is that we have $\Omega \in \Lambda$ no matter what value a_{Ω} is. This gap can be filled by considering (4.2), which can be interpreted as $(X - X^*, 1) = 0$ in our present situation. []

<u>Corollary 3.2.1</u>. Let $X^* \in M$. Then X^* is the isotonic regression of X if and only if X^* satisfies

(3.2)
$$\sum_{\omega \in [X^{*}=\gamma]} X(\omega) W(\omega) = \sum_{\omega \in [X^{*}=\gamma]} X^{*}(\omega) W(\omega) \text{ for each } \gamma$$

and

(3.3)
$$\sum_{\omega \in U} X(\omega) W(\omega) \leq \sum_{\omega \in U} X^{*}(\omega) W(\omega)$$

for each upper set U.

<u>Proof.</u> Let X* be an isotonic function which assumes values $\gamma_1 < \gamma_2 < \ldots < \gamma_k$ and let $a_i = \gamma_i - \gamma_{i-1}$, $i = 1, \ldots, k$ where $\gamma_0 = 0$. Then $X* = \sum_{i=1}^k a_i l_{[X* \ge \gamma_i]}$ with a_i positive for i > 1. Let X* satisfy (3.2) and (3.3). Then

$$(\mathbf{X}-\mathbf{X}^*,\mathbf{X}^*) = \sum_{i=1}^{k} \gamma_i \sum_{\omega \in [\mathbf{X}^*=\gamma_i]} (\mathbf{X}(\omega)-\mathbf{X}^*(\omega)) \mathbf{W}(\omega) = 0$$

and $(X-X^*, -1) = 0$. Since M is the cone generated by upper sets and -1, by Theorem 2.12, X* is the isotonic regression of X.

On the other hand, let X^* be the isotonic regression of X. Since each upper set is the disjoint union of its components, by Theorem 3.2 we have $(X-X^*, 1_U) \leq 0$ for each upper set U. Similarly, $(X-X^*, 1_{[X^* \geq \gamma_i]}) = 0$ for $i = 1, \ldots, k$. For each γ different from $\gamma_1, \ldots, \gamma_k$, X* satisfies (3.2) automatically. For each i,

$$(X-X^{*}, 1_{[X^{*}=\gamma_{i}]}) = (X-X^{*}, 1_{[X^{*}\geq\gamma_{i}]}) - (X-X^{*}, 1_{[X^{*}>\gamma_{i}]}) = 0.$$

This completes the proof.

<u>Corollary 3.2.2.</u> Let X* be the isotonic regression of X. If ν is minimal in $[\alpha \leq X^* \leq \beta]$, then $X^*(\nu) \leq X(\nu)$. If μ is maximal in $[\alpha \leq X^* \leq \beta]$, then $X^*(\mu) \geq X(\mu)$.

<u>Proof.</u> If ν is minimal in $[a \le X^* \le \beta]$, then ν is minimal in $[a \le X^*]$. By the minimality of ν in $[a \le X^*]$, U = { $\omega: X^*(\omega) \ge a, \ \omega \ne \nu$ } is an upper set. From Corollary 3.2.1,

$$(X - X^*, 1[\alpha \le X^*]) = 0$$

and

$$(\mathbf{X} - \mathbf{X}^*, \mathbf{1}_{\mathbf{U}}) \leq \mathbf{0},$$

which implies $X(v) - X^*(v) \ge 0$. Similarly we have the second statement.

<u>Corollary 3.2.3.</u> Let X* be the isotonic regression of X, let $[\mu \leq \nu]$ be an immediately comparable pair and let $X(\mu) \geq X(\nu)$. Then either there exists an immediate successor ω of μ such that $X*(\omega) = X*(\mu)$ or there exists an immediate predecessor τ of ν such that $X*(\tau) = X*(\nu)$.

<u>Proof.</u> Suppose on the contrary that neither $X^*(\omega) = X^*(\mu)$ for any immediate successor ω of μ nor $X^*(\tau) = X^*(\nu)$ for any immediate predecessor τ of ν . Then $X^*(\nu) > X^*(\mu)$, ν is minimal in $[X^* \ge X^*(\nu)]$ and μ is maximal in $[X^* \le X^*(\mu)]$. By Corollary 3.2.2, we have $X(\nu) \ge X^*(\nu) \ge X^*(\mu) \ge X(\mu)$. This contradicts that $X(\mu) \ge X(\nu)$. []

<u>Theorem 3.3.</u> Let Ω be a partially ordered set, let X* be the isotonic regression of X and let μ be an immediate predecessor of ν . If any one of the following three statements is true:

- (1) μ is the only immediate predecessor of ν , ν is the only immediate successor of μ , and $X(\mu) \ge X(\nu)$,
- (2) μ is the only immediate predecessor of ν and $X(\nu) \leq X(\omega)$ for every $\omega \geq \mu$,
- (3) ν is the only immediate successor of μ and $X(\mu) \ge X(\omega)$ for every $\omega \le \nu$,

then $X^*(\mu) = X^*(\nu)$.

The theorem is an extension of Theorem 2.6 given by Barlow and coworkers (1972). That $X^*(\mu) = X^*(\nu)$ follows from the first statement is an immediate result of Corollary 3.2.3. If Ω is a linearly ordered set, then each element has at most one immediate predecessor and at most one immediate successor. Thereofre, statment (1) is enough to provide an algorithm in such a situation. If Ω is a tree, then each element has at most one immediate predecessor, and if Ω is a reversed tree, then each element has at most one immediate successor. Statement (2) and statement (3) will be sufficient for obtaining isotonic regressions in such situations. Statement (2) and statement (3) are symmetric with respect to reversal, because $M(\Omega) = -M(\Omega_r)$ and $P(X|M(\Omega)) = -P(-X|M(\Omega_r))$.

<u>Proof.</u> Only the result following from statement (2) will be considered. Let μ be the only immediate predecessor of ν and $X(\nu) \leq X(\omega)$ for every $\omega \geq \mu$. Suppose $X^*(\mu) < X^*(\nu)$. Then ν is minimal in $[X^*(\nu) \leq X^*]$. By Corollary 3.2.2, $X^*(\nu) \leq X(\nu)$. On the other hand, $U(\mu) = \{\omega: \omega \geq \mu\}$ is an upper set and $X(\omega) \geq X(\nu)$ for each $\omega \in U(\mu)$. By Corollary 4.15.2 which does not depend on Theorem 3.3, $X^*(\mu) \geq X(\nu)$ and hence $X^*(\mu) \geq X(\nu) \geq X^*(\nu)$. This contradicts the assumption $X^*(\mu) < X^*(\nu)$.

III. 3 On a Linearly Ordered Set

The minimization problem discussed here is to minimize f(Z)subject to $z_i \leq z_{i+1}$ for i = 1, 2, ..., n-1 where $Z = (z_1, ..., z_n)$ and

$$f(Z) = \sum_{i=1}^{n} (x_i - z_i)^2 w_i$$

with given $X = (x_1, \ldots, x_n)$ and given $w_i > 0$, $i = 1, \ldots, n$. The problem is known as an isotonic regression over the linearly ordered set $\{1, 2, \ldots, n\}$. It can be solved by the Pool-Adjacent-Violators algorithm which was introduced by Ayer and coworkers (1955). The algorithm, which is an immediate result of Theorem 3.3, will be presented later in this section.

Let $M = \{Z: z_i \leq z_{i+1}, i = 1, 2, ..., n-1\}$ be the feasible region, let M^* be the dual cone of M with respect to the inner product $(X, Y) = \sum_{i=1}^{n} x_i y_i w_i$ and let $Y_1, Y_2, ..., Y_{n-1}$ be n-component vectors such that for each k, $Y_k = (y_{k1}, ..., y_{kn})$ is defined by $y_{kk} = w_k^{-1}$, $y_{k,k+1} = -w_{k+1}^{-1}$ and $y_{ki} = 0$ if $i \neq k, k+1$. The dual cone M^* is the cone generated by the linearly independent vectors $Y_1, ..., Y_{n-1}$, and it is a subset of the linear space $S_0 = \{Z: \sum_{i=1}^{n} z_i w_i = 0\}$. The inner products of Y_i 's among themselves are $(Y_{j-1}, Y_j) = -w_j^{-1}$, $(Y_j, Y_j) = w_j^{-1} + w_{j+1}^{-1}$, $(Y_{j+1}, Y_j) = -w_{j+1}^{-1}$ and $(Y_i, Y_j) = 0$ if |i-j| > 1.

The isotonic regression of X, P(X|M), is X - P(X|M*) as given by Theorem 2.9. By the smoothing propery, $P(X|M*) = P(P(X|S_0)|M*)$. The projection of X on S₀ is

$$P(X|S_0) = X - \overline{x}$$
 where $\overline{x} = \sum_{i=1}^{n} x_i w_i / \sum_{i=1}^{n} w_i$

Let $X_0 = X - \overline{x}$. Since S_0 is the linear space generated by Y_1, \dots, Y_{n-1} , there exists a unique set of real numbers a_1, \dots, a_{n-1} such that

$$\mathbf{X}_{0} = \boldsymbol{\Sigma}_{i=1}^{n-1} \boldsymbol{\alpha}_{i} \boldsymbol{Y}_{i}.$$

The projection $P(X_0 | M^*)$ is in M*, so there exists a unique set

of non-negative real numbers a_1^*, \ldots, a_{n-1}^* such that

$$P(X_0 | M^*) = \sum_{i=1}^{n-1} \alpha_i^* Y_i$$
.

Let $\Lambda = \{i: a_i^* > 0\}$ and let $N_{\Lambda} := \{Y = \sum_{i \in \Lambda} \beta_i Y_i : \beta_i \ge 0\}$. Then $P(X_0 | M^*) = P(X_0 | N_{\Lambda})$. We shall soon see that $P(X_0 | M^*)$ can be obtained by successive projections on linear spaces, each of which is generated by a subset of $\{Y_1, \ldots, Y_{n-1}\}$.

Let Γ be a subindex set of $\{1, 2, \ldots, n-1\}$. The linear space generated by $\{Y_i : i \in \Gamma\}$, is denoted by T_{Γ} . If $j \in \Gamma$, the projection of Y_j on T_{Γ} , $P(Y_j | T_{\Gamma})$, is Y_j itself. For fixed i, j, k with i < j < k, let $\Gamma_1 = \{i, i+1, \ldots, j-1, j+1, \ldots, k\}$, let $\Gamma_2 = \{1, 2, \ldots, i-2, k+2, \ldots, n-1\}$ and let $\Gamma_3 = \Gamma_1 \cup \Gamma_2$. The linear spaces T_{Γ_1} and T_{Γ_2} are orthogonal and Y_j is orthogonal to T_{Γ_2} . It follows that $P(Y_j | T_{\Gamma_3}) = P(Y_j | T_{\Gamma_1})$.

If Γ is such that $\Gamma_1 \subset \Gamma \subset \Gamma_3$, then by the smoothing property we have $P(Y_j | T_{\Gamma}) = P(P(Y_j | T_{\Gamma_3}) | T_{\Gamma}) = P(P(Y_j | T_{\Gamma_1}) | T_{\Gamma})$ $= P(Y_j | T_{\Gamma_1})$. Let

(3.4)
$$a_q = \sum_{m=i}^{q} w_m / \sum_{m=i}^{j} w_m, \quad b_q = \sum_{m=q+1}^{k+1} w_m / \sum_{m=j+1}^{k+1} w_m$$

and let $Y = -\sum_{q=i}^{j-1} a_q Y_q - \sum_{q=j+1}^{k} b_q Y_q$. Then for each $h = i, i+1, \dots, j-2$, we have

$$(Y_{j}-Y, Y_{h}) = -(Y, Y_{h})$$

$$= a_{h-1}(Y_{h-1}, Y_{h}) + a_{h}(Y_{h}, Y_{h}) + a_{h+1}(Y_{h+1}, Y_{h})$$

$$= -a_{h-1}w_{h}^{-1} + a_{h}w_{h}^{-1} + a_{h}w_{h+1}^{-1} - a_{h+1}w_{h+1}^{-1}$$

$$= w_{h}^{-1}(a_{h}-a_{h-1}) - w_{h+1}^{-1}(a_{h+1}-a_{h})$$

$$= 0$$

and similarly for each h = j+2, ..., k, we have $(Y_j - Y, Y_h) = 0$ by changing the a's to b's. For h = j-1, we have

$$(Y_{j}-Y, Y_{j-1}) = (Y_{j}, Y_{j-1}) + a_{j-2}(Y_{j-2}, Y_{j-1}) + a_{j-1}(Y_{j-1}, Y_{j-1})$$
$$= -w_{j}^{-1} - a_{j-2}w_{j-1}^{-1} + a_{j-1}(w_{j-1}^{-1}+w_{j}^{-1})$$
$$= w_{j-1}^{-1}(a_{j-1}-a_{j-2}) - w_{j}^{-1}(1-a_{j-1})$$
$$= 0$$

and similarly, we have $(Y_j - Y, Y_{j+1}) = 0$. Therefore, Y is the projection $P(Y_j | T_{\Gamma_1})$ and

(3.5)
$$P(Y_j | T_{\Gamma}) = -\Sigma_{q=i}^{j-1} a_q Y_q - \Sigma_{q=j+1}^k b_q Y_q$$
.

If $\Gamma_1 = \{j+1, \ldots, k\}$, $\Gamma_3 = \Gamma_1 \cup \{1, \ldots, j-2, k+2, \ldots, n-1\}$ and Γ is such that $\Gamma_1 \subset \Gamma \subset \Gamma_3$, then following the same procedure we have

$$(3.6) \qquad P(Y_j | T_{\Gamma}) = -\Sigma_{q=j+1}^k b_q Y_q .$$

Similarly, if $\Gamma_1 = \{i, i+1, \dots, j-1\}, \quad \Gamma_3 = \Gamma_1 \cup \{1, \dots, i-2, j+2, \dots, n-1\}$ and Γ is such that $\Gamma_1 \subset \Gamma \subset \Gamma_3$, then

(3.7)
$$P(Y_j | T_{\Gamma}) = -\Sigma_{q=i}^{j-1} a_q Y_q$$

If j-1, j and j+1 are not in Γ , then $P(Y_j | T_{\Gamma}) = 0$. Let $Z = \sum_{j=1}^{n-1} \beta_j Y_j$. The projection of Z on T_{Γ} , for some subindex set Γ , is $P(Z | T_{\Gamma}) = \sum_{j=1}^{n-1} \beta_j P(Y_j | T_{\Gamma})$. If $j \in \Gamma$, $P(Y_j | T_{\Gamma}) = Y_j$. Let $j \notin \Gamma$. If $j-1 \in \Gamma$, let i be the smallest index such that $i, i+1, \ldots, j-1$ are in Γ and if $j+1 \in \Gamma$, let k be the largest index such that $j+1, j+2, \ldots, k$ are in Γ . If j-1and j+1 are not in Γ , we have $P(Y_j | T_{\Gamma}) = 0$. We identify Γ_1 as $\{i, i+1, \ldots, j-1\}$, $\{j+1, \ldots, k\}$ or their union according as only $j-1 \in \Gamma$, only $j+1 \in \Gamma$ or both j-1 and j+1 in Γ . Therefore, the projection of Z on T_{Γ} can be calculated by (3.5), (3.6) and (3.7).

Let $P(Z | T_{\Gamma}) := \Sigma_{h \in \Gamma} \beta_{h \wedge h}^{*} Y_{h}$. For each $h \in \Gamma$, let $p = \max\{m:m < h, m \notin \Gamma\}$ and let $q = \min\{m:m > h, m \notin \Gamma\}$. Suppose p and q exist. Then for each $j \neq h$ which is less than p, greater than q or belonging to Γ , we have $P(Y_{j} | T_{\Gamma}) = \Sigma_{m \in \Gamma} \gamma_{m} Y_{m}$ with $\gamma_{h} = 0$. It follows that

(3.8)
$$\beta_h^* = \beta_h - (\beta_p b_n + \beta_q a_h)$$

where $a_h = \sum_{m=p+1}^h w_m / \sum_{m=p+1}^q w_m$ and $b_h = \sum_{m=h+1}^q w_m / \sum_{m=p+1}^q w_m$. The a_h and the b_h are the same as in (3.4); for the a_h , we identify i and j as p+1 and q respectively; for the b_h , we identify j and k as p and q-1 respectively. If p does not exist, set p = 0 and $\beta_p = 0$, and if q does not exist, set q = n and $\beta_q = 0$. Then β_h^* is still given by (3.8).

<u>Theorem 3.4.</u> Let $X_0 = \sum_{i=1}^{n-1} a_i Y_i$ and let $P(X_0 | M^*) = \sum_{i=1}^{n-1} a_i^* Y_i$. If $a_h = \min\{a_1, \dots, a_{n-1}\}$ and $a_h < 0$, then $a_h^* = 0$.

<u>Proof.</u> Let $\Lambda = \{i: a_i^* > 0\}$. Theorem 2.13 shows that $P(X_0 | M^*) = P(X_0 | T_\Lambda)$. Suppose $h \in \Lambda$. By (3.8), $a_h^* = a_h - (a_p b_h + a_q a_h)$. Since $a_p \ge a_h$, $b_h > 0$, $a_q \ge a_h$, $a_h > 0$ and $b_h + a_h = 1$, we have $a_h^* \le 0$. This contradicts that $h \in \Lambda$. []

The <u>Projection of Minimum Violators</u> is an algorithm to obtain the isotonic regression of X over the linearly ordered set $\{1, 2, ..., n\}$ through the dual cone M*. By a <u>violator</u> we mean an index j such that $a_j \leq 0$. Step 1. Set $\overline{x} = \sum_{i=1}^{n} x_i w_i / \sum_{i=1}^{n} w_i$.

Set $a_1 = w_1(x_1 - \overline{x}), a_i = a_{i-1} + w_i(x_i - \overline{x})$ for i = 2, ..., n-1. Set $\Gamma = \{1, ..., n-1\}.$

If a_1, \ldots, a_{n-1} are non-negative, go to Step 4; Step 2. otherwise, choose j such that $a_{j} = \min\{a_{1}, \ldots, a_{n-1}\}$. Set $\Gamma = \Gamma - \{j\}$. If $j-1 \in \Gamma$, then find the smallest i such that i, i+1,..., j-1 $\in \Gamma$ and set $\gamma_m = 0$ for m = 1, ..., i-1and set $\gamma_m = \Sigma_{h=i}^m w_h / \Sigma_{h=i}^j w_h$ for m = i, i+1, ..., j-1;otherwise, set $\gamma_m = 0$ for $m = 1, \dots, j-1$. If $j+1 \in \Gamma$, then find the largest k such that $j+1, \ldots, k \in \Gamma$ and set $\gamma_m = \Sigma_{h=m+1}^{k+1} w_h / \Sigma_{h=j+1}^{k+1} w_h$ for $m = j+1, \dots, k$ set $\gamma_m = 0$ for $m = k+1, \dots, n-1$; otherwise, set $\gamma_m = 0$ for m = j+1, ..., n-1. Set $\gamma_i = 1$. Set $a = a - a_j \gamma_m$ for $m = 1, \ldots, n-1$. Step 3. Go to Step 2.

Step 4. Set
$$y_1 = w_1^{-1} a_1$$
, $y_i = w_i^{-1} (a_i - a_{i-1})$ for $i = 2, ..., n-1$,
 $y_n = -w_n^{-1} a_{n-1}$.
Set $x_i^* = x_i - y_i$ for $i = 1, 2, ..., n$.
End.

When we have more than one j such that

 $a_j = \min\{a_1, \ldots, a_{n-1}\}$, we may compute γ_m 's simultaneously for such a set of j's. The algorithm is not efficient as evidenced by Example 3.1; yet, it is the method developed through the dual cone. The justification of the algorithm can be described as follows. Let

 $X_0 = \sum_{i=1}^{n-1} a_i Y_i$, where $X_0 = P(X | S_0)$. If a_1, \dots, a_{n-1} are non-negative, then $X_0 = P(X_0 | M^*) = P(X | M^*)$. In any case, let $\Lambda_0 := \{i: a_i \ge 0 \text{ or } a_i > a_i \text{ for some } j \neq i\}.$ Theorem 3.4 shows that $\Lambda_0 \supset \Lambda$, where $\Lambda = \{i: a_i^* > 0\}$ and $P(X_0 | M^*) = \sum_{i=1}^{n-1} a_i^* Y_i$. Since $P(X_0 | M^*) = P(X_0 | N_\Lambda)$, by the smoothing property $P(X_0 | M^*) = P(P(X_0 | T_{\Lambda_0}) | N_{\Lambda})$ Let $X_1 := P(X_0 | T_{\Lambda_0})$ be represented by $\sum_{i \in \Lambda_{n}} a_{1i}Y_{i}$ and let $\Lambda_{1} := \{i \in \Lambda_{0}; a_{1i} \ge 0 \text{ or } a_{1i} > a_{1i}\}$ for some $j \neq i$. If $a_{1i} \geq 0$ for each $i \in \Lambda_0$, then $\Lambda_1 = \Lambda_0$ and $X_1 \in N_{\Lambda}$. Since $P(X_0 | M^*) = P(X_0 | N_{\Lambda})$, the smoothing property shows that $P(X_0 | N_{\Lambda_1}) = P(P(X_0 | T_{\Lambda_0}) | N_{\Lambda_1}) = P(X_1 | N_{\Lambda_1})$. It follows that $X_1 = P(X_0 | M^*)$. If $a_{1i} < 0$ for some $i \in \Lambda_0$, then $\Lambda_0 \neq \Lambda_1 \supset \Lambda$. Since $P(X_0 | M^*) = P(X_1 | N_{\Lambda_1})$, the set Λ stays the same, and $P(X_0 | M^*) = P(P(X_1 | T_{\Lambda_1}) | N_{\Lambda})$. Therefore, applying the above procedure inductively, we have a strictly decreasing sequence $\Lambda_0 \supset \Lambda_1 \supset \Lambda_2 \supset \ldots \supset \Lambda$. The sequence will terminate at some index k such that $\Lambda_k = \Lambda_{k+1}$. At that stage, $a_{ki} \ge 0$ for each $i \in \Lambda_k$, $X_k \in N_{\Lambda_k}$ and $X_k = P(X_0 | M^*)$. The index set Λ_k need not be the same as Λ , and we may reach the situation $\Lambda_k = \Lambda = \phi$ which indicates $X_k = 0$ and $N_{\Lambda_L} = T_{\Lambda_L} = \{0\}$. The situation occurs only when $X \in M$. Once $P(X_0 | M^*)$ is obtained, $P(X | M) = X - P(X | M*) = X - P(X_0 | M*).$

A similar technique can be applied to M also. For each i, let 1; be the vector such that the ith,...,nth entries have values one with zeros elsewhere. The vectors l_1, l_2, \ldots, l_n are linearly independent, and every vector in M is a linear combination of these indicators such that the coefficients are non-negative except the one corresponding to l_1 .

<u>Theorem 3.5.</u> Let $X = \sum_{i=1}^{n} a_{i} l_{i}$, let the isotonic regression of X, X*, be represented by $X* = \sum_{i=1}^{n} a_{i}^{*} l_{i}$ where $a_{i}^{*} \ge 0$ for i > 1 and let $\Lambda = \{i: a_{i}^{*} > 0 \text{ or } i = 1\}$. If $a_{j} \le 0$ for some j > 1, then $j \notin \Lambda$.

The theorem is essentially the same as Theorem 2.1 of Ayer and coworkers (1955), and it is a consequence of Theorem 3.2, Corollary 3.2.2 or Theorem 3.3. Suppose there is an index j > 1such that $a_j \leq 0$ and $a_j^* > 0$. Let $X = (x_1, \dots, x_n)$ and let $X^* = (x_1^*, \dots, x_n^*)$. Then $a_j = x_j - x_{j-1}$ and $a_j^* = x_j^* - x_{j-1}^*$ for j > 1. Since $a_j^* > 0$, j is minimal in $[X^* \geq x_j^*]$ and j - 1 is maximal in $[X^* \leq x_{j-1}^*]$, by Corollary 3.2.2 we have $x_j \geq x_j^*$ and $x_{j-1}^* \geq x_{j-1}$. This contradicts that $x_j^* > x_{j-1}^*$ and $x_j \leq x_{j-1}$. For each subindex set Γ of $\{1, 2, \dots, n\}$, let S_{Γ} denote the linear space generated by $\{1_i : i \in \Gamma\}$ and let $M_{\Gamma} = S_{\Gamma} \cap M$. Let $\Lambda_0 = \{i : a_i > 0 \text{ or } i = 1\}$. Then $\Lambda_0 \supset \Lambda$, $P(P(X|S_{\Lambda_0})|M_{\Lambda}) = P(X|M)$ and $P(X|S_{\Lambda_0}) = \sum_{i=1}^n a_i P(1_i|S_{\Lambda})$. If one computes $P(1_i|S_{\Lambda_0})$ for each i and then does a succession of projections on certain linear spaces S_{Γ} with $\Gamma \supset \Lambda$ in a manner similar to the Projection of Minimum Violators algorithm, one obtains the <u>Projection of Violators</u> algorithm. However, there is a better method to compute projections on the linear spaces S_{Γ} .

Let Γ be a subindex set of $\{1, 2, ..., n\}$. Suppose p and q are two consecutive indices in Γ . If $Y = (y_1, ..., y_n)$ is the projection of $Z = (z_1, ..., z_n)$ on S_{Γ} , then $(Z - Y, 1_j) = 0$ for each $j \in \Gamma$ and $y_p = y_{p+1} = ... = y_{q-1}$. Therefore, $(Z - Y, 1_q - 1_p) = 0$ or equivalently

$$\sum_{i=p}^{q-1} (z_i - y_i) w_i = 0$$

and hence

$$y_p = \Sigma_{i=p}^{q-1} z_i w_i / \Sigma_{i=p}^{q-1} w_i.$$

This modified version of the Projection of Violators algorithm is known as the <u>Pool-Adjacent-Violators</u> algorithm, which was introduced by Ayer and coworkers (1955). Let $X = (x_1, x_2, ..., x_n)$ be a vector of n entries.

Step 1. If $x_1 \le x_2 \le \cdots \le x_n$, stop; otherwise, divide the sequence into several monotone decreasing sequences from left to right such that each subsequence is as long as possible, i.e.,

$$x_{1} \ge x_{2} \ge \cdots \ge x_{i_{1}-1}, x_{i_{1}} \ge x_{i_{1}+1} \ge \cdots \ge x_{i_{2}-1}, \cdots, x_{i_{k}} \ge x_{i_{k}+1} \ge \cdots \ge x_{n} \quad \text{with} \quad x_{i_{j}-1} \le x_{i_{j}} \quad \text{for each } j = 1, \dots, k.$$

Set $\Gamma = \{1, i_{1}, i_{2}, \dots, i_{k}, n+1\}.$
Step 2. Set $x_{h} = \sum_{i=p}^{q-1} x_{i} w_{i} / \sum_{i=p}^{q-1} w_{i} \quad \text{for } h = 1, \dots, n \quad \text{where } p$
and q are two consecutive indices in Γ such that
 $p \le h \le q.$
Go to Step 1.

The justification of the algorithm is similar to that of the Projection of Minimum Violators algorithm. Let X, X* and Λ be defined as in Theorem 3.5. Let $\Lambda_0 = \{i: a_i > 0 \text{ or } i = 1\}$. The condition $a_i > 0$ is the same as $x_{i-1} < x_i$. Theorem 3.5 shows that $\Lambda_0 \supset \Lambda$. If $x_1 \leq x_2 \leq \cdots \leq x_n$, then $X \in M$ and $X^* = X$. Otherwise, $X * = P(X | M) = P(X | M_{\Lambda})$. By the smoothing property, we have $P(X | M_{\Lambda}) = P(P(X | S_{\Lambda_0}) | M_{\Lambda})$. Let $X_1 := P(X | S_{\Lambda_0})$ be represented by $X_1 = \sum_{i \in \Lambda_0} a_{1i} a_i$ and let $\Lambda_1 = \{i \in \Lambda_0: a_{1i} > 0 \text{ or } i_i > 0 \}$ i = 1. If $x_{11} \le x_{12} \le \cdots \le x_{1n}$, then $X_1 \in M$ and $X^* = X_1$ as indicated by Corollary 2.13.1. Otherwise, $\Lambda_0 \neq \Lambda_1 \supset \Lambda$. Applying the above procedure inductively, we have a strictly decreasing sequence $\Lambda_0 \supset \Lambda_1 \supset \Lambda_2 \supset \ldots \supset \Lambda_k \supset \Lambda$. The sequence will terminate at some index k such that $x_{k1} \leq x_{k2} \leq \cdots \leq x_{kn}$. The situation that $\Lambda_k = \{1\}$ is the case when $X - \overline{x} \in M^*$ where $\overline{\mathbf{x}} = \Sigma_{i=1}^{n} \mathbf{x}_{i} \mathbf{w}_{i} / \Sigma_{i=1}^{n} \mathbf{w}_{i}.$

The similarity of the Projection of Minimum Violators algorithm and the Pool-Adjacent-Violators algorithm can be observed from Theorem 2.16 by considering M^* and $M_0 := M \cap S_0$. The closed convex cone M_0 is the dual cone of M^* with respect to the linear space S_0 . There are n-1 extreme vectors in M_0 each of which is determined by $P(1_i | S_0)$ for some i = 2, ..., n. It is obvious that for each i, $(Y_i, 1_j) = (Y_i, P(1_j | S_0)) = 0$ if $j \neq i+1$ and $(Y_i, 1_{i+1}) = (Y_i, P(1_{i+1} | S_0)) = -1$.

<u>Example 3.1</u>. Let X = (3, 2, 7, 8, 5) and let

 $w_1 = w_2 = w_3 = w_4 = w_5 = 10$. The isotonic regression of X will be the same if we let $w_1 = w_2 = w_3 = w_4 = w_5 = 1$. By the Projection of Minimum Violators algorithm, we have $\bar{x} = 5$ and X - $\bar{x} = (-2, -3, 2, 3, 0)$.

Step 1. $a_1 = -2$, $a_2 = -5$, $a_3 = -3$, $a_4 = 0$ $\Gamma = \{1, 2, 3, 4\}$ Step 2. j = 2, $\Gamma = \{1, 3, 4\}$, i = 1, k = 4 $\gamma_1 = 1/2$, $\gamma_2 = 1$, $\gamma_3 = 2/3$, $\gamma_4 = 1/3$ Step 3. $a_1 = 1/2$, $a_2 = 0$, $a_3 = 1/3$, $a_4 = 5/3$ Now, a_1 , a_2 , a_3 and a_4 are non-negative Step 4. $y_1 = 1/2$, $y_2 = -1/2$, $y_3 = 1/3$, $y_4 = 4/3$, $y_5 = -5/3$ $x_1^* = 2\frac{1}{2}$, $x_2^* = 2\frac{1}{2}$, $x_3^* = 6\frac{2}{3}$, $x_4^* = 6\frac{2}{3}$, $x_5^* = 6\frac{2}{3}$ (Y = P(X | M*), X* = P(X | M)) By the Pool-Adjacent-Violators algorithm, we have

Step 1. {3,2}, {7}, {8,5} $\Gamma = \{1,3,4,6\}$ Step 2. $X_1 = (2\frac{1}{2}, 2\frac{1}{2}, 7, 6\frac{1}{2}, 6\frac{1}{2})$ Step 1. { $2\frac{1}{2}, 2\frac{1}{2}$ }, {7, $6\frac{1}{2}, 6\frac{1}{2}$ } $\Gamma = \{1,3,6\}$ Step 2. $X_2 = (2\frac{1}{2}, 2\frac{1}{2}, 6\frac{2}{3}, 6\frac{2}{3}, 6\frac{2}{3})$ $X_2 = P(X | M)$. []

III.4 On<u>a Tree</u>

A partially ordered set is a tree if each non-comparable pair has a greatest lower bound but does not have an upper bound. A finite tree has a unique minimal element r, called the <u>root</u>, and each element other than r has exactly one immediate predecessor. In a partially ordered set, if $\mu \leq \nu$, then there exists a <u>chain</u> $\omega_0 \leq \omega_1 \leq \cdots \leq \omega_{j+1}$ such that $[\omega_i \leq \omega_{i+1}]$ is an immediately comparable pair, $i = 0, 1, \ldots, j$, where $\mu = \omega_0$ and $\nu = \omega_{j+1}$. The chain connecting μ and ν need not be unique. The intersection of chains need not be a chain. Let Ω be a finite tree, let m_1, m_2, \ldots, m_k be its maximal elements and let $Ch(m_1), Ch(m_2), \ldots, Ch(m_k)$ be the chains which connect r and m_i , $i = 1, \ldots, k$. The chain $Ch(m_i)$ connecting r and m_i is unique. The intersection $Ch(m_i) \cap Ch(m_j)$ is also a chain which connects r and $m_i \wedge m_j$. The union $\bigcup_{i=1}^k Ch(m_i)$ is simply Ω itself. The basic sets are upper sets with a unique minimal element; they are denoted by $U(\omega)$ with ω the minimal element in the set. The indicator of $U(\omega)$ is denoted by l_{ω} . An example of a tree will be given below.

Let Y be a function defined on Ω , let a(r) = Y(r) and for each ν other than r, let $a(\nu) = Y(\nu) - Y(\mu)$ where μ is the immediate predecessor of ν . Let ω be an element other than r and let $Ch(\omega) = \{r, \omega_1, \omega_2, \dots, \omega_i, \omega\}$ be the chain connecting r and $\omega. \text{ The sum } \sum_{\nu \in Ch(\omega)} a(\nu) \text{ is } Y(\omega). \text{ Let } Z = \sum_{\omega \in \Omega} a(\omega) l_{\omega}. \text{ Then}$ $Z(\omega) = \sum_{\nu < \omega} \alpha(\nu)$. The set $\{\nu:\nu \leq \omega\}$ is the chain $Ch(\omega)$. Therefore, we have $Y = \sum_{\omega \in \Omega} \alpha(\omega) \mathbf{1}_{\omega}$. Since every function Y on Ω has a unique linear representation by $\{1_{\omega} : \omega \in \Omega\}$, it follows that $\{1_{\omega} \colon \omega \in \Omega\}$ is linearly independent. If Z is isotonic, then Zis a linear combination of $\{1_{\omega}: \omega \in \Omega\}$ such that the coefficients are non-negative except possibly the one corresponding to l_r . The coefficient of l_{ν} is $Z(\nu) - Z(\mu)$ where μ is the immediate predecessor of ν provided that ν differs from r. If Ω has elements, then there are n-l constraints (immediately comparable pairs) and M_{+} has n extreme vectors $\{1_{\omega}: \omega \in \Omega\}$.

The isotonic regression over a tree can be obtained very easily as we shall soon see in Example 3.2. Thompson (1962) introduced the

Minimum Violator algorithm. Let X be a given function defined on and let X^* be the isotonic regression of X. If ν is an Ω immediate successor of μ and $X(\nu) \leq X(\mu)$, then ν is said to be a violator. Among all the violators, if X(v) attains the minimum value, then ν is called a <u>minimum violator</u>. Let ν be a minimum violator and let μ be its immediate predecessor. Then by Theorem 3.3 $X^{*}(\mu) = X^{*}(\nu)$. By the smoothing property, we may group μ and ν as a single element μ whose X value is the weighted average of those at μ and ν and whose weight is the sum of the weights at μ and ν . Now, we have n-1 elements in $\Omega - \{v\}$. The ordering on $\Omega - \{v\}$ is determined by all immediately comparable pairs in Ω except that we take out the pair $[\mu \leq \nu]$ and change all the pairs $[\nu \leq \omega]$ into $[\mu \leq \omega]$. Applying this procedure inductively, there will eventually be no violator. Let Ω_0 be the set of remaining elements and let X_0^* be the function on Ω_0 at this stage. Define the function X^* on Ω by $X^{*}(\omega) = \max\{X_{0}^{*}(\mu): \mu \leq \omega, \mu \in \Omega_{0}\}$. Then X^{*} is the isotonic regression of X.

A modified version of the Minimum Violator algorithm is the <u>Minimum Upper Set</u> algorithm which involves an improvement in the method of grouping. Let ν be a violator and let μ be its immediate predecessor. A violator ν is said to be <u>pivotal</u> if either $X(\nu) = \min\{X(\omega): \omega \ge \mu\}$ or ν is the only immediate successor of μ . By Theorem 3.3 a pivotal element and its immediate predecessor will have the same X^* value. Instead of using minimum violators, we use pivotal elements in the Minimum Upper Set algorithm.

<u>Example 3.2.</u> Let $\Omega = \{(1,4), (2,3), (2,4), (3,1), (3,2), (3,4), (4,3), (4,4), (5,4)\}$ be a tree with its ordering being specified by the five chains $\{(3,1), (3,2), (2,3), (1,4)\}$, $\{(3,1), (3,2), (2,3), (2,4)\}$, $\{(3,1), (3,2), (2,3), (3,4)\}$, $\{(3,1), (3,2), (4,3), (4,4)\}$ and $\{(3,1), (3,2), (4,3), (5,4)\}$, let the weight function W have the value one at each element in Ω and let X be the function given as follows.



Let us say the pivotal element ν is type I, if ν is the only immediate successor of its immediate predecessor and is type II otherwise. The grouping of a pivotal element ν and its immediate predecessor μ into a single element μ is not necessary as long as we keep the pair $[\mu \leq \nu]$ having the same function value. The pair $[\mu \leq \nu]$ will be connected by double line segments to indicate their special relation.

The violators are (1,4), (3,2), (4,3), (4,4) and (5,4). The element (3,2) is a type I pivotal element and elements (1,4) and (4,4) are type II pivotal elements. The first iteration yields the function X_1 given below:



The element (4,3) is the only violator. Therefore, the average is taken over $\{(3,1), (3,2), (4,3), (4,4)\}$ and the second iteration yields the function X_2 .



A new violator occurs at (5,4). We take the average over the set $\{(3,1), (3,2), (4,3), (4,4), (5,4)\}$. We see that the function X_3 is isotonic and hence it is the isotonic regression of X.



If Ω is a reversed tree, then its reversal Ω_r is a tree. One may use the reversal technique based on the relation, $P(X|M(\Omega)) = -P(-X|M(\Omega_r))$, to obtain the isotonic regression. The procedure is that we change the ordering from less than to greater than and replace $X(\omega)$ by $-X(\omega)$ for each $\omega \in \Omega$. After we have the isotonic regression Y of the latter problem, we define $X^*(\omega) = -Y(\omega)$ for each $\omega \in \Omega$ and X^* is the isotonic regression of X over Ω . For convenience, we may use the <u>Maximum Lower</u> <u>Set</u> algorithm. The method is the same as the Minimum Upper Set algorithm except that pivotal elements are μ 's such that either $X(\mu) = \max\{X(\omega): \omega \leq \nu\}$ or μ is the only immediate predecessor of ν , where ν is the immediate successor of μ . In this situation, we group $[\mu \leq \nu]$ into a single element ν . Applications of the Maximum Lower Set algorithm will be found in Example 3.3 and Example 3.5.

III. 5 On a Partially Ordered Set

The structure of a general connected partially ordered finite set is very complicated and hence algorithms for isotonic regression of this type will not be easy to develop. Alexander (1970) introduced a method which rewrites a partial ordering into a consistent linear ordering. The method is much more complicated than the Pool-Adjacent-Violators algorithm and the Minimum Upper Set algorithm. In the remainder of this section we will try to reduce the minimization problem into several small problems, whenever it is possible, such that the combination of the solutions to the small problems is the desired solution. In a smaller problem, Alexander's algorithm will be easier to apply.

Let Ω be a connected partially ordered finite set, let W be a given positive weight function, let X be a given function and let Γ be a subset of Ω . If Z is a function defined on Ω , we denote

$$\mathbf{f}(\mathbf{Z};\Gamma) := \sum_{\omega \in \Gamma} [\mathbf{X}(\omega) - \mathbf{Z}(\omega)]^2 \mathbf{W}(\omega).$$

Let $\Gamma_1, \Gamma_2, \dots, \Gamma_k$ be a <u>partition</u> of Ω , i.e., $\Gamma_1, \Gamma_2, \dots, \Gamma_k$ are mutually disjoint and $\bigcup_{i=1}^k \Gamma_i = \Omega$, let Y_1, Y_2, \dots, Y_k be the
restricted isotonic regression of X to $\Gamma_1, \Gamma_2, \ldots, \Gamma_k$ respectively and let Y be defined by $Y(\omega) = Y_i(\omega)$ if $\omega \in \Gamma_i$. For any isotonic function Z, $f(Z; \Gamma_i) \ge f(Y; \Gamma_i)$, $i = 1, 2, \ldots, k$ and hence $f(Z) \ge f(Y)$. Therefore, if Y is isotonic, it is the isotonic regression of X.

Let X* be the isotonic regression of X, let $[a \leq X* \leq \beta]$ be non-empty, let C be a component of $[a \leq X* \leq \beta]$ and let Y be the restriction of X* to C, X*|C. Let U_c be an <u>upper set</u> <u>in C</u>, i.e., $\mu \in U_c$, $\nu \in C$ and $\mu \leq \nu$ imply $\nu \in U_c$. For any U_c, define $U = U_c \cup [X* > \beta]$. Then U is an upper set. By (3.2) and (3.3) we have

$$\sum_{\omega \in U_{c}} \mathbf{X}(\omega) \mathbf{W}(\omega) \leq \sum_{\omega \in U_{c}} \mathbf{X}^{*}(\omega) \mathbf{W}(\omega) = \sum_{\omega \in U_{c}} \mathbf{Y}(\omega) \mathbf{W}(\omega)$$

For any γ between α and β , let $V_c = \{\omega \notin C: \gamma \leq X^*(\omega) \leq \beta\}$ and let $V = V_c \cup [X^* > \beta]$. Then V is an upper set and similarly

$$\sum_{\omega \in V_{c}} X(\omega) W(\omega) \leq \sum_{\omega \in V_{c}} X^{*}(\omega) W(\omega).$$

The set $\{\omega \in C : X^*(\omega) \ge \gamma\}$ is an upper set in C and $V_{C} \cup \{\omega \in C : X^*(\omega) \ge \gamma\}$ is the level set $[\gamma \le X^* \le \beta]$. Therefore, by (3.2) we have

$$\Sigma_{\omega} \in \mathbf{C}, \mathbf{X}^{*}(\omega) \geq \gamma^{\mathbf{X}(\omega)\mathbf{W}(\omega)} = \Sigma_{\omega} \in \mathbf{C}, \mathbf{X}^{*}(\omega) \geq \gamma^{\mathbf{X}^{*}(\omega)\mathbf{W}(\omega)}$$
$$= \Sigma_{\omega} \in \mathbf{C}, \mathbf{X}^{*}(\omega) \geq \gamma^{\mathbf{Y}(\omega)\mathbf{W}(\omega)}$$

and it follows that Y is the restricted isotonic regression of X to C. The restriction of X* to any non-empty level set $[a \leq X^* \leq \beta]$, to one of its component or to the union of some of its components is the restricted isotonic regression of X to that set. Therefore, the following theorem has been established.

<u>Theorem 3.6.</u> Let X^* be the isotonic regression of X, let the non-empty level set $[a \leq X^* \leq \beta]$ have components C_1, \ldots, C_k and let $\Gamma = \bigcup_{i=1}^{j} C_i$ where $1 \leq j \leq k$ Then the restriction of X^* to Γ is the restricted isotonic regression of Xto Γ .

Theorem 3.7. Let Γ_2 be a non-empty level set, let $\Gamma_1 \subset \Gamma_2^c$ be a lower set such that $\omega \in \Gamma_2^c$ and $\omega \leq \mu$ for some $\mu \in \Gamma_2$ imply that $\omega \in \Gamma_1$, let $\Gamma_3 \subset \Gamma_1^c \cap \Gamma_2^c$ be an upper set such that $\omega \in \Gamma_2^c$ and $\omega \geq \mu$ for some $\mu \in \Gamma_2$ imply that $\omega \in \Gamma_3$ and let Υ_1 , Υ_2 and Υ_3 be the restricted isotonic regressions of X to Γ_1 , Γ_2 and Γ_3 respectively. If there exist a and β such that $\Upsilon_1 \leq \alpha, \alpha \leq \Upsilon_2 \leq \beta$ and $\beta \leq \Upsilon_3$, then $X^*(\omega) \leq \Upsilon_1(\omega)$ for each $\omega \in \Gamma_1$, $X^*(\omega) = \Upsilon_2(\omega)$ for each $\omega \in \Gamma_2$ and $X^{*}(\omega) \ge Y_{3}(\omega)$ for each $\omega \in \Gamma_{3}$ where X^{*} is the isotonic regression of X.

<u>Proof</u>. Let $\Gamma_4 = \Omega - (\Gamma_1 \cup \Gamma_2 \cup \Gamma_3)$, let Z be an isotonic function and let Y be defined by $Y(\omega) = Z(\omega) \wedge Y_{1}(\omega)$ on Γ_{1} , $Y(\omega) = Y_2(\omega)$ on Γ_2 , $Y(\omega) = Z(\omega) \checkmark Y_3(\omega)$ on Γ_3 and $Y(\omega) = Z(\omega)$ on Γ_4 . We shall show that Y is isotonic and $f(Y) \leq f(Z)$. It will then follow that $X^{*}(\omega) \leq Y_{1}(\omega)$ for each $\omega \in \Gamma_{1}$, $X^{*}(\omega) = Y_{2}(\omega)$ for each $\omega \in \Gamma_{2}$ and $X^{*}(\omega) \ge Y_{3}(\omega)$ for each $\omega \in \Gamma_3$. The order-preserving property of Y on each Γ_i is trivial. The sets Γ_2 and Γ_4 are unrelated. Since Γ_1 is a lower set, elements in Γ_1 will be no greater than ay element in $\Gamma_2 \cup \Gamma_3 \cup \Gamma_4$. Since Γ_3 is an upper set, elements in Γ_3 will be no less than any element in $\Gamma_1 \cup \Gamma_2 \cup \Gamma_4$. Any element greater than an element in Γ_2 is in $\Gamma_2 \cup \Gamma_3$ and any element less than an element in Γ_2 is in $\Gamma_1 \cup \Gamma_2$. Since $Y \leq a$ on Γ_1 , $\alpha \leq Y \leq \beta$ on Γ_2 and $Y \geq \beta$ on Γ_3 , it follows that Y is isotonic. By (4.7), we have $f(Y; \Gamma_1) \leq f(Z; \Gamma_1)$ and $f(Y;\Gamma_3) \leq f(Z;\Gamma_3)$ when we regard Γ_1 or Γ_3 as our given partially ordered set. Since $f(Y;\Gamma_2) \leq f(Z;\Gamma_2)$, $f(Y;\Gamma_4) = f(Z;\Gamma_4)$ and $f(Z) = \sum_{i=1}^{4} f(Z; \Gamma_i)$, we have $f(Y) \leq f(Z)$. This completes the proof. []

Corollary 3.7.1. Let U be a non-empty upper set, let $\Gamma_1 \subset U^C$ be a lower set such that $\omega \in U^C$ and $\omega \leq \mu$ for some $\mu \in U$ imply that $\omega \in \Gamma_1$ and let Y_U and Y_1 be the restricted isotonic regressions of X to U and Γ_1 respectively. If there exists an α such that $Y_1 \leq \alpha$ and $\alpha \leq Y_U$, then $X^*(\omega) = Y_U(\omega)$ for all $\omega \in U$ and $X^*(\omega) \leq Y_1(\omega)$ for all $\omega \in \Gamma_1$. Similarly, let L be a non-empty lower set, let $\Gamma_3 \subset L^C$ be an upper set such that $\omega \in L^C$ and $\omega \geq \mu$ for some $\mu \in L$ imply that $\omega \in \Gamma_3$ and let Y_L and Y_3 be the restricted isotonic regressions of X to L and Γ_3 respectively. If there exists a β such that $Y_L \leq \beta$ and $\beta \leq Y_3$, then $X^*(\omega) = Y_L(\omega)$ for each $\omega \in L$ and $X^*(\omega) \geq Y_3(\omega)$ for each $\omega \in \Gamma_3$.

<u>Proof.</u> For the first statement, let $\Gamma_2 = U$ and $\Gamma_3 = \phi$. For the second statement, let $\Gamma_1 = \phi$ and $\Gamma_2 = L$. The results follow from Theorem 3.7. []

<u>Corollary 3.7.2.</u> Let Γ_2 be a non-empty level set, let $\Gamma_1 = \{\omega \in \Gamma_2^{\mathbb{C}} : \omega \leq \mu \text{ for some } \mu \in \Gamma_2\}$ and let $\Gamma_3 = \{\omega \in \Gamma_2^{\mathbb{C}} : \omega \geq \mu \text{ for some } \mu \in \Gamma_2\}$. If there exist α and β such that $X \leq \alpha$ on Γ_1 , $\alpha \leq X \leq \beta$ on Γ_2 and $X \geq \beta$ on Γ_3 , then the isotonic regression of X can be obtained by considering Γ_2 and $\Gamma_2^{\mathbb{C}}$ independently, i.e., the restriction of the isotonic regression of X to Γ_2 is the restricted isotonic regression of X to Γ_2 .

<u>Proof.</u> Γ_1 is a lower set, Γ_3 is an upper set and $\Gamma_1 \cap \Gamma_3 = \phi$. Let Y_1, Y_2 , and Y_3 be the restricted isotonic regressions of X to Γ_1, Γ_2 and Γ_3 respectively. Then by the monotonicity in Theorem 4.5, we have $Y_1 \leq \alpha, \ \alpha \leq Y_2 \leq \beta$ and $Y_3 \geq \beta$. It follows from Theorem 3.7 that $X^*(\omega) = Y_2(\omega)$ for each $\omega \in \Gamma_2$. []

<u>Corollary 3.7.3.</u> Let U be a non-empty upper set and let $\Gamma_1 = \{\omega \in U^C : \omega \leq \mu \text{ for some } \mu \in U\}$. If there exists an a such that $X \geq a$ on U and $X \leq a$ on Γ_1 , then the isotonic regression of X can be obtained by considering U and U^C independently. Similarly, let L be a non-empty lower set and let $\Gamma_3 = \{\omega \in L^C : \omega \geq \mu \text{ for some } \mu \in L\}$. If there exists a β such that $X \leq \beta$ on L and $X \geq \beta$ on Γ_3 , then the isotonic regression of X can be obtained by considering L and L^C independently.

In particular, if a maximal element has a maximum X value or a minimal element has a minimum X value, then the value of the isotonic regression of X at that element is the value of X at that element.

<u>Proof</u>. For the first statement, let $\Gamma_2 = U$ and $\Gamma_3 = \phi$. For

the second statement, let $\Gamma_1 = \phi$ and let $\Gamma_2 = L$. The results follow from Corollary 3.7.2. []

The level set Γ_2 in Theorem 3.7, if it is connected, will be a component of the level set $[a \leq X^* \leq \beta]$ where X^* is the isotonic regression of X. Such a component may be obtained by inspection of X values as indicated by Corollary 3.7.2 or Corollary 3.7.3. Corollary 3.2.3 and Theorem 3.3 may be able to take care of violations locally; Theorem 3.7 and its corollaries may be able to take care of the problem globally. Theoretically, these results are enough to obtain the isotonic regression of X. The difficulties are how to determine such a component Γ_2 and how to find the restricted isotonic regression of X to Γ_2 . Presumably, if the longest chain in Ω has a small number of elements or upper sets tend to have larger X values than lower sets, then the difficulties mentioned above will not be serious.

When difficulties arise, the reader may want to refer to Alexander's algorithm. But Corollary 3.2.3, Theorem 3.3, Theorem 3.7 and its corollaries may be greatly helpful in reducing the problem, as demonstrated by the following example which appears in Alexander (1970).

Example 3.3. Let $\Omega = \{(i,j): i, j = 1,2,3,4\}$ be a partially ordered set with the ordering $(i,j) \leq (h,k)$ if $i \leq h$ and $j \leq k$. For each element (i, j), the function value x and the weight w are ij respectively the fraction and the denominator as given below.

	ji	1	2	3	4
	1	1/16	1/7	4/39	1/6
X	2	1/8	1/10	1/7	1/2
	3	1/6	1/7	1/8	1/3
	4	1/5	2/11	1/2	1/3

By inspection, one may find out that $U = \{(3, 4), (4, 2), (4, 3), (4, 4)\}$ is an upper set and $X \ge 1/3$ on U, $X \le 1/5$ on U^C; (1,1) is the minimal element with the minimum X value; $\Gamma_2 = \{(1,4), (2,4)\}$ is a level set with $\Gamma_3 = \{(3,4), (4,4)\}$ and $\Gamma_1 = \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (2, 3)\}$ such that X < 2/11 on Γ_1 , $2/11 \le X \le 1/5$ on Γ_2 and X > 1/5 on Γ_3 ; and also $\Lambda_2 = \{(4,1)\}$ is a level set with $\Lambda_3 = \{(4,2), (4,3), (4,4)\}$ and $\Lambda_1 = \{(1, 1), (1, 2), (1, 3)\}$ such that X < 1/6 on Λ_1 , X = 1/6 on Λ_2 and X > 1/6 on Λ_3 . From Corollary 3.7.2 and Corollary 3.7.3, $x_{11}^* = 1/16$ where X* is the isotonic regression of X and we may consider U, Γ_2 and Λ_2 independently. The set U is a reversed tree and by the Maximum Upper Set algorithm we have $x_{34}^* = x_{42}^* = x_{43}^* = x_{44}^* = 2/5$. The set Γ_2 is a linearly ordered set and by the Pool-Adjacent-Violators algorithm we have $x_{14}^* = x_{24}^* = 3/16$. The set Λ_2 is a singleton and hence $x_{41}^* = 1/6$.

Therefore, the remaining problem is the following:

Since (1,3) and (3,2) have the largest X value 1/6, Theorem 3.3. shows that $x_{13}^* = x_{23}^*$ and $x_{32}^* = x_{33}^*$. By the smoothing property as described in the Maximum Lower Set algorithm, we may replace 1/7 and 1/8 at (2,3) and (3,3) by 2/13 and 2/14 respectively and cross out elements (1,3) and (3,2). The weights at (2,3) and (3,3) now are 13 and 14.

At this stage we have six elements. Since $X_2 \ge 1/7$ in the upper set $U = \{(2,3), (3,3)\}$ and $X_2 \le 1/7$ on U^c , by Corollary 3.7.3 we may consider U and U^c independently. The set U is linearly ordered and by the Pool-Adjacent-Violators algorithm we have $x_{23}^* = x_{33}^* = 4/27$. Recall that $x_{13}^* = x_{23}^*$ and $x_{32}^* = x_{33}^*$, so

$$\mathbf{x}_{13}^{*} = \mathbf{x}_{23}^{*} = \mathbf{x}_{32}^{*} = \mathbf{x}_{33}^{*} = 4/27.$$

$$\mathbf{x}_{3} = \frac{2}{1/8} + \frac{1}{10}$$

$$\mathbf{x}_{3} = \frac{1}{1/7} + \frac{1}{1/7} + \frac{1}{1/39}$$

$$\frac{1}{1/7} + \frac{1}{1/7} + \frac{1}{1/39}$$

For a set with four elements, the problem can always be solved very easily. If $[\mu \leq \nu]$ is an immediately comparable pair and $X(\mu) \ge X(\nu)$, then by Corollary 3.2.3 we shall group μ with at least one of its immediate successors or group ν with at least one of its immediate predecessors. It follows that $x_{12}^* = x_{22}^*$, $x_{21}^* = x_{22}^*$ or $x_{21}^* = x_{33}^*$. Let us consider U = {(1,2), (2,2)}. This is an upper set and the restricted isotonic regressions Y and Z of X_3 to U and U^{c} are $y_{12} = y_{22} = 2/18$ and $z_{21} = z_{31} = 5/46$. The value 2/18 is larger than 5/46 and from Corollary 3.7.1 we have $x_{12}^* = x_{22}^* = 2/18$ and $x_{21}^* = x_{22}^* = 5/46$. Therefore, the isotonic regression X^* of X has been obtained. Suppose we group (2, 1) and (2,2) first. Then we have a linearly ordered set with function values from the smallest element to the largest element being 1/8, 2/17 and 4/39. The Pool-Adjacent-Violators algorithm shows that $x_{12}^* = x_{21}^* = x_{22}^* = x_{31}^* = 7/64$. But $x_{12}w_{12} + x_{22}w_{22} = 2$ and $x_{12}^* w_{12} + x_{22}^* w_{22} = 63/32$ which violates (3.3). Therefore, the proper grouping should be $\{(1,2), (2,2)\}$ or $\{(2,1), (3,1)\}$ and either one will lead to the final solution. The isotonic regression X^* of

X is given below. []

X*	4	3/16	3/16	2/5	2/5
	3	4/27	4/27	4/27	2/5
	2	2/18	2/18	4/27	2/5
	1	1/16	5/46	5/46	1/6
	j i	1	2	3	4

III. 6 Some Related Problems

Let Ω be a partially ordered set and let W be a given weight function. For any X there is an isotonic regression X* of X. Let $\{\alpha_1, \ldots, \alpha_k\}$ be the range of X* and let $\Gamma_i = [X* = \alpha_i], i = 1, \ldots, k$. The sequence $\Gamma_1, \ldots, \Gamma_k$ is a partition of Ω . Such a partition may possibly be determined by Theorem 3.3, Corollary 3.7.2 and Corollay 3.7.3 without regarding what the weight function is. Therefore, we will be able to select another weight function W_0 of interest such that the function Y defined by $Y(\omega) = \sum_{\omega \in \Gamma_i} X(\omega) W_0(\omega) / \sum_{\omega \in \Gamma_i} W_0(\omega)$ for $\omega \in \Gamma_i$, is isotonic. Our interest could be the weight function, the isotonic regression or both.

The problem we are going to study is to minimize

$$f(Z, W) = \sum_{\omega \in \Omega} [Z(\omega) - X(\omega)]^2 W(\omega)$$

subject to $Z \in M$ and subject to some conditions on the weight

function W. For each weight function W, there is an isotonic regression $X^{*}(W)$. Since $f(Z, W) \ge f(X^{*}(W), W)$, the minimization problem can be studied as a function of W, i.e.,

$$h(W) = f(X * (W), W).$$

The isotonic regression $X^{*}(W)$ may not heavily depend on W. If that is the case, the problem can be solved very easily.

Example 3.4. The problem is to minimize

$$f(Z, W) = (z_{11}^{-1/7})^2 w_{11}^{-1} + (z_{21}^{-1/8})^2 w_{21}^{-1/3} + (z_{31}^{-1/3})^2 w_{31}^{-1/3} + (z_{12}^{-2/11})^2 w_{12}^{-1/2} + (z_{22}^{-1/2})^2 w_{22}^{-1/3} + (z_{32}^{-1/3})^2 w_{32}^{-1/3}$$

subject to

$$\mathbf{z}_{ij} \leq \mathbf{z}_{hk} \quad \text{if } i \leq h \quad \text{and} \quad j \leq k \\ \mathbf{w}_{ij} \geq 1/5 \quad i = 1, 2, 3; \quad j = 1, 2$$

and

$$w_{i1} + w_{i2} = 1$$
, $i = 1, 2, 3$

Let X be given by

For each W, let $X^{*}(W) = (x_{ij}^{*})$ be the isotonic regression of X with respect to W. Then $x_{11}^{*} = x_{21}^{*} = (w_{11}^{7} + w_{12}^{8})/(w_{11}^{1} + w_{12}^{1}),$ $x_{31}^{*} = 1/3, x_{12}^{*} = 2/11$ and $x_{22}^{*} = x_{32}^{*} = (w_{22}^{2} + w_{32}^{8})/(w_{22}^{2} + w_{23}^{8}).$ Therefore,

h(W) = f(X*(W), W)= $(1/7 - x_{11}^*)^2 w_{11} + (1/8 - x_{21}^*)^2 w_{21} + (1/2 - x_{22}^*)^2 w_{22}$ + $(1/3 - x_{32}^*)^2 w_{32}$ = $w_{11} w_{21} (w_{11} + w_{21})^{-1} / 3136 + w_{22} w_{32} (w_{22} + w_{32})^{-1} / 36$

For fixed w_{21} , w_{22} and w_{32} , h is an increasing function of w_{11} over the range [1/5, 4/5] and hence $w_{11}^* = 1/5$ where W* is the optimal solution. Similarly, $w_{32}^* = 1/5$. Let $\lambda = w_{21}$. We have

$$h(\lambda) = \lambda/3136(1+5\lambda) + (1-\lambda)/36(6-5\lambda)$$
 where $\lambda \in [1/5, 4/5]$.

Since h is monotone decreasing, $\lambda = 4/5$. The optimal solution (X^*, W^*) has been obtained and it is given below. []

X*
 W*

 2
 2/11

$$5/12$$
 2
 $4/5$
 $1/5$
 $1/5$

 1
 9/70
 9/70
 $1/3$
 1
 $1/5$
 $4/5$
 $4/5$

 j
 i
 1
 2
 3
 j
 i
 1
 2
 3

Another problem we are interested in is the situation when the ordering imposed on the set Ω could be one of k given types. We want to minimize

$$f(Z) = \sum_{\omega \in \Omega} [Z(\omega) - X(\omega)]^2 W(\omega)$$

subject to $Z \in \bigcup_{i=1}^{k} M(\Omega, \leq_{i})$ where $M(\Omega, \leq_{i})$ is the family of isotonic functions with respect to the ordering \leq_{i} , $i = 1, \ldots, k$. Let X_{i}^{*} be the isotonic regression of X with respect to the ordering \leq_{i} . Since $f(Z) \geq f(X_{i}^{*})$ whenever $Z \in M(\Omega, \leq_{i})$, the optimal solution is X_{i}^{*} such that $f(X_{i}^{*}) = \min\{f(X_{i}^{*}): i = 1, \ldots, k\}$.

Example 3.5. The problem is to minimize

$$f(Z) = (z_1 - 3)^2 + (z_2 - 2)^2 + (z_3 - 7)^2 + (z_4 - 8)^2 + (z_5 - 5)^2$$

subject to Z being <u>unimodal</u>, i.e., $z_1 \leq z_2 \leq \cdots \leq z_j$ and $z_j \geq z_{j+1} \geq \cdots \geq z_5$ for j from one to five. For each ordering \leq_i , the pair (Ω, \leq_i) is a reversed tree where $\Omega = \{1, 2, 3, 4, 5\}$. The isotonic regressions of X = (3, 2, 7, 8, 5) with respect to these orderings can be found very easily and the optimal solution is (2.5, 2.5, 7, 8, 5).

<u>Bounded isotonic regression</u> is isotonic regression with an addition constraint:

$$Z_1 \leq Z \leq Z_2$$

where Z_1 and Z_2 are two given functions. The problem was introduced by van Eeden (cf. Barlow and coworkers (1972)). Let X^* and \hat{X} be the isotonic regression and the bounded isotonic regression of a given X over a partially ordered finite set and let $\hat{M} := \{Z \in M : Z_1 \leq Z \leq Z_2\}$. Since \hat{M} is a cosed convex set, if it is non-empty then the bounded isotonic regression exists and is unique

By the monotonicity (cf. Theorem 4.5) of the isotonic regression, if $Z_1 \leq Z \leq Z_2$ and $Z \in M$, then $Z_1^* \leq Z \leq Z_2^*$ where Z_1^* and Z_2^* are respectively the isotonic regressions of Z_1 and Z_2 . Without loss of generality, we may assume that Z_1 and Z_2 are isotonic. If there is an ω such that $Z_2(\omega) < Z_1(\omega)$, then there is no feasible solution. If $Z_1 \leq Z_2$ and $Z_1(\omega) = Z_2(\omega)$ for some ω , then $X(\omega) = Z_1(\omega)$. In the remainder of the section, we shall assume that Z_1 and Z_2 are isotonic and $Z_1 \leq Z_2$.

<u>Theorem 3.8.</u> If Z_1 and Z_2 are constant functions with values a and β respectively, then $X = (X * \sqrt{a}) \wedge \beta$.

<u>Proof</u>. Let $U = [X^* > \beta]$, $L = [X^* < \alpha]$ and $\Gamma = [\alpha \le X^* \le \beta]$. From Theorem 3.6, $X^* | U$, $X^* | L$ and $X^* | \Gamma$ are the restricted isotonic regressions of X to U, L and Γ respectively. Let $Z \in \widehat{M}$, then $Z \le \beta$ on U and $X^* > \beta$ on U. By (4.8), we have $f(Z \lor \beta; U) \leq f(Z; U)$ where $f(Z; U) := \sum_{\omega \in U} [X(\omega) - Z(\omega)]^2 W(\omega)$. Similarly, $f(Z \land \alpha; L) \leq f(Z; L)$. Since $f(Z; \Gamma) \geq f(X^*; \Gamma)$, $Z \lor \beta = \beta$, $Z \land \alpha = \alpha$ and $f(Z) = f(Z; U) + f(Z; L) + f(Z; \Gamma)$, it follows that $\hat{X} = X^*$ on Γ , $\hat{X} = \alpha$ on L and $\hat{X} = \beta$ on U.

Let
$$a = \min\{X^*(\omega) : \omega \in \Omega\}$$
, $b = \max\{X^*(\omega) : \omega \in \Omega\}$,
 $a = \min\{Z_2(\omega) : \omega \in \Omega\}$ and $\beta = \max\{Z_1(\omega) : \omega \in \Omega\}$ Let
 $U = \{\omega : Z_1(\omega) = \beta\}$, $U_b = \{\omega : Z_1(\omega) \ge b\}$, $L = \{\omega : Z_2(\omega) = a\}$ and
 $L_a = \{\omega : Z_2(\omega) \le a\}$.

<u>Lemma 3.1.</u> If U_b is non-empty, then $\hat{X}(\omega) = Z_1(\omega)$ for each $\omega \in U$. If L_a is non-empty, then $\hat{X}(\omega) = Z_2(\omega)$ for each $\omega \in L$.

<u>Proof.</u> If U_b is non-empty, then $\beta \ge b$. For each $\omega \in U$, $\widehat{X}(\omega) \ge \beta$. We are going to show $\widehat{X}(\omega) = \beta$. Note that $X^* \le b$ and hence $X^* \le \beta$. If $Z \in \widehat{M}$, then $Z \land \beta \in \widehat{M}$. By (4.8), $f(Z) \ge f(Z \land \beta)$. Therefore $\widehat{X} = \widehat{X} \land \beta$. Similarly if L_a is nonempty then $\widehat{X}(\omega) = Z_2(\omega)$ for each $\omega \in L$. []

<u>Lemma 3.2</u>. Let X* be a constant function. If L_a is empty, then $\hat{X} \ge X^*$. If U_b is empty, then $X^* \ge \hat{X}$.

<u>Proof.</u> If $L_a = \phi$, then $Z_2 > a = X^*$ and it follows that $Z \in \widehat{M}$ implies $Z \vee X^* \in \widehat{M}$. By (4.7), $f(Z) \ge f(Z \vee X^*)$. Therefore $\hat{\mathbf{X}} = \hat{\mathbf{X}} \vee \mathbf{X}^*$. Similarly, if $\mathbf{U}_{\mathbf{h}}$ is empty, then $\mathbf{X}^* \ge \hat{\mathbf{X}}$. []

<u>Theorem 3.9.</u> Let $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ be a linearly ordered set with the ordering $\omega_i \leq \omega_{i+1}$, $i = 1, 2, \dots, n-1$. The bounded isotonic regression of X can be obtained by considering each level set $[X^* = c]$ independently, where X^* is the isotonic regression of X.

 $\begin{array}{lll} \underline{\operatorname{Proof}} & \operatorname{Let} & [\mathbf{X}^* = \mathbf{a}] = \{\omega_i, \omega_{i+1}, \ldots, \omega_j\} & \operatorname{and} \\ [\mathbf{X}^* = \mathbf{b}] = \{\omega_{j+1}, \ldots, \omega_k\} & \operatorname{be} \text{ two adjacent level sets with } \mathbf{a} < \mathbf{b} & \operatorname{and} \\ \operatorname{let} \ & \widehat{\mathbf{X}}_{\mathbf{a}} & \operatorname{and} & \widehat{\mathbf{X}}_{\mathbf{b}} & \operatorname{be} \text{ the bounded restricted isotonic regressions of} \\ \mathbf{X} & \operatorname{to} & [\mathbf{X}^* = \mathbf{a}] & \operatorname{and} & [\mathbf{X}^* = \mathbf{b}]. & \operatorname{If} \text{ we can show} & \widehat{\mathbf{X}}_{\mathbf{a}}(\omega_j) \leq \widehat{\mathbf{X}}_{\mathbf{b}}(\omega_{j+1}), \\ \operatorname{then} & \widehat{\mathbf{X}}(\omega) = \widehat{\mathbf{X}}_{\mathbf{a}}(\omega) & \operatorname{for} \text{ each } \omega \in [\mathbf{X}^* = \mathbf{a}] & \operatorname{and} & \widehat{\mathbf{X}}(\omega) = \widehat{\mathbf{X}}_{\mathbf{b}}(\omega) & \operatorname{for} \\ \operatorname{each} & \omega \in [\mathbf{X}^* = \mathbf{b}]. & \operatorname{By} \text{ Theorem 3.6, the restricted isotonic regressions of} \\ \mathbf{X} & \operatorname{to} \text{ each level set} & [\mathbf{X}^* = \mathbf{c}] & \operatorname{is the constant} & \mathbf{c}. & \operatorname{There-fore,} & \operatorname{if} & \mathbf{Z}_1(\omega_j) \geq \mathbf{a} & \operatorname{then} \operatorname{by} \operatorname{Lemma} 3.1 & \widehat{\mathbf{X}}_{\mathbf{a}}(\omega_j) = \mathbf{Z}_1(\omega_j); & \operatorname{other} - \\ & \operatorname{wise,} & \operatorname{by} \operatorname{Lemma} 3.2 & \widehat{\mathbf{X}}_{\mathbf{a}}(\omega_j) \leq \mathbf{X}^*(\omega_j) = \mathbf{a}. & \operatorname{Similarly,} & \operatorname{if} \\ \mathbf{Z}_2(\omega_{j+1}) \leq \mathbf{b} & \operatorname{we} & \operatorname{have} & \widehat{\mathbf{X}}_{\mathbf{b}}(\omega_{j+1}) = \mathbf{Z}_2(\omega_{j+1}) & \operatorname{and} & \operatorname{otherwise} \\ & \widehat{\mathbf{X}}_{\mathbf{b}}(\omega_{j+1}) \geq \mathbf{b} & \operatorname{If} & \mathbf{Z}_1(\omega_j) \geq \mathbf{a}, & \operatorname{then} & \widehat{\mathbf{X}}_{\mathbf{a}}(\omega_j) = \mathbf{Z}_1(\omega_j) \leq \mathbf{Z}_1(\omega_{j+1}) \\ & \leq \widehat{\mathbf{X}}_{\mathbf{b}}(\omega_{j+1}). & \operatorname{If} & \mathbf{Z}_2(\omega_{j+1}) \leq \mathbf{b}, & \operatorname{then} \\ & \widehat{\mathbf{X}}_{\mathbf{a}}(\omega_j) \leq \mathbf{Z}_2(\omega_j) \leq \mathbf{Z}_2(\omega_{j+1}) = \widehat{\mathbf{X}}_{\mathbf{b}}(\omega_{j+1}). & \operatorname{Otherwise} \\ & \widehat{\mathbf{X}}_{\mathbf{a}}(\omega_j) \leq \mathbf{a} < \mathbf{b} \leq \widehat{\mathbf{X}}_{\mathbf{b}}(\omega_{j+1}). & \operatorname{This} & \operatorname{completes} & \operatorname{the} & \operatorname{proof.} \end{array} \right]$

<u>Example 3.6.</u> Let X = (25, 13, 2, 15, 14, 21, 9, 33, 25, 15),

82

 $Z_1 = (10, 11.5, 13, 14.5, 16, 17.5, 19, 20.5, 22, 23.5)$ and $Z_2 = (13, 14.5, 16, 17.5, 19, 20.5, 22, 23.5, 25, 26.5)$. The problem is to minimize

$$f(Z) = \Sigma_{i=1}^{10} (x_i - z_i)^2$$

subject to $z_1 \leq z_2 \leq \cdots \leq z_{10}$ and $Z_1 \leq Z \leq Z_2$.

The isotonic regression X^* of X is $X^* = (13.3, 13.3, 13.3, 14.5, 14.5, 15, 15, 24.3, 24.3, 24.3)$. Theorem 3.9 shows that the bounded isotonic regression $\hat{\mathbf{X}}$ of \mathbf{X} can be obtained by considering the set {25, 13, 2}, {15, 14}, {21, 9} and $\{33, 25, 15\}$ of X values independently. For the first partition $\{25, 13, 2\}$, the average is 13.3 and $L_a = \{1\}$. By Lemma 3.1, $\hat{x}_1 = z_{21} = 13$. Therefore, \hat{x}_2 and \hat{x}_3 are the optimal solution to the following problem. Minimize $(13-z_2)^2 + (2-z_3)^2$ subject to $13 \le z_2 \le 14.5$, $13 \le z_3 \le 16$ and $z_2 \le z_3$. Hence $\hat{x}_2 = 13$ and $\hat{x}_3 = 13$. For the second partition {15, 14}, the average is 14.5 and $U_{b} = \{4, 5\}$. By Lemma 3.1, $\hat{x}_{5} = z_{15} = 16$. Therefore, \hat{x}_{4} is the optimal solution to the problem: minimize $(15-z_A)^2$ subject to $14.5 \le z_4 \le 16$. Hence $x_4^* = 15$. Similarly, for the third partition {21,9} we have $\hat{\mathbf{x}}_7 = 19$ and it follows that $\hat{\mathbf{x}}_6 = 19$. For the last partition $\{33, 25, 15\}$, we have $\hat{x}_8 = 23.5$ and it follows that $\hat{x}_9 = \hat{x}_{10} = 23.5$. The bounded isotonic regression is therefore

 $\hat{\mathbf{X}} = (13, 13, 13, 15, 16, 19, 19, 23, 5, 23, 5, 23, 5)$. In order to check that $\hat{\mathbf{X}}$ has the correct values, one may refer to the Kuhn-Tucker condition (cf. Appendix IV). []

IV. CONDITIONAL EXPECTATION GIVEN A σ-LATTICE

IV.1 σ -Lattices and Σ -Measurable Random Variables

Let Σ be a family of subsets of a given set Ω . It is said to be a <u>lattice</u> if it contains Ω and ϕ' and it is closed under union and closed under intersection. A <u>field</u> is a lattice which contains complements of sets in the family. A <u> σ -lattice</u> is a lattice which is closed under countable union and closed under countable intersection. A <u> σ -field</u> is a σ -lattice which is also a field. A <u>complete lattice</u> is a lattice which is closed under arbitrary union and closed under arbitrary intersection. The family Σ is said to be a <u>monotone class</u> if whenever $\{A_n\}$ is a monotone sequence in Σ , both $\bigcup_{n=1}^{\infty} A_n$ and $\bigcap_{n=1}^{\infty} A_n$ are in Σ .

The collection of complements of a monotone class is a monotone class and the collection of complements of a σ -lattice (lattice, or complete lattice) Σ is a σ -lattice (lattice or complete lattice); such a collection is denoted by Σ^{C} . The intersection of an arbitrary collection of σ -lattices is a σ -lattice. In particular, the intersection of a σ -lattice Σ with its complement Σ^{C} is a σ -field. Let Γ be a family of subsets of Ω . The σ -lattice generated by Γ is the intersection of all σ -lattices which contain Γ . Similarly for the definition of σ -field generated by a family of subsets in Ω . Let \leq be a quasi-ordering on Ω and let Σ be the family of all upper sets. Then Σ is a complete lattice, such a complete lattice is said to be <u>induced</u> by \leq . On the other hand, let Γ be a family of subsets of Ω and let \leq be the binary relation defined on Ω so that $\mu \leq \nu$ if $\mu \in U$ and $U \in \Gamma$ imply $\nu \in U$; if no set in Γ contains μ , then $\mu \leq \nu$ for every $\nu \in \Omega$. Then the binary relation \leq is a quasi-ordering on Ω ; such an ordering is said to be <u>induced</u> by Γ .

Let Δ be the collection of all families of subsets of Ω each of which induces the same quasi-ordering \leq . Let $\Gamma \in \Delta$. For each $\omega \in \Omega$, let $U(\omega) = \bigcap \{U: \omega \in U, U \in \Gamma\}$ and let $L(\omega) = \bigcap \{L: \omega \in L, L^{C} \in \Gamma\}$; and if there is no set in Γ which contains ω , set $U(\omega) = \Omega$ and if every set in Γ contains ω , set $L(\omega) = \Omega$. It is clear from the definitions, for each pair μ and ν in Ω , $\mu \leq \nu$, $\nu \in U(\mu)$ and $\mu \in L(\nu)$ are equivalent. Since $U(\mu)$ is the set of all elements ν such that $\mu \leq \nu$, it follows that for each $\mu \in \Omega$, $U(\mu)$ is the same for all families in Δ .

Let $\Gamma \in \Delta$, and let $U \in \Gamma$. Then for each $\omega \in U$, we have $\omega \in U(\omega) \subset U$ and hence $U = \bigcup_{\omega \in U} U(\omega)$. Let Σ be the family of arbitrary unions of $U(\omega)$'s. If $U \in \Sigma$, then for each $\omega \in U$ we have $U(\omega) \subset U$. Since Σ contains every $U(\omega)$, $\Sigma \in \Delta$ and if $\Gamma \in \Delta$ then we have $\Gamma \subset \Sigma$. The family Σ is maximal in Δ in the sense of set inclusion. By the convention, a void union of sets is

86

the empty set. The family Σ contains Ω and ϕ , and it is closed under arbitrary union. We shall show that Σ is also closed under arbitrary intersection. It will then follow that Σ is a complete lattice. Let $\{U_{\alpha}: \alpha \in \Lambda\}$ be a subfamily of Σ and let $A = \bigcap_{\alpha \in \Lambda} U_{\alpha}$. If $\omega \in A$, then $\omega \in U_{\alpha}$ for each $\alpha \in \Lambda$ and hence $U(\omega) \subset U_{\alpha}$ for each $\alpha \in \Lambda$. It follows that

$$\mathsf{A} = igcup_{\omega \in \mathbf{A}} \{\omega\} \subset igcup_{\omega \in \mathbf{A}} \mathsf{U}(\omega) \subset \mathsf{A}$$

and hence $A \in \Sigma$. If $A = \phi$, then $A \in \Sigma$. If Σ' is a complete lattice in Δ , then Σ' contains $\{U(\omega): \omega \in \Omega\}$. It follows that $\Sigma \subset \Sigma'$. But we have just shown that $\Gamma \subset \Sigma$ for each $\Gamma \in \Delta$. Therefore, the collection Δ contains at most one complete lattice.

Let \leq be a quasi-ordering in Ω , let Σ' be the complete lattice induced by \leq and let \leq' be the quasi-ordering induced by Σ' . Let μ and ν be in Ω . If $\mu \leq \nu$, then every upper set $U \in \Sigma'$ which contains μ , contains ν . It follows that $\mu \leq' \nu$. On the other hand, if $\mu \leq \nu$ then there exists an upper set $U \in \Sigma'$ such that $\mu \in U$ and $\nu \notin U$. It follows that $\mu \leq' \nu$ and therefore \leq and \leq' are identical. The complete lattice Σ' is in Δ , thus we have proved the following theorem.

<u>Theorem 4.1.</u> Let Ω be a given set. Quasi-orderings on Ω and complete lattices of Ω are in one-to-one correspondence such that if \leq corresponds to Σ , then \leq induces Σ and Σ induces \leq .

Let Δ be the collection of all families of subsets of Ω each of which induces the same quasi-ordering \leq . We have shown that the maximal element in Δ is the complete lattice induced by the quasi-ordering \leq . However, minimal elements in Δ need not be unique.

Let \leq_1 and \leq_2 be two quasi-orderings on Ω . \leq_1 is said to be finer than \leq_2 if for each pair μ and ν in Ω , $\mu \leq_1 \nu$ implies $\mu \leq_2 \nu$. Let Γ_1 and Γ_2 be two families of subsets of Ω and let \leq_1 and \leq_2 be the quasi-orderings induced respectively by Γ_1 and Γ_2 . If $\Gamma_2 \subset \Gamma_1$ and $\mu \leq v$, then every $U \in \Gamma_1$ which contains μ , contains ν ; so does every $U \in \Gamma_2$. It follows that \leq_1 is finer than \leq_2 . But if \leq_1 is finer than \leq_2 , it need not be true that $\Gamma_2 \subset \Gamma_1$. On the other hand, let Σ_1 and Σ_2 be the complete lattices induced respectively by two quasi-orderings \leq_1 and \leq_2 on Ω . Let \leq_1 be finer than \leq_2 . For each $\omega \in \Omega$, let $U_i(\omega)$ be the set of ν 's such that $\omega \leq v$, i = 1, 2. If $\mu \in U_2(\omega)$, then $U_1(\mu) \subset U_2(\omega)$ and hence $U_2(\omega) = \bigcup_{\mu \in U_2(\omega)} U_1(\mu)$. Since Σ_i is the family of arbitrary unions of $U_i(\omega)$'s for i = 1,2, $\Sigma_2 \subset \Sigma_1$. Thus we have proved the following theorem.

<u>Theorem 4.2.</u> Let Ω be a given set, let \leq_1 and \leq_2 be two quasi-orderings on Ω and let Σ_1 and Σ_2 be the complete lattices induced respectively by \leq_1 and \leq_2 . Then \leq_1 is finer than \leq_2 if an only if $\Sigma_2 \subset \Sigma_1$.

The relation "finer than" on quasi-orderings of Ω is by itself a partial ordering. The finest quasi-ordering is the one induced by the power set of Ω , and the least fine quasi-ordering is the one induced by $\{\Omega, \phi\}$. Let $\{\leq a \in \Lambda\}$ be a collection of quasiorderings on Ω , and let $\Delta = \{ \leq : \leq \text{ is finer than } \leq_{\alpha} \text{ for each } \}$ $a \in \Lambda$ }. We claim that Δ has a unique minimal element \leq_m the sense that $\leq \epsilon \Delta$ implies \leq is finer than $\leq \dots$. For each $a \in \Lambda$, let Σ_a be the complete lattice induced by \leq_a . Let $\Phi = \{\Sigma \colon \Sigma \text{ is a complete lattice, } \Sigma \supseteq \Sigma_{\mathfrak{a}} \text{ for each } \mathfrak{a} \in \Lambda\}$ and let $\Sigma_m = \bigcap_{\Phi} \Sigma$. Then Σ_m is a complete lattice. The quasi-ordering \leq_{m} induced by \sum_{m} is finer than \leq_{a} for each $a \in \Lambda$. If \leq is finer than \leq_{a} for each $a \in \Lambda$, then $\Sigma \supset \Sigma_{a}$ for each $a \in \Lambda$ where Σ is the complete lattice induced by \leq . Since $\Sigma \supset \Sigma_m$, \leq is finer than \leq_m . Similarly, if $\Delta = \{\leq : \leq_a \text{ is finer than } \leq d \in \{ \leq : \leq_n \}$ for each $a \in \Lambda$, then Δ has a unique maximal element $\leq m$ in the sense that $\leq \epsilon \Delta$ implies $\leq m$ is finer than $\leq \infty$. Thus, we have proved the following theorem.

<u>Theorem 4.3.</u> Let Ω be a given set. The relation "finer than" on quasi-orderings of Ω is a partial ordering such that every non-empty collection of quasi-orderings of Ω has a greatest lower bound and a least upper bound with respect to the partial ordering "finer than".

Let \leq_1 and \leq_2 be two quasi-orderings induced respectively by complete lattices Σ and Σ^c . Then $\mu \leq_1 \nu$ if and only if $\nu \leq_2 \mu$ for each pair μ and ν in Ω . If $U \in \Sigma \cap \Sigma^c$, then U and U^c are unrelated with respect to \leq_1 . A complete lattice Σ is a field if and only if $U(\omega)$ is an equivalence class for each ω , i.e., $\nu \in U(\mu)$ implies $\mu \in U(\nu)$ for each pair μ and ν where $U(\omega)$ is the set of all ν 's such that $\omega \leq \nu$ with the quasi-ordering \leq induced by Σ . A quasi-ordering \leq is partially ordered if and only if $U(\mu) = U(\nu)$ implies $\mu = \nu$ for each pair μ and ν . A partial ordering has a tree structure if and only if for each pair μ and ν , we have $U(\mu) \cap U(\nu) = \phi$, $U(\mu) \subset U(\nu)$ or $U(\nu) \subset U(\mu)$. A tree structured partial ordering is linearly ordered if and only if $U(\mu) \cap U(\nu) \neq \phi$ for each pair μ and ν .

When Ω is finite, every lattice is a complete lattice. Let $\Omega = R$ and let \leq be the natural ordering of real numbers. The complete lattice corresponding to \leq is the σ -lattice generated by $(\alpha, +\infty)$ for all real α . Let $\Omega = R^2$, let \leq be the partial ordering defined by $(\alpha, \beta) \leq (a, b)$ if $\alpha \leq a$ and $\beta \leq b$, let Σ be the complete lattice corresponding to \leq , let $U \in \Sigma$ with $U \neq \phi$ and let f be the extended real-valued function defined on R such that $f(x) = \inf\{y: (x, y) \in U\}$. It is clear by the definition that if $(x, y) \in U$ and $(x, y) \leq (\alpha, \beta)$ then $(\alpha, \beta) \in U$. Let x_1 and x_2 be two real numbers such that $x_1 < x_2$. Since $y > f(x_1)$ implies $(x_1, y) \in U$ and $(x_2, y) \in U$, $f(x_1) \geq f(x_2)$ and hence f is monotone decreasing. It follows that $\{(x, y): y > f(x)\} \subset U \subset \{(x, y): y \geq f(x)\}$. Let Σ_0 be the family of Borel measurable sets in Σ . Since the family of Borel measurable sets in \mathbb{R}^2 is a σ -lattice, Σ_0 is a σ -lattice. It is clear that the partial ordering induced by Σ_0 is the same as that induced by Σ . It follows that Σ_0 is the σ -lattice generated by $U(\alpha, \beta) = \{(a, b): a \geq \alpha, b \geq \beta\}$ for all $(\alpha, \beta) \in \mathbb{R}^2$.

As far as measurability is concerned, our interest will be in σ -lattices rather than complete lattices. An element of a σ -lattice Σ is an upper set with respect to the quasi-ordering induced by Σ . However, an upper set with respect to the quasi-ordering induced by a σ -lattice Σ need not be in Σ .

A σ -lattice is a monotone class. By a proof analogous to that of the Monotone Class Theorem, one can show that a monotone class which contains a lattice Σ_0 contains the σ -lattice generated by Σ_0 . It follows that a monotone class which is a lattice is a σ -lattice. Let Γ be a family of subsets of Ω . The σ -lattice Σ generated by Γ can be described as follows. Since Ω and ϕ are in Σ and the σ -lattice generated by $\Gamma \cup \{\Omega, \phi\}$ is Σ , without loss of generality we may assume Ω and ϕ in Γ . Let $\Gamma_0 = \Gamma$ and for each ordinal number $\alpha > 0$, let Γ_{α} be defined inductively by

$$\Gamma_{a} = \{(\bigcup_{m=1}^{\infty} A_{m}) \cup (\bigcap_{n=1}^{\infty} B_{n}): A_{n}, B_{m} \in \bigcup_{\gamma < a} \Gamma_{\gamma}\}.$$

Let $\Lambda = \bigcup_{\alpha < \beta} \Gamma_{\alpha}$ where β is the first uncountable ordinal number. Since $\Gamma_{\alpha} \subset \Sigma$ implies $\Gamma_{\alpha+1} \subset \Sigma$ and since $\Gamma_{0} \subset \Sigma$, by the transfinite induction $\Lambda \subset \Sigma$. On the other hand, if $\{C_n\} \subset \Lambda$, then for each n there is $a_n < \beta$ such that $C_n \in \Lambda_a$. Let $a = \sup_n a_n$. Then $a < \beta$ and $C_n \in \Lambda_a$ for each n. Therefore, $\bigcup_{n=1}^{\infty} C_n$ and $\bigcap_{n=1}^{\infty} C_n$ are in $\Gamma_{\alpha+1}$ and hence they are in Λ . It follows that Λ is a σ -lattice containing Γ and thus $\Sigma = \Lambda$.

A σ -lattice Σ is said to be <u>linearly ordered</u> if whenever $U_1, U_2 \in \Sigma$ then either $U_1 \subset U_2$ or $U_2 \subset U_1$. If Σ is a linearly ordered σ -lattice, then so is Σ^c and in such a situation $\Sigma \cap \Sigma^c$ is the trivial σ -field $\{\Omega, d\}$. If Ω is a linearly ordered set, then the collection of all upper sets is a linearly ordered σ -lattice. However, the ordering induced by a linearly ordered σ -lattice need not be linear. Let (Ω, F, P) be a <u>probability space</u>, i.e., F is a σ -field of subsets of Ω and P is a probability measure on F. A <u>random variable</u> is an equivalence class of extended real-valued F-measurable functions defined on Ω such that each pair of functions in the class differ by a set of P measure zero. Therefore, the term "almost everywhere" will be omitted from the context. Whenever we say that a random variable satisfies a property, it means that one of its representations satisfies that property. It is implicit that if a random variable is integrable, then it is finite.

Let Σ be a sub- σ -lattice of F. A random variable X is said to be Σ -measurable if $[X \ge a] \in \Sigma$ for each real number α or equivalently $[X > a] \in \Sigma$ for each real number a. The family of Σ -measurable random variables is denoted by $R(\Sigma)$. Indicators of upper sets in Σ are Σ -measurable. If X is a random variable, then the family of all sets $[X \ge a]$ and [X > a] for all extended real numbers a is a σ -lattice. Such a σ -lattice, denoted by $\Sigma(\mathbf{X})$, is said to be induced by X. The σ -lattice $\Sigma(X)$ is linearly ordered. The family of all sets $[X \in B]$ for all Borel measurable sets in the extended real line is a σ -field. Such a σ -field, denoted by F(X), is said to be <u>induced</u> by X. It is obvious that the intersection of all σ -fields containing $\Sigma(X)$ is the σ -field F(X). Therefore F(X) is the σ -field generated by $\Sigma(\mathbf{X})$.

93

If X_1 and X_2 are Σ -measurable random variables, then so are $X_1 + X_2$, $X_1 \lor X_2$, $X_1 \land X_2$ and δX_1 for each $\delta \ge 0$ (cf. Barlow and coworkers (1972)). Let $\{X_n\}$ be a monotone increasing sequence of Σ -measurable random variables and let X be defined by $X(\omega) = \lim_{n \to \infty} X_n(\omega)$ for each ω . Then $X = \bigvee_{n=1} X_n$ and $[X \leq a] = \bigcap_{n=1}^{\infty} [X_n \leq a]$ for each a. Therefore, $[X > a] = \bigcup_{n=1}^{\infty} [X_n > a] \in \Sigma$ for each a and X is Σ -measurable. Similarly, the limit of a monotone decreasing sequence of Σ -measurable random variables is Σ -measurable. It follows that the upper limit and the lower limit of any sequence of Σ -measurable random variables are Σ -measurable. In particular, the limit of a pointwise convergent sequence of Σ -measurable random variables is Σ measurable. Thus, we have shown that $R(\Sigma)$ is a convex cone which is closed under countable meet, closed under countable join and closed under pointwise convergence.

A random variable is <u>simple</u> if its range is finite. A simple random variable is Σ -measurable if and only if it is a finite linear combination of indicators of upper sets in Σ such that the coefficients are non-negative except the one that corresponds to 1. A non-negative Σ -measurable random variable is the limit of a nondecreasing sequence of simple Σ -measurable random variables. An arbitrary Σ -measurable random variable X is the limit of a sequence of some Σ -measurable random variables {X is the limit of a $|X_n| \leq |X|$ for each n. Let X be a Σ -measurable random variable and let f be a non-decreasing function defined on the extended real line. Then the random variable f $\circ X$ is Σ -measurable.

IV.2 Conditional Expectation as a Generalized Projection

Let $L_1(\Omega, F, P)$ and $L_2(\Omega, F, P)$ be the linear spaces of integrable random variables and of square-integrable random variables respectively. When there is no ambiguity, we shall use L_1 and L_2 . The linear space L_2 is a subspace of L_1 . Let E be the operator defined by

$$(4.1) EXY = \int XYdP$$

for each pair of random variables X and Y, provided that the integral on the right-hand side exists. If the random variable Y is the constant random variable 1, then we shall use EX, the integral of X, instead of EXY when the latter exists. The operator E restricted to $L_2 \times L_2$ is an inner product. The linear space L_2 with the inner product E is known to be a Hilbert space. The norm of $X \in L_2$, ||X||, is the square root of EX^2 .

The convex cones $L_1 \cap R(\Sigma)$ and $L_2 \cap R(\Sigma)$ are denoted

briefly by $L_1(\Sigma)$ and $L_2(\Sigma)$ respectively. Let $\{Z_n\}$ be a sequence in $L_2(\Sigma)$ which <u>converges to Z in L</u>₂, i.e., $||Z_n-Z|| \rightarrow 0$ as $n \rightarrow \infty$. Then $Z \in L_2$ and there exists a subsequence $\{Z_{n'}\}$ which converges to Z pointwise (cf. Ash (1972)). It follows that $Z \in R(\Sigma)$ and hence $L_2(\Sigma)$ is a closed convex cone in L_2 . The uniqueness and existence of the projection $P(X|L_2(\Sigma))$ for each $X \in L_2$ follows from Theorem 2.1. Theorem 2.7 shows that a random variable $X^* \in L_2(\Sigma)$ is $P(X|L_2(\Sigma))$ if and only if $E[(X-X^*)X^*] = 0$ and $E[(X-X^*)Z] \leq 0$ for each $Z \in L_2(\Sigma)$. Let $X^* = P(X|L_2(\Sigma))$. Since constant random variables are in $L_2(\Sigma)$, we have

(4.2)
$$EX^* = EX$$
.

For each $Z \in L_2(\Sigma)$, there is a sequence of simple random variaables $\{Z_n\} \subset L_2(\Sigma)$ and $|Z_n| \leq |Z|$ for each n such that Z_n converges to Z pointwise. For each n, $|(X-X*)Z_n| \leq |X-X*||Z|$. Since |X-X*||Z| is integrable, by the Dominated Convergence Theorem $E[(X-X*)Z] = \lim_{n \to \infty} E[(X-X*)Z_n]$. It follows that the conn $\to \infty$ dition $E[(X-X*)Z] \leq 0$ for each $Z \in L_2(\Sigma)$ is equivalent to $E[(X-X*)I_U] \leq 0$ for each upper set $U \in \Sigma$. Brunk (1965) showed that if $X* = P(X|L_2(\Sigma))$ and $g(X*) \in L_2$ for a real-valued function g, then E[(X-X*)g(X*)] = 0. In particular, $E[(X-X*)I_B(X*)] = 0$ for each Borel set B Therefore, a necessary and sufficient condition for $P(X | L_2(\Sigma))$ can be stated as follows.

<u>Theorem 4.4.</u> A random variable $X^* \in L_2(\Sigma)$ is $P(X | L_2(\Sigma))$ if and only if X^* satisfies

(4.3)
$$E[(X-X*)]_{B}(X*)] = 0$$
 for each Borel set B

and

(4.4)
$$\mathbb{E}[(X-X^*)1_{U}] \leq 0$$
 for each $U \in \Sigma$.

Theorem 2.3 shows that $E(X-Z)^2 \ge E(X-X*)^2 + E(X*-Z)^2$ for each $Z \in L_2(\Sigma)$. By (4.2), we have E(X-X*) = 0 and hence E(X-Z) = E(X*-Z). It follows that

$$\operatorname{Var}(Z-Z) \ge \operatorname{Var}(X-X^*) + \operatorname{Var}(X^*-Z)$$
 for each $Z \in L_2(\Sigma)$.

In particular, $\operatorname{Var} X \ge \operatorname{Var} X^* + \operatorname{Var}(X - X^*)$ where the <u>variance</u> of a random variable Z is defined by $\operatorname{Var} Z = E(Z - EZ)^2$. Theorem 2.7 and (4.2) show that

(4.5)
$$Cov(X, X^*) = Var X^*$$

and

(4.6)
$$\operatorname{Cov}(X, Z) \leq \operatorname{Cov}(X^*, Z)$$
 for each $Z \in L_2(\Sigma)$

where the <u>covariance</u> of a pair of random variables X and Z is defined by Cov(X, Z) = E(X-EX)(Z-EZ).

If $Cov(X, Z) \leq 0$ for each $Z \in L_2(\Sigma)$, then by Theorem 4.4

the projection $P(X \mid L_2(\Sigma))$ is the constant random variable EX. Let $X \in L_2$, let $Y \in L_2(\Sigma)$ and let $X^* = P(X \mid L_2(\Sigma))$. If $E[(X-Y)1_B(Y)] = 0$ for each Borel set B, then E(X-Y)Y = 0 and EX = EY. Since $E[(X-X^*)Y] \leq 0$, we have $EY^2 = EXY \leq EX^*Y \leq (EX^*^2EY^2)^{1/2}$ and hence $EY^2 \leq EX^*^2$. By the fact $EX^* = EY$, $Var Y \leq Var X^*$. Let $T \in L_2$. If $E[(X-T)1_U] \leq 0$ for each $U \in \Sigma$, then $E[(X-T)Z] \leq 0$ for each $Z \in L_2(\Sigma)$. In particular, $E[(X-T)X^*] \leq 0$. By the fact that $EX^*^2 = EXX^*$, we have $EX^*^2 = EXX^* \leq ETX^* \leq (EX^*^2ET^2)^{1/2}$ and hence $EX^{*2} \leq ET^2$.

If Z_1 and Z_2 are in $L_2(\Sigma)$, then so are $Z_1 \lor Z_2$ and $Z_1 \land Z_2$. Since $X^* + Z = X^* \lor Z + X^* \land Z$ for each $Z \in L_2$, by (4.5) we have $Cov(X - X^*, Z) = Cov(X - X^*, X^* \lor Z) + Cov(X - X, X^* \land Z)$. For each $Z \in L_2(\Sigma)$, by (4.6) we have $Cov(X - X^*, Z) \leq 0$, $Cov(X - X^*, X^* \lor Z) \leq 0$ and $Cov(X - X^*, X^* \land Z) \leq 0$. It follows that $Cov(X - X^*, Z) \leq Cov(X - X^*, X^* \lor Z)$ and $Cov(X - X^*, Z) \leq Cov(X - X^*, X^* \land Z)$. It is clear that $|X^* - Z_{\lor} X^*| \leq |X^* - Z|$ and $|X^* - Z \land X^*| \leq |X^* - Z|$. By the identity $E(X - Z)^2 = E(X - X^*)^2 + E(X^* - Z)^2 - 2E[(X - X^*)Z]$ and the identities when Z is replaced respectively by $Z \lor X^*$ and $Z \land X^*$, we have

(4.7)
$$E(X-Z_{\sqrt{X}})^{2} \leq E(X-Z)^{2} \text{ and } E(X-Z_{\sqrt{X}})^{2} \leq E(X-Z)^{2}$$

for each $Z \in L_2(\Sigma)$.

By (4.2) and by an argument similar to that described above, $C(X-X^*, Z) \leq Cov(X-X^*, Z \lor \gamma) \leq 0$ and $Cov(X-X^*, Z) \leq Cov(X-X^*, Z \land \gamma) \leq 0$ for each $Z \in L_2(\Sigma)$ and for each real number γ . If $\alpha \leq X \leq \beta$, then by the monotonicity in Theorem 4.5 we have $\alpha \leq X^* \leq \beta$. Suppose $\alpha \leq X^* \leq \beta$ for a pair of real number α and β . It is obvious that $|X^*-Z \lor \alpha|$, $|X^*-Z \land \beta|$ are bounded by $|X^*-Z|$. By a proof analogous to that of (4.7), we have

(4.8)
$$E(X-Z \checkmark \alpha)^2 \leq E(X-Z)^2 \text{ and } E(X-Z \land \beta)^2 \leq E(X-Z)^2$$

for each $Z \in L_2(\Sigma)$. Consequently, $E[X-(Z \vee \alpha) \wedge \beta]^2 \leq E(X-Z)^2$.

 $(2.6) \text{ shows that } \mathbf{E}(\mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma)) - \mathbf{P}(\mathbf{Y} \mid \mathbf{L}_{2}(\Sigma)))^{2} \leq \mathbf{E}(\mathbf{X} - \mathbf{Y})^{2} \text{ and}$ hence $\operatorname{Var}(\mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma)) - \mathbf{P}(\mathbf{Y} \mid \mathbf{L}_{2}(\Sigma))) \leq \operatorname{Var}(\mathbf{X} - \mathbf{Y}) \text{ for each pair } \mathbf{X},$ $\mathbf{Y} \in \mathbf{L}_{2}.$ Let Σ_{0} be a sub- σ -lattice of Σ . Then $\mathbf{L}_{2}(\Sigma_{0})$ is a closed convex cone contained in $\mathbf{L}_{2}(\Sigma)$. (2.7) shows that $\mathbf{E}(\mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma)) - \mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma_{0})))^{2} \leq \mathbf{E}(\mathbf{X} - \mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma_{0})))^{2} - \mathbf{E}(\mathbf{X} - \mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma)))^{2},$ (2.17) shows that $\operatorname{Var}(\mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma)) - \mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma_{0}))) \leq \operatorname{Var}(\mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma)))$ $- \operatorname{Var}(\mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma_{0}))) \text{ and } (2.16) \text{ shows that}$ $\operatorname{Var} \mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma_{0})) \leq \operatorname{Var} \mathbf{P}(\mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma)) \mid \mathbf{L}_{2}(\Sigma_{0})) \leq \operatorname{Var} \mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma)) \leq \operatorname{Var} \mathbf{X}.$ Some other properties of the generalized projection $\mathbf{P}(\mathbf{X} \mid \mathbf{L}_{2}(\Sigma))$ will be given in Section IV.4.

IV. 3 Conditional Expectation Given a σ -Lattice

Conditional expectation given a σ -lattice is an extension of generalized projection on the closed convex cone $L_2(\Sigma)$. Let X be a random variable defined on the probability space (Ω, F, P) and let Σ be a sub- σ -lattice of F. A random variable $X^* \in R(\Sigma)$ is said to be the <u>conditional expectation</u> of X given Σ if X^* satisfies (4.3) and (4.4). The conditional expectation of X given Σ need not exist. If it exists, then it is unique (cf. Theorem 4.6); such a random variable is denoted by $E(X \mid \Sigma)$.

If $X \in L_1$, then by Theorem 4.7, $E(X \mid \Sigma)$ exists. Let $X^* = E(X \mid \Sigma)$ for an $X \in L_1$. The conditions (4.3) and (4.4) can be represented as

(4.9)
$$\int_{[X*\in B]} X \, dP = \int_{[X*\in B]} X* \, dP \text{ for each Borel set } B$$

and

(4.10)
$$\int_{U} X \, dP \leq \int_{U} X^* \, dP \quad \text{for each } U \in \Sigma.$$

It follows that $X^* \in L_1$ and $EX = EX^*$. (4.9) shows that X^* is the Radon-Nikodym derivative of the measure μ with respect to P restricted to the σ -field $F(X^*)$ where $\mu(A) = E(X \mid_A)$ for each $A \in F$. (4.10) implies that $EXZ \leq EX^*Z$ for every $Z \in R(\Sigma)$ provided that both integrals exist. If Σ is by itself a σ -field, then $EXI_U = EX*I_U$ for each $U \in \Sigma$ and hence X* is the Radon-Nikodym derivative of the measure μ given above with respect to P restricted to Σ . It follows that conditional expectation given a σ -field is a special case of conditional expectation given a σ -lattice.

<u>Example 4.1.</u> Let $\Omega = (-1/2, 1/2)$, let F be the family of all Borel subsets of Ω , let P be the Lebesgue measure on F and let Σ be the σ -lattice induced by Z where $Z(\omega) = \omega$ for each $\omega \in \Omega$. Let X be the random variable defined by $X(\omega) = 1/|\omega|$ if $\omega \neq 0$ and X(0) = 0. By the definition, one can show that the conditional expectation of X given Σ does not exist.

Another way to see the non-existence of $E(X \mid \Sigma)$ is by (4.13). Let $Z_s = \inf\{(t-s)^{-1}(EX1_{(s,t)}): s < t < 1/2\} I_{(s,1/2)}$ and let $Y_s = \sup\{(s-r)^{-1}(EX1_{(r,s)}): -1/2 < r < s\} I_{(-1/2,s)} + \infty I_{(s,1/2)}$ for each $s \in \Omega$. For s > 0, $Z_s = (-\ln 2s)^{-1} I_{(s,1/2)}$ and $Y_s = +\infty$ For s < 0, $Z_s = |s|^{-1} I_{(s,1/2)}$ and $Y_s = |s|^{-1} I_{(-1/2,s)} + \infty I_{(s,1/2)}$. For s = 0, $Z_0 = +\infty I_{(0,1/2)}$ and $Y_0 = +\infty$. Let $Z = \bigvee_{s \in \Omega} Z_s$ and let $Y = \bigwedge_{s \in \Omega} Y_s$. Then Y = Zand $Z(\omega) = X(\omega)$ if $\omega < 0$ and $Z(\omega) = +\infty$ if $\omega \ge 0$. If $E(X \mid \Sigma)$ exists, then it must be Z. However, $E(X-Z) = -\infty$. Therefore $E(X \mid \Sigma)$ does not exist.

For each $s \in (-1/2, 0)$, let Y_s be the random variable defined by $Y_s(\omega) = X(\omega)$ if $\omega \le s$ and $Y_s(\omega) = +\infty$ if $\omega > s$. Then $Y_s \in R(\Sigma)$ and Y_s satisfies (4.9) and (4.10) for each s $\in (-1/2, 0)$. Therefore, conditions (4.3) and (4.4) for $E(\mathbf{X} \mid \Sigma)$ can not be replaced by (4.9) and (4.10). []

Let X^* be the conditional expectation of X given Σ . For each real number a, we have

$$\int (\mathbf{X} - \mathbf{X}) d\mathbf{P} = 0$$

and hence for each $L \in \Sigma^{C}$,

$$\begin{split} \int_{L} \sum_{n} [X \ast > \alpha] &= \int_{[X \ast > \alpha]} (X - X \ast) dP = \int_{L} c \sum_{n} (X - X \ast) dP \\ &= - \int_{L} c \sum_{n} [X \ast > \alpha] (X - X \ast) dP. \end{split}$$

Since $L^{C} \cap [X^{*>a}] \in \Sigma$, by (4.4) we have

(4.11)
$$\int_{L \cap [X^{*} > a]} X dP \ge \int_{L \cap [X^{*} > a]} X^{*} dP.$$

Similarly, if $U \in \Sigma$ then we have

(4.12)
$$\int_{U \cap [X^{*} < \alpha]} X dP \leq \int_{U \cap [X^{*} < \alpha]} X^{*} dP$$

for each real number a. The inequality (4.11) was introduced by
Brunk (1965).

<u>Theorem 4.5.</u> Let $X_i^* = E(X_i | \Sigma)$, i = 1, 2 for a pair of random variables X_1 and X_2 such that $X_1 \leq X_2$. Then $X_1^* \leq X_2^*$.

<u>Proof.</u> We shall show that $P(X_1^* > X_2^*) = 0$. The event $[X_1^* > X_2^*]$ can be represented by $[X_1^* > X_2^*] = \bigcup_{\alpha \text{ rational}} [X_1^* > \alpha > X_2^*]$. Therefore, we need only to show that $P(X_1^* > \alpha > X_2^*) = 0$ for each rational number α . By (4.11) and (4.12), we have

$$\begin{split} \int_{[\mathbf{X}_{2}^{*} < \alpha] \cap [\mathbf{X}_{1}^{*} > \alpha]} \mathbf{X}_{1}^{*} d\mathbf{P} &\leq \int_{[\mathbf{X}_{2}^{*} < \alpha] \cap [\mathbf{X}_{1}^{*} > \alpha]} \mathbf{X}_{1} d\mathbf{P} \\ &\leq \int_{[\mathbf{X}_{2}^{*} < \alpha] \cap [\mathbf{X}_{1}^{*} > \alpha]} \mathbf{X}_{2} d\mathbf{P} \\ &\leq \int_{[\mathbf{X}_{2}^{*} < \alpha] \cap [\mathbf{X}_{1}^{*} > \alpha]} \mathbf{X}_{2}^{*} d\mathbf{P} \end{split}$$

for each real number α . Since $X_2^* - X_1^* < 0$ on $[X_1^* > \alpha > X_2^*]$, it follows that $P(X_2^* < \alpha < X_1^*) = 0$. This completes the proof. []

The above property is called <u>monotonicity</u>, which appears in Barlow and coworkers (1972).

<u>Theorem 4.6.</u> Let X be a random variable. The conditional

expectation of X given Σ is unique provided that it exists.

<u>Proof.</u> Suppose Y_1 and Y_2 are two Σ -measurable random variables such that each satisfies (4.3) and (4.4). The inequalities (4.11) and (4.12) hold when we replace X^* by Y_1 or Y_2 . Since $X \leq X$, by a similar argument in the proof of Theorem 4.5 we have $P(Y_1 > Y_2) = 0$ and $P(Y_2 > Y_1) = 0$. Therefore, $Y_1 = Y_2$. []

The uniqueness of $E(\mathbf{X} \mid \Sigma)$ can also be obtained by (4.13). The proof of the theorem is through communication with Professor H.D. Brunk. The following existence theorem is given by Brunk (1963).

<u>Theorem 4.7.</u> If $X \in L_1$, then the conditional expectation of X given Σ exists.

Let $\{Z_{\alpha}: \alpha \in \Lambda\}$ be a collection of random variables. The random variable Z is said to be the <u>essential supremum</u> of $\{Z_{\alpha}: \alpha \in \Lambda\}$ if for any random variable Y, $Z \leq Y$ if and only if $Z_{\alpha} \leq Y$ for each $\alpha \in \Lambda$. Similarly for the definition of the <u>essential</u> <u>infimum</u> of $\{Z_{\alpha}: \alpha \in \Lambda\}$. They are denoted by $\bigvee_{\alpha \in \Lambda} Z_{\alpha}$ and $\bigwedge_{\alpha \in \Lambda} Z_{\alpha}$ respectively. It is known that any family of random variables has an essential supremum and an essential infimum and $\Lambda_{a \in \Lambda} Z_a = -V_{a \in \Lambda}(-Z)$. The following identity derived from (4.11) and (4.12) was introduced by Brunk and Johansen (1970).

$$(4.13) \qquad \mathbf{E}(\mathbf{X} \mid \Sigma) = \bigvee_{\mathbf{U} \in \Sigma} \{ \inf_{\substack{\mathbf{L} \in \Sigma^{\mathbf{C}} \\ \mathbf{P}(\mathbf{L}\mathbf{U}) > 0}} [\mathbf{P}(\mathbf{L}\mathbf{U})^{-1}\mathbf{E}\mathbf{X}]_{\mathbf{L}\mathbf{U}}]^{1} \mathbf{U}^{-\infty 1} \mathbf{U}^{\mathbf{C}} \}$$
$$= \bigwedge_{\mathbf{L} \in \Sigma^{\mathbf{C}}} \{ \sup_{\substack{\mathbf{U} \in \Sigma \\ \mathbf{U} \in \Sigma} \\ \mathbf{P}(\mathbf{L}\mathbf{U}) > 0} [\mathbf{P}(\mathbf{L}\mathbf{U})^{-1}\mathbf{E}\mathbf{X}]_{\mathbf{L}\mathbf{U}}]^{1} \mathbf{L}^{+\infty 1} \mathbf{L}^{\mathbf{C}} \}$$
$$= (\mathbf{L}\mathbf{U}) = 0$$

provided that $E(X \mid \Sigma)$ exists.

Let $a \leq X \leq \beta$, let

$$Z = \bigvee_{\substack{U \in \Sigma}} \{ \inf_{\substack{L \in \Sigma^{C} \\ P(LU) > 0}} [P(LU)^{-1}EX_{LU}]_{U}^{+\alpha_{1}} \}$$

and let

$$Y = \bigwedge_{L \in \Sigma^{c}} \{ \sup_{\substack{U \in \Sigma \\ U \in \Sigma \\ P(LU) > 0}} [P(LU)^{-1}EX_{LU}]_{L}^{1} + \beta_{1} \}$$

Then $Z \leq Y$ and (4.13) can be represented as $E(X \mid \Sigma) = Z = Y$. If $Z \neq Y$, then $E(X \mid \Sigma)$ does not exist. If Z = Y, $E(X \mid \Sigma)$ need not exist as shown in Example 4.1.

If Ω if finite, then every upper set with respect to the quasiordering induced by Σ is in Σ . Suppose $P(\{\omega\}) > 0$ for each $\omega \in \Omega$. (4.13) may be represented by the following identity.

(4.14)
$$X * (\omega) = \max \min M(LU)$$

 $\omega \in U \quad \omega \in L$
 $= \min \max M(LU)$
 $\omega \in L \quad \omega \in U$

where X* is the isotonic regression of X, U is an upper set, L is a lower set, M(LU) is the weighted average of X over $L \cap U$ and the weight function W is defined by $W(\omega) = P(\{\omega\})$ for each $\omega \in \Omega$.

The identity (4.14) appears in Barlow and coworkers (1972). Ayer and coworkers (1955) introduced (4.14) for the case that Ω is a linearly ordered set.

IV.4 Properties of Conditional Expectation

Our interest in this section and the following section is the case when random variables are integrable, although most of the properties hold in general. If $X \in L_1$, then X is finite and $E(X \mid \Sigma)$ exists.

It is trivial that $E(X+a|\Sigma) = E(X|\Sigma) + a$ for each real a, $E(\delta X | \Sigma) = \delta E(X | \Sigma)$ for each real $\delta \ge 0$, and $E(X | \Sigma^{C}) = -E(-X | \Sigma)$. The monotonicity shows that if $X_{1} \le X_{2}$, then $E(X_{1} | \Sigma) \le E(X_{2} | \Sigma)$. It follows that $E(X \land Y | \Sigma) \le EX | \Sigma) \le E(X \lor Y | \Sigma)$ for each pair X, $Y \in L_{1}$. By a proof analogous to Theorem 4.5, if $X_{1} \le X_{2}$ then $E(X_{1} | \Sigma) \le E(X_{2} | \Sigma)$. The following three convergence theorems which appear in Barlow and coworkers (1972) are given below without proof.

<u>Theorem 4.8.</u> Let X, $X_n \in L_1$ and let $\{X_n\}$ be a monotone sequence such that it converges to X. Then $\{E(X_n \mid \Sigma)\}$ converges to $E(X \mid \Sigma)$.

 $\frac{\text{Theorem 4.9. If } |X_n| \leq Y \text{ for each } n \text{ with } Y \in L_1 \text{ and}}{\{X_n\} \text{ converging to } X, \text{ then } \{E(X_n | \Sigma)\} \text{ converges to } E(X | \Sigma).}$

<u>Theorem 4.10.</u> Let $\{\Sigma_n\}$ be a monotone sequence of sub- σ -lattices of F and let $X \in L_1$. Then $\{E(X \mid \Sigma_n)\}$ converges to $E(X \mid \Sigma)$ where Σ is the σ -lattice generated by $\bigcup_{n=1}^{\infty} \Sigma_n$ if $\{\Sigma_n\}$ is monotone increasing and $\Sigma = \bigcap_{n=1}^{\infty} \Sigma_n$ if $\{\Sigma_n\}$ is monotone decreasing.

A similar result to Corollary 2.11.1 can be applied to the operator of conditional expectation.

<u>Theorem 4.11</u>. Let $X \in L_1$ and let Σ_1 and Σ_2 be two sub- σ -lattices of F with $\Sigma_1 \subset \Sigma_2$. Then $E(E(X \mid \Sigma_2) \mid \Sigma_1) = E(X \mid \Sigma_1)$ if and only if

$$\int_{U} E(X \mid \Sigma_{2}) dP \leq \int_{U} E(X \mid \Sigma_{1}) dP \text{ for each } U \in \Sigma_{1}.$$

Proof. Let
$$Y_i = E(X | \Sigma_i)$$
, $i = 1, 2$. If $Y_1 = E(Y_2 | \Sigma_1)$,

then by (4.4) we have

$$\int_{U} Y_2 dP \leq \int_{U} Y_1 dP \quad \text{for each} \quad U \in \Sigma_1.$$

Conversely, if the above inequality holds for each $U \in \Sigma_1$, then

$$\int_{[Y_1 > a]} Y_2 dP \leq \int_{[Y_1 > a]} Y_1 dP \quad \text{for each real} \quad a$$

By (4.3) and (4.4) we have

$$\int_{[Y_1>a]} Y_1 dP = \int_{[Y_1>a]} X dP \leq \int_{[Y_1>a]} Y_2 dP \text{ for each real } a$$

since $[Y_1 > a] \in \Sigma_1 \subset \Sigma_2$. Combining the last two inequalities, we have

$$\int_{[Y_1>a]} Y_2 dP = \int_{[Y_1>a]} Y_1 dP \text{ for each real } a$$

It follows that

$$\int_{[Y_1 \in B]} Y_2 dP = \int_{[Y_1 \in B]} Y_1 dP \quad \text{for each Borel set } B.$$

Since Y_1 satisfies (4.3) and (4.4) when we replace X and Σ by Y_2 and Σ_1 respectively, $Y_1 = E(Y_2 | \Sigma)$. This completes the proof. [] <u>Theorem 4.12.</u> Let $X \in L_1$ and let Σ_1 and Σ_2 be two sub- σ -lattices of F with $\Sigma_1 \subset \Sigma_2$. If either Σ_1 or Σ_2 is a σ -field, then $E(E(X | \Sigma_2) | \Sigma_1) = E(X | \Sigma_1)$.

<u>Proof</u>. Let $Y_i = E(X | \Sigma_i)$, i = 1, 2. If Σ_1 is a σ -field, then $U^C \in \Sigma_1$ if $U \in \Sigma_1$. Therefore, if we apply both U and U^C to (4.4) then we have

$$\int_{U} Y_{l} dP = \int_{U} X dP = \int_{U} Y_{2} dP$$

for each $U \in \Sigma_1$ because $\Sigma_1 \subset \Sigma_2$. It follows from Theorem 4.11 that $Y_1 = E(Y_2 | \Sigma_1)$. If Σ_2 is a σ -field, then by a similar argument applying to Σ_2 we have

$$\int_{U} Y_{2} dP = \int_{U} X dP \leq \int_{U} Y_{1} dP \quad \text{for each} \quad U \in \Sigma_{1}.$$

By Theorem 4.11, $Y_1 = E(Y_2 | \Sigma_1)$. []

The theorem which appears in Robertson (1968), is called the <u>smoothing property</u>. It is parallel to Corollary 2.4.1. If F_0 is the trivial σ -field $\{\Omega, \phi\}$, then $E(\mathbf{X} | F_0) = E\mathbf{X}$. Therefore, the operator E can be regarded as conditional expectation given F_0 . By the smoothing property with Σ_1 replaced by F_0 , we have $(4.2), i.e., E(E(\mathbf{X} | \Sigma)) = E\mathbf{X}$. Let I be an open interval in the real line. The monotonicity implies that if $P(X \in I) = 1$ then $P(E(X \mid \Sigma) \in I) = 1$. Let $X^* = E(X \mid \Sigma)$. Then by (4.3), $X^* = E(X \mid F(X^*))$. Therefore the following version of the <u>Jensen's inequality</u> follows (cf. Ash (1972)).

<u>Theorem 4.13.</u> Let $X \in L_1$ and let I be an open interval in the real line such that $P(X \in I) = 1$. If g is a real-valued convex function defined on I, then

$$\mathbf{E}(\mathbf{g}(\mathbf{X}) \mid \mathbf{F}(\mathbf{X}^*)) \geq \mathbf{g}(\mathbf{X}^*)$$

where $X^* = E(X \mid \Sigma)$.

By the smoothing property, we have $Eg(X) \ge Eg(X^*)$, and hence $E|X| \ge E|X^*|$ and $EX^2 \ge EX^{*2}$ where $X^* = E(X|\Sigma)$, provided $X \in L_1$.

The following version of the <u>Jensen's inequality</u> appears in Barlow and coworkers (1972).

<u>Theorem 4.14.</u> Under the same assumption as in Theorem 4.13, if Xg'(X*), X*g'(X*) and g(X*) are in L_1 , then

$$\mathbf{E}(\mathbf{g}(\mathbf{X}) \mid \Sigma) \geq \mathbf{g}(\mathbf{X}^*)$$

where g' is a determination of the derivative of g.

Immediate results of the theorem are

$$\mathbb{E}(\mathbf{X} \mid \Sigma) \mid \leq \mathbb{E}(\mid \mathbf{X} \mid \mid \Sigma) \text{ and } (\mathbb{E}(\mathbf{X} \mid \Sigma))^2 \leq \mathbb{E}(\mathbf{X}^2 \mid \Sigma).$$

Let $\Gamma \in F$ such that $P(\Gamma) > 0$, let $F | \Gamma := \{A \cap \Gamma : A \in F\}$ and let P_{Γ} be defined by $P_{\Gamma}(A) := P(A \cap \Gamma)/P(\Gamma)$ for each $A \in F$. Then $(\Gamma, F | \Gamma, P_{\Gamma})$ is a probability space. Let $\Sigma | \Gamma := \{U \cap \Gamma : U \in \Sigma\}$. Then $\Sigma | \Gamma$ is a sub- σ -lattice of $F | \Gamma$. Let $R(\Sigma | \Gamma)$ be the family of $\Sigma | \Gamma$ -measurable random variables defined on $(\Gamma, F | \Gamma, P_{\Gamma})$. The <u>restricted conditional expectation</u> of X given Σ to Γ is the conditional expectation of $X | \Gamma$ given $\Sigma | \Gamma$ with respect to $(\Gamma, F | \Gamma, P_{\Gamma})$. Such a random variable is denoted by $E(X | \Sigma, \Gamma)$ if it exists.

Let $X^* = E(X \mid \Sigma)$ and let $\Gamma = [X^* > \alpha]$ for some real α such that $P(\Gamma) > 0$. It is obvious that

$$\int_{[X*\epsilon B]\cap\Gamma} XdP = \int_{[X*\epsilon B]\cap\Gamma} X*dP \text{ for each Borel set } B.$$

and

$$\int_{U \cap \Gamma} XdP \leq \int_{U \cap \Gamma} X*dP \text{ for each } U \in \Sigma$$

provided that $X \in L_1$. It follows that $X^* | \Gamma = E(X | \Sigma, \Gamma)$.

Suppose $\Gamma \in \Sigma \cap \Sigma^{C}$ such that $P(\Gamma)P(\Gamma^{C}) > 0$. By (4.11) and (4.4), we have

$$\int_{[X^*>a] \cap \Gamma} XdP = \int_{[X^*>a] \cap \Gamma} X^*dP \text{ for each real } a$$

and hence

$$\int_{[X*\epsilon B]\cap\Gamma} XdP = \int_{[X*\epsilon B]\cap\Gamma} X*dP$$
for each Borel set B.

It is trivial that

$$\int_{U \cap \Gamma} X dP \leq \int_{U \cap \Gamma} X * dP \text{ for each } U \in \Sigma.$$

Therefore $X * | \Gamma = E(X | \Sigma, \Gamma)$ and similarly $X * | \Gamma^{c} = E(X | \Sigma, \Gamma^{c})$. It follows that for each $X \in L_{1}$ we have

(4.15)
$$E(\mathbf{X} \mid \boldsymbol{\Sigma}) = E(\mathbf{X} \mid \boldsymbol{\Sigma}, \boldsymbol{\Gamma})^{1} + E(\mathbf{X} \mid \boldsymbol{\Sigma}, \boldsymbol{\Gamma}^{c})^{1} \Gamma^{c}$$

where $E(X \mid \Sigma, \Gamma)_{\Gamma}^{1}$ and $E(X \mid \Sigma, \Gamma^{c})_{\Gamma}^{1}$ are extensions of Γ^{c} $E(X \mid \Sigma, \Gamma)$ and $E(X \mid \Sigma, \Gamma^{c})$ to Ω .

<u>Theorem 4.15.</u> Let $U \in \Sigma$ and let $L \in \Sigma^{c}$. Then

 $E(X \mid \Sigma, U) \leq E(X \mid \Sigma) \mid U$ and $E(X \mid \Sigma, L) \geq E(X \mid L) \mid L$

provided that $P(U)P(U^{c})P(L)P(L^{c}) > 0$. If any equality holds, then $E(X \mid \Sigma)$ can be represented by (4.15) with Γ replaced by U or L according as $E(X \mid \Sigma, U) = E(X \mid \Sigma) \mid U$ or $E(X \mid \Sigma, L) = E(X \mid \Sigma) \mid L$. <u>Proof</u>. This is a similar proof to that of Theorem 4.5. Let $X * = E(X | \Sigma), Y_1 = E(X | \Sigma, U)$ and $Y_2 = X * | U$. We are going to show $P_U(Y_1 > a > Y_2) = 0$ for each real a. The first inequality will then follow. Similarly for the second inequality. The relation between P and P_U is that $dP_U = P(U)^{-1} l_U dP$.

By (4.11) and (4.12), we have

$$\begin{split} \int_{[Y_1 > \alpha > Y_2]} Y_1 dP_U &\leq \int_{[Y_1 > \alpha > Y_2]} X dP_U \\ &= P(U)^{-1} \int_{[Y_1 > \alpha > Y_2]} X dP \\ &= P(U)^{-1} \int_{[Y_1 > \alpha] \frown [X * < \alpha]} X dP \\ &\leq P(U)^{-1} \int_{[Y_1 > \alpha] \frown [X * < \alpha]} X * dP \\ &\leq P(U)^{-1} \int_{[Y_1 > \alpha] \frown [X * < \alpha]} X * dP \\ &= \int_{[Y_1 > \alpha > Y_2]} Y_2 dP_U \end{split}$$

for each real
$$\alpha$$
, because $\{\omega: Y_1(\omega) > \alpha\} \in \Sigma$ and
 $[Y_1 > \alpha] \cap [X \ll \alpha] = [Y_1 > \alpha > Y_2]$. Since $Y_2 - Y_1 < 0$ on
 $[Y_1 > \alpha > Y_2]$, $P_U(Y_1 > \alpha > Y_2) = 0$.
Suppose $Y_1 = Y_2$. Then
 $E(X \mid \Sigma) = X \approx 1_U + X \approx 1_U \leq E(X \mid \Sigma, U) = 1_U + E(X \mid \Sigma, U^c) = 1_U^c$.

Applying E to both sides of the inequality, by the monotonicity of E we have

$$E\mathbf{X} \leq E\{E(\mathbf{X} \mid \Sigma, U)_{U}\} + E\{E(\mathbf{X} \mid \Sigma, U^{C})_{U}\}$$
$$= E(\mathbf{X}_{U}) + E(\mathbf{X}_{U})$$
$$= E\mathbf{X}.$$

It follows that $E(\mathbf{X} \mid \Sigma) = E(\mathbf{X} \mid \Sigma, U)_{U}^{1} + E(\mathbf{X} \mid \Sigma, U^{c})_{U}^{1}$. []

<u>Corollary 4.15.1.</u> Let $U \in \Sigma$, $L \in \Sigma^{c}$ such that $P(U)P(U^{c})P(L)P(L^{c}) > 0$ and let $X \ge 0$. Then

$$\mathbf{E}(\mathbf{X}_{U} \mid \Sigma) = \mathbf{E}(\mathbf{X} \mid \Sigma, U)_{U} \leq \mathbf{E}(\mathbf{X} \mid \Sigma)_{U}$$

and

$$\mathbb{E}(\mathbf{X}_{L} \mid \Sigma) = \mathbb{E}(\mathbf{X} \mid \Sigma, L)_{L} \ge \mathbb{E}(\mathbf{X} \mid \Sigma)_{L}.$$

<u>Proof</u>. It is trivial that $E(Xl_U | \Sigma, U^c) = 0$. Let Y = $E(Xl_U | \Sigma)$. Then

$$\mathbf{EX1}_{\mathrm{U}} = \int_{\mathrm{U}} \mathbf{XdP} \leq \int_{\mathrm{U}} \mathbf{YdP} \leq \int \mathbf{YdP} = \mathbf{EY}$$

Since $E(\mathbf{X}_{U}^{1}) = EY$ and $Y \ge 0$, $Y | U^{c} = 0$, i.e., $E(\mathbf{X}_{U}^{1} | \Sigma) | U^{c} = 0$. By Theorem 4.15,

$$\mathbf{E}(\mathbf{X}_{U} \mid \Sigma) = \mathbf{E}(\mathbf{X} \mid \Sigma, U)_{U}$$
.

The monotonicity implies that $E(X|_U | \Sigma) \leq E(X | \Sigma)$. The first statement will then follow. Similarly we have the second statement. []

<u>Corollary 4.15.2.</u> If $X \ge \alpha$ on U for a set $U \in \Sigma$, then $E(X \mid \Sigma) \ge \alpha$ on U. If $X \le \beta$ on L for a set $L \in \Sigma^{c}$, then $E(X \mid \Sigma) \le \beta$ on L.

<u>Proof.</u> If $X | U \ge a$, then by the monotonicity we have $E(X | \Sigma, U) \ge a$. From Corollary 4.15.1, we have $E(X | \Sigma) | U \ge E(X | \Sigma, U)$. Therefore, $E(X | \Sigma) | U \ge a$. Similarly for the second statement. []

IV.5 On a Linearly Ordered σ -Lattice

Let Y be a random variable. The σ -lattice $\Sigma(Y)$ induced by Y is linearly ordered. Let X be a random variable and let $X* = E(X \mid \Sigma(Y))$. Since X* is $\Sigma(Y)$ -measurable, X* if F(Y)measurable. It has been shown that an F(Y)-measurable random variable is a function of Y, i.e., there is an extended real-valued function f defined on the extended real line such that $X* = f \circ Y$ (cf. Ash (1972)).

<u>Theorem 4.16.</u> Let $E(X \mid \Sigma(Y)) = f \circ Y$. Then f is monotone increasing on the range of Y.

<u>Proof.</u> Let ω_1 and ω_2 in Ω such that $Y(\omega_1) < Y(\omega_2)$. Since $X * \in R(\Sigma(Y))$, $[X * \ge X * (\omega_1)] \in \Sigma(Y)$. Upper sets in $\Sigma(Y)$ are either [Y > a] or $[Y \ge a]$ for each a. It follows that $[Y \ge Y(\omega_1)]$ is the smallest element in $\Sigma(Y)$ which contains ω_1 . Since $\omega_1 \in [X * \ge X * (\omega_1)]$, $[X * \ge X * (\omega_1)] \supset [Y \ge Y(\omega_1)]$. The element ω_2 is in $[Y \ge Y(\omega_1)]$, so $\omega_2 \in [X * \ge X * (\omega_1)]$ and hence $X * (\omega_2) \ge X * (\omega_1)$. In other words, $f(Y(\omega_2)) \ge f(Y(\omega_1))$ whenever $Y(\omega_2) > Y(\omega_1)$. This completes the proof. []

Let Σ be a sub- σ -lattice of F. Let a binary relation \leq_{Σ} be defined on the linear space of random variables by $X \leq_{\Sigma} Y$ if

$$\int_{U} XdP \leq \int_{U} YdP \text{ for each } U \in \Sigma.$$

The binary relation \leq_{Σ} is a quasi-ordering. For each X we have $X \leq_{\Sigma} E(X \mid \Sigma)$ provided that $X \in L_1$. If the σ -field generated by Σ is F, then \leq_{Σ} is a partial ordering.

<u>Theorem 4.17.</u> Let Σ be a linearly ordered sub- σ -lattice of F and let Σ_1 and Σ_2 be two sub- σ -lattices of Σ such that $\Sigma_1 \subset \Sigma_2$. For each $X \in L_1$, we have

$$\mathbf{E}(\mathbf{X} \mid \Sigma_1) \leq \sum_{\Sigma} \mathbf{E}(\mathbf{E}(\mathbf{X} \mid \Sigma_2) \mid \Sigma_1) \leq \sum_{\Sigma} \mathbf{E}(\mathbf{X} \mid \Sigma_2) \ .$$

<u>Proof</u>. Let $Y_i = E(X | \Sigma_i)$, i = 1, 2, and let $Y_0 = E(Y_2 | \Sigma_1)$. Suppose it is not true that $Y_1 \leq \Sigma Y_0$. Then there exists a U such that

$$\int_{U} Y_{1} dP > \int_{U} Y_{0} dP$$

Let $\alpha = \inf\{Y_1(\omega) : \omega \in U\}$ and let $A = [X_1 \ge \alpha]$ and $B = [Y_1 > \alpha]$. By the linear ordering property of Σ , we have $A \supset U \supset B$. For each $C \in \Sigma_1 \cap F(Y_1)$,

$$\int_{C} Y_{1} dP = \int_{C} X dP \leq \int_{C} Y_{2} dP \leq \int_{C} Y_{0} dP$$

In particular for C is A or B. Therefore,

$$\int_{U-B} Y_1 dP = \alpha P(U-B) > \int_{U-B} Y_0 dP$$

and hence $P(U \cap [Y_0 < a]) > 0$. By the linear property of Σ again, $[Y_0 \ge a] \subset U$. It follows that

$$\begin{split} \int_{\mathbf{A}} \mathbf{Y}_{0} d\mathbf{P} &= \int_{\mathbf{U}} \mathbf{Y}_{0} d\mathbf{P} + \int_{\mathbf{A} - \mathbf{U}} \mathbf{Y}_{0} d\mathbf{P} \\ &< \int_{\mathbf{U}} \mathbf{Y}_{1} d\mathbf{P} + \alpha \mathbf{P} (\mathbf{A} - \mathbf{U}) \\ &= \int_{\mathbf{A}} \mathbf{Y}_{1} d\mathbf{P}. \end{split}$$

This contradicts that $\int_{A} Y_1 dP \leq \int_{A} Y_0 dP$. Thus $Y_1 \leq \sum Y_0$. By a similar argument, we have $Y_0 \leq \sum Y_2$.

A sequence $\{X_n, \Sigma_n\}$ is said to be a <u>submartingale</u> with respect to \leq_{Σ} if $\{\Sigma_n\}$ is a monotone increasing sequence of σ lattices, X_n is Σ_n -measurable for each n and $X_n \leq_{\Sigma} E(X_{n+k} | \Sigma_n)$ for each non-negative integer k. The following corollary is an immediate result of Theorem 4.17.

<u>Corollary 4.17.1</u>. Let Σ be linearly ordered and let $\{\sum_{n}\}$ be a monotone increasing sequence of sub- σ -lattices of Σ . For each $X \in L_1$, the sequence $\{X_n, \Sigma_n\}$ is a submartingale with respect to \leq_{Σ} where $X_n = E(X \mid \Sigma_n)$ for each n.

By a similar argument in the proof of Theorem 4.17, one may obtain the following result.

<u>Corollary 4.17.2.</u> If Σ is linearly ordered and $X \in L_1$, then $E(X \mid \Sigma)$ is a minimal element in the class $\{Z \in L_1(\Sigma) : X \leq \Sigma Z\}$ with respect to \leq_{Σ} .

<u>Corollary 4.17.3.</u> Under the same assumptions as in Corollary 4.17.1, if $X \leq_{\Sigma} X_n$ for some n, then $X_{n+k} = E(X \mid \Sigma)$ for each positive integer k.

Proof. Apply Corollary 4.17.2 and Theorem 4.17. []

<u>Theorem 4.18.</u> Let Σ be a linearly ordered sub- σ -lattice of F, let Σ_1 be a sub- σ -lattice of Σ and let $X \in L_1(\Sigma)$. Then $E(X | \Sigma_1) = E(X | F_1)$ where F_1 is the σ -field generated by Σ_1 .

<u>Proof</u> Let $X^* = E(X | \Sigma_1)$. Then $X \leq_{\Sigma_1} X^*$. But on the other hand $X = E(X | \Sigma)$ and $\Sigma \supseteq \Sigma_1$. By Theorem 4.17, $X^* \leq_{\Sigma_1} X$. Therefore $E(X - X^*) |_U = 0$ for each $U \in \Sigma_1$ and hence $X^* = E(X | F_1)$. []

<u>Corollary 4.18.1.</u> Let Σ be a linearly ordered sub- σ -lattice of F and let $\{\Sigma_n\}$ be a monotone increasing sequence of sub- σ lattices of Σ . For each $X \in L_1(\Sigma)$, the sequence $\{X_n, \Sigma_n\}$ is a martingale where $X_n = E(X \mid \Sigma_n)$.

<u>Proof</u>. It is obvious that $\{E(X \mid F_n), F_n\}$ is a martingale where F_n is the σ -field generated by Σ_n for each n. By Theorem 4.18, $E(X \mid F_n) = E(X \mid \Sigma_n)$ for each n. It follows that $E(X_{n+k} \mid \Sigma_n) = E(E(X \mid \Sigma_{n+k}) \mid \Sigma_n) = E(E(X \mid F_{n+k}) \mid \Sigma_n) = E(X \mid \Sigma_n) = X_n$. This completes the proof. []

V. MULTIVARIATE ISOTONIC REGRESSION

V.1 Introduction

Let Ω be a finite set and let H be the linear space of vector-valued functions $Y: \Omega \to \mathbb{R}^m$ for a fixed positive integer m. For convenience, let $\Omega = \{1, 2, ..., n\}$ and for each $j \in \Omega$ let Y(j) be the function value of Y at j which is an m-component column vector. The function Y in this case is an m x n matrix (y_{ij}) . For each i, we denote the n-component row vector $(y_{i1}, y_{i2}, ..., y_{in})$ by Y_i . Therefore, Y_i is a function from Ω to R, i = 1, ..., m. For each i, let \leq_i be a quasi-ordering defined on Ω and let M_i be the family of real-valued isotonic functions with respect to the ordering \leq_i . Let M be defined by $Y \in M$ if $Y_i \in M_i$ for each i = 1, ..., m. The minimization problem we are interested in is to minimize

(5.1)
$$f(Z) = \sum_{j=1}^{n} w_{j}(X(j) - Z(j))^{t} V(X(j) - Z(j))$$

subject to $Z \in M$ where X is a given $m \ge n$ matrix, V is a given $m \ge m$ positive definite matrix and w_1, w_2, \ldots, w_n are positive real numbers.

Let us define a bilinear functional (\cdot, \cdot) on H by

$$(\mathbf{Y}, \mathbf{Z}) = \Sigma_{j=1}^{n} \mathbf{w}_{j} \mathbf{Y}(j)^{t} \mathbf{V} \mathbf{Z}(j)$$

for each Y and Z in H. Then the bilinear functional (\cdot, \cdot) is an inner product and the linear space H with (\cdot, \cdot) is a Hilbert space. Since M_i is finitely generated as described in Chapter III for i = 1, ..., m, so is M and hence M is a closed convex cone in H. Existence and uniqueness of the optimal solution to (5.1) follow from Theorem 2.1. Such an optimal solution is called the <u>multivariate isotonic regression</u> of X and is denoted by P(X|M). For convenience, we denote $P(X_i|M_i)$ as the isotonic regression of X_i , i = 1, ..., m, i.e., $P(X_i|M_i)$ minimizes $\sum_{j=1}^{n} (x_{ij} - z_{ij})^2 w_j$ subject to $Z_i \in M_i$.

A necessary and sufficient condition for an $m \times n$ matrix X^* to be the multivariate isotonic regression of X is given by Theorem 2.7. Let J^i be an $m \times n$ matrix such that each entry at the ith row has value one with zeros elsewhere, i = 1, ..., m. Then J^i and $-J^i$ are in M, i = 1, ..., m. If $X^* = P(X|M)$, then by (2.11) we have

$$\sum_{j=1}^{n} w_{j} (X(j) - X^{*}(j))^{t} V J^{i}(j) = 0 \quad i = 1, ..., m.$$

The matrix V is positive definite. Solving the above equations, we shall obtain

(5.2)
$$\Sigma_{j=1}^{n} \mathbf{w}_{j} \mathbf{x}_{ij} = \Sigma_{j=1}^{n} \mathbf{w}_{j} \mathbf{x}_{ij}^{*} \quad i = 1, \dots, m.$$

If V is a diagonal matrix, i.e., $v_{ij} = 0$ if $i \neq j$, then (5.1) can be written as

$$f(Z) = \Sigma_{j=1}^{n} \mathbf{w}_{j} \Sigma_{i=1}^{m} \mathbf{v}_{ii} (\mathbf{x}_{ij} - \mathbf{z}_{ij})^{2}$$
$$= \Sigma_{i=1}^{m} \mathbf{v}_{ii} \{\Sigma_{j=1}^{n} (\mathbf{x}_{ij} - \mathbf{z}_{ij})^{2} \mathbf{w}_{j}\}.$$

Therefore, $P(X | M)_i = P(X_i | M_i)$ for i = 1, ..., m. If V is not a diagonal matrix, the minimization problem could be very complicated.

In this thesis, we shall treat m = 2 and $w_1 = w_2 = \dots = w_n = 1$. A 2 x 2 positive matrix V can be represented by $v_{11} = v_1^2$, $v_{22} = v_2^2$ and $v_{12} = v_{21} = -\rho v_1 v_2$ where v_1 and v_2 are positive and $-1 < \rho < 1$. The objective function is

$$f(Z) = \sum_{j=1}^{n} \sum_{i=1}^{2} \{ v_i^2(x_{ij}^{-}z_{ij}^{-})^2 - \rho v_1 v_2(x_{1j}^{-}z_{1j}^{-})(x_{2j}^{-}z_{2j}^{-}) \} .$$

Let Y be the 2 x n matrix such that $Y_i = v_i X_i$, i = 1, 2 and let Y* be the multivariate isotonic regression of Y with respect to the positive definite matrix T such that $t_{11} = t_{22} = 1$ and $t_{12} = t_{21} = -\rho$. Let X* be defined by $X_i^* = v_i^{-1} Y_i^*$, i = 1, 2. Then X* is the multivariate isotonic regression of X with respect to the positive definite matrix V. Without loss of generality, we may assume that $v_1 = v_2 = 1$.

Let X* be the multivariate isotonic regression of X. If M_2 is the family of constant functions, i.e., the quasi-ordering \leq_2 is such that $j \leq_2 k$ and $k \leq_2 j$ for each j and k between one and n, then by (5.2) we have $X_2^* = \overline{x}_2$ where $\overline{x}_2 = \sum_{j=1}^n x_{2j}/n$. For each Z such that $Z_2 = \overline{x}_2$, we have

$$f(Z) = \sum_{j=1}^{n} \{ [x_{1j} - \rho(x_{2j} - \overline{x}_2)] - z_{1j} \}^2 + (1 - \rho^2) \sum_{j=1}^{n} (x_{2j} - \overline{x}_2)^2 \}$$

It follows that $\mathbf{X}_{1}^{*} = \mathbf{P}(\mathbf{X}_{1} - \rho(\mathbf{X}_{2} - \mathbf{x}_{2}) | \mathbf{M}_{1}).$

Let M_2 be the family of all functions, i.e., the quasiordering \leq_2 is such that neither $j \leq_2 k$ nor $k \leq_2 j$ for each j and k between one and n with $j \neq k$. Let U^j be a $2 \times n$ matrix such that the (2, j) entry has value one with zeros elsewhere. Then U^j and $-U^j$ are in M for $i = 1, \ldots, n$. By (2.11), we have $\rho(x_{1j} - x_{1j}^*) = x_{2j} - x_{2j}^*$, $j = 1, \ldots, n$. For each Z such that $\rho(x_{1j} - z_{1j}) = x_{2j} - z_{2j}$, $j = 1, \ldots, n$, we have

$$f(Z) = (1 - \rho^2) \Sigma_{j=1}^n (x_{1j} - z_{1j})^2$$
.

It follows that $X_1^* = P(X_1 | M_1)$ and $X_2^* = X_2 - \rho(X_1 - X_1^*)$.

V.2 Bivariate Isotonic Regression

Let H be the linear space of $2 \times n$ matrices, let M be the family of matrices such that $Z \in M$ if $z_{ij} \leq z_{ij+1}$ for i = 1, 2and for j = 1, ..., n-1. The problem in the remainder of the chapter is to minimize

(5.3)
$$f(\mathbf{Z}) = \sum_{j=1}^{n} \sum_{i=1}^{2} \{ (\mathbf{x}_{ij} - \mathbf{z}_{ij})^{2} - \rho(\mathbf{x}_{1j} - \mathbf{z}_{1j}) (\mathbf{x}_{2j} - \mathbf{z}_{2j}) \}$$

subject to $Z \in M$ where X is a given $2 \times n$ matrix and ρ is a given real number, $-1 < \rho < 1$. The optimal solution to the problem is called the <u>bivariate isotonic regression</u> of x with respect to ρ and is denoted by $P(X \mid M, \rho)$.

Let $M_{id} := \{Z: Z_1 \in M_1, -Z_2 \in M_2\}$, i.e., $Z \in M_{id}$ if the first row of Z is monotone increasing and the second row of Z is monotone decreasing. Let Y be such that $Y_1 = X_1$ and $Y_2 = -X_2$, let $Y = P(Y|M, -\rho)$ and let X* be defined by $X_1^* = Y_1$ and $X_2^* = -Y_2^*$. Then $X = P(X|M_{id}, \rho)$. Similarly for the situations M_{di} and M_{dd} , where $M_{di} = \{Z: -Z_1 \in M_1, Z_2 \in M_2\}$ and $M_{dd} = \{Z: -Z_1 \in M_1, -Z_2 \in M_2\}$. The sign of ρ will play the most important role in analyzing the properties of $P(X|M, \rho)$ as we shall soon see. Therefore $P(X|M, \rho)$, $P(X|M_{id}, -\rho)$, $P(X|M_{di}, -\rho)$ and $P(X|M_{dd}, \rho)$ will have the same properties. Let 1_{ij} be the $2 \ge n$ matrix such that the jth, \ldots, nth entries at the ith row have values one with zeros elsewhere, $i = 1, 2, j = 1, 2, \ldots, n$. The set $\{1_{ij}: i = 1, 2, j = 1, 2, \ldots, n\}$ is linearly independent, and the family M is the cone generated by the set such that $Z \in M$ if and only if $Z = \sum_{i=1}^{2} \sum_{j=1}^{n} \beta_{ij} 1_{ij}$ with $\beta_{ij} \ge 0$ if j > 1. Let $X = \sum_{i=1}^{2} \sum_{j=1}^{n} \alpha_{ij}^{*} 1_{ij}$ with $\alpha_{ij}^{*} \ge 0$ if j > 1. Theorem 2.12 and Theorem 2.13 show that $X = P(X \mid M, \rho)$ if and only if X = satisfies

(5.4)
$$\Sigma_{h=j}^{n}(x_{1h}^{-\rho}x_{2h}^{-\rho}) \leq \Sigma_{h=j}^{n}(x_{1h}^{*}-\rho x_{2h}^{*}) \quad j = 1, \dots, n$$
$$\Sigma_{h=j}^{n}(x_{2h}^{-\rho}-\rho x_{1h}^{-\rho}) \leq \Sigma_{h=j}^{n}(x_{2h}^{*}-\rho x_{1h}^{*}) \quad j = 1, \dots, n$$

(5.5)
$$\Sigma_{h=j}^{n}(x_{1h}-\rho x_{2h}) = \Sigma_{h=j}^{n}(x_{1h}^{*}-\rho x_{2h}^{*}) \quad \text{if} \quad a_{1j}^{*} > 0$$
$$\Sigma_{h=j}^{n}(x_{2h}-\rho x_{1h}) = \Sigma_{h=j}^{n}(x_{2h}^{*}-\rho x_{1h}^{*}) \quad \text{if} \quad a_{2j}^{*} > 0$$

and (5.2), i.e., $\sum_{h=1}^{n} x_{ij} = \sum_{h=1}^{n} x_{ih}^{*}$ i = 1,2.

<u>Theorem 5.1.</u> If $\rho \ge 0$ and $X_1 - \rho X_2$ is monotone increasing then $X_2^* = P(X_2 | M_2)$ and $X_1^* = X_1 + \rho(X_2^* - X_2)$ where $X^* = P(X | M, \rho)$. Similarly if we interchange indices 1 and 2.

<u>Proof.</u> Let Y be defined by $Y_2 = P(X_2 | M_2)$ and $Y_1 = X_1 + \rho(Y_2 - X_2)$. We are going to show that $Y = P(X | M, \rho)$. By the assumption that $X_1 - \rho X_2$ is monotone increasing and $\rho \ge 0$, we have $Y \in M$. Note that $X_1 - \rho X_2 = Y_1 - \rho Y_2$. Therefore Ysatisfies (5.4) and (5.5). By (4.2), $\sum_{j=1}^n y_{2j} = \sum_{j=1}^n x_{2j}$ and it follows that $\sum_{j=1}^n y_{1j} = \sum_{j=1}^n x_{1j}$. This completes the proof. []

<u>Theorem 5.2.</u> Let $\rho \leq 0$. If $X_1 - \rho X_2$ is monotone increasing and $P(X_2 \mid M_2)$ is constant, then $X_2^* = P(X_2 \mid M_2)$ and $X_1^* = X_1 + \rho(X_2^* - X_2)$ where $X^* = P(X \mid M, \rho)$. Similarly if we interchange indices 1 and 2. If $P(X_1 \mid M_1)$ and $P(X_2 \mid M_2)$ are constant, then $X_i^* = P(X_i \mid M_i)$, i = 1, 2.

<u>Proof.</u> The first statement is similar to Theorem 5.1 except that we need $P(X_2 | M_2)$ be constant to ensure that $X_1 + \rho(X_2^* - X_2)$ is isotonic. Let us consider the second statement. Let $P(X_1 | M_1)$ and $P(X_2 | M_2)$ be constant with value γ_1 and γ_2 respectively. Then $\gamma_i = \sum_{k=1}^n x_{ij}/n$ and $\sum_{k=j}^n x_{ij} \leq (n-j+1)\gamma_i$ for i = 1, 2 and for j = 1, ..., n. Let $Y = \gamma_1 1_{11} + \gamma_2 1_{21}$. By the assumption $\rho \leq 0$, Y satisfies (5.2), (5.4) and (5.5). Therefore $Y = P(X | M, \rho)$. []

The average property which plays the most important role in the isotonic regression and its algorithms, is the one we are interested in. Let X^* be the bivariate isotonic regression and let $x_{ij}^* < x_{ij+1}^*$ and $x_{ik}^* < x_{ik+1}^*$ for some j and k with j < k, i = 1, 2. By (5.5) and (5.2), we have

$$\begin{split} \Sigma_{h=1}^{j} (\mathbf{x}_{1h}^{-} - \rho \mathbf{x}_{2h}^{-}) &= \Sigma_{h=1}^{j} (\mathbf{x}_{1h}^{*} - \rho \mathbf{x}_{2h}^{*}) \\ \Sigma_{h=1}^{j} (\mathbf{x}_{2h}^{-} - \rho \mathbf{x}_{1h}^{-}) &= \Sigma_{h=1}^{j} (\mathbf{x}_{2h}^{*} - \rho \mathbf{x}_{1h}^{*}) \\ \Sigma_{h=j+1}^{k} (\mathbf{x}_{1h}^{-} - \rho \mathbf{x}_{2h}^{-}) &= \Sigma_{h=j+1}^{k} (\mathbf{x}_{1h}^{*} - \rho \mathbf{x}_{2h}^{*}) \\ \Sigma_{h=j+1}^{k} (\mathbf{x}_{2h}^{-} - \rho \mathbf{x}_{1h}^{-}) &= \Sigma_{h=j+1}^{k} (\mathbf{x}_{2h}^{*} - \rho \mathbf{x}_{1h}^{*}) \\ \Sigma_{h=k+1}^{n} (\mathbf{x}_{1h}^{-} - \rho \mathbf{x}_{2h}^{-}) &= \Sigma_{h=k+1}^{n} (\mathbf{x}_{1h}^{*} - \rho \mathbf{x}_{2h}^{*}) \end{split}$$

and

$$\Sigma_{h=k+1}^{n}(x_{2h}^{}-\rho x_{1h}^{}) = \Sigma_{h=k+1}^{n}(x_{2h}^{*}-\rho x_{1h}^{*})$$

The first two equations show that

$$\Sigma_{h=1}^{j} x_{ih} = \Sigma_{h=1}^{j} x_{ih}^{*}, \quad i = 1, 2,$$

the next two equations show that

$$\Sigma_{h=j+1}^{k} x_{ih} = \Sigma_{h=j+1}^{k} x_{ih}^{*}, \quad i = 1, 2,$$

and the last two equations show that

$$\Sigma_{h=k+1}^{n} x_{ih} = \Sigma_{h=k+1}^{n} x_{ih}^{*}, \quad i = 1, 2.$$

If $x_{1j}^* < x_{1j+1}^*$ and $x_{1k}^* < x_{1k+1}^*$ for a pair of j and k with j < k, then by (5.5) and (5.2) we have

$$\Sigma_{h=1}^{j} (\mathbf{x}_{1h} - \rho \mathbf{x}_{2h}) = \Sigma_{h=1}^{j} (\mathbf{x}_{1h}^{*} - \rho \mathbf{x}_{2h}^{*})$$

$$\Sigma_{h=j+1}^{k} (\mathbf{x}_{1h} - \rho \mathbf{x}_{2h}) = \Sigma_{h=j+1}^{k} (\mathbf{x}_{1h}^{*} - \rho \mathbf{x}_{2h}^{*})$$

and

$$\Sigma_{h=k+1}^{n}(x_{1h}^{-}\rho x_{2h}^{-}) = \Sigma_{h=k+1}^{n}(x_{1h}^{*}-\rho x_{2h}^{*}).$$

Similarly if we interchange indices 1 and 2.

<u>Theorem 5.3.</u> If $\rho \ge 0$ and $VX \le VY$, then $P(X \mid M, \rho) \le P(Y \mid M, \rho).$

<u>Proof.</u> Let U^{1k} be the 2 x n matrix such that the (1,k) entry has value one, the (2,k) entry has value ρ with zeros elsewhere, $k = 1, \ldots, n$. Similarly for the $U^{21}, U^{22}, \ldots, U^{2n}$. We shall show $P(X+\delta U^{ij}|M,\rho) \ge P(X|M,\rho)$ for any $\delta > 0$ and for each i and j. Since Y - X can be represented as a non-negative linear combination of U^{ij} , $i = 1, 2; j = 1, \ldots, n$, it will then follow $P(X|M,\rho) \le P(Y|M,\rho)$.

Let $X * = P(X | M, \rho)$ and let $Y * = P(X+U | M, \rho)$ where $U = \delta U^{lk}$ for a fixed positive real number δ . Then by (2.3)

$$(X+U-Y*, Y*-X*) > 0$$

and

$$(X - X^*, Y^* - X^*) \leq 0.$$

The difference of the above two inequalities is

$$(U - Y * + X *, Y * - X *) > 0$$

and hence

$$(U, Y*-X*) \ge (Y*-X*, Y*-X*)$$

The left-hand side of the above inequality is $\delta(1-\rho^2)(y_{1k}^* - x_{1k}^*)$ and hence $y_{1k}^* \ge x_{1k}^*$.

It is trivial that

$$\| \mathbf{X} + \mathbf{U} - \mathbf{Z} \|^2 = \| \mathbf{X} - \mathbf{Z} \|^2 + g(\mathbf{Z})$$

where $g(Z) = 2\delta(1-\rho^2)(x_{1k}-z_{1k}+\delta/2)$. Since $y_{1k}^* \ge x_{1k}^*$, $g(X*\wedge Y*) = g(X*)$ where the (i,j) entry of $X*\wedge Y*$ has the value $x_{ij}^* \wedge y_{ij}^*$. By a similar argument as in (4.7), we have

$$(\mathbf{X}+\mathbf{U}-\mathbf{Y}^*,\,\mathbf{Y}^*-\mathbf{X}^*) \geq (\mathbf{X}+\mathbf{U}-\mathbf{Y}^*,\,\mathbf{Y}^*-\mathbf{X}^*\wedge\,\mathbf{Y}^*)$$

Since

$$\| \mathbf{Y}^{*} - \mathbf{X}^{*} \|^{2} = \Sigma_{j=1}^{n} (\mathbf{Y}^{*}(j) - \mathbf{X}^{*}(j))^{t} \mathbf{V} (\mathbf{Y}^{*}(j) - \mathbf{X}^{*}(j))$$
$$= \Sigma_{j=1}^{n} \{ (\mathbf{y}_{1j}^{*} - \mathbf{x}_{1j}^{*})^{2} + (\mathbf{y}_{2j}^{*} - \mathbf{x}_{2j}^{*})^{2}$$
$$- 2\rho (\mathbf{y}_{1j}^{*} - \mathbf{x}_{ij}^{*}) (\mathbf{y}_{2j}^{*} - \mathbf{x}_{2j}^{*}) \}$$

and

$$(Y*(j)-X*(j))^{t}V(Y*(j)-X*(j)) \ge (Y*(j)-X*(j)\wedge Y*(j))^{t}V(Y*(j)-X*(j)\wedge Y*(j)),$$

j = 1,...,n provided that
$$\rho \ge 0$$
, we have
 $\|Y^*-X^*\|^2 \ge \|Y^*-X^* \wedge Y^*\|^2$. Therefore,

$$\|\mathbf{X}+\mathbf{U}-\mathbf{X}*\|^{2} = \|\mathbf{X}+\mathbf{U}-\mathbf{Y}*\|^{2} + \|\mathbf{Y}*-\mathbf{X}*\|^{2} + 2(\mathbf{X}+\mathbf{U}-\mathbf{Y}*,\mathbf{Y}*-\mathbf{X}*)$$

$$\geq \|\mathbf{X}+\mathbf{U}-\mathbf{Y}*\|^{2} + \|\mathbf{Y}*-\mathbf{X}*\wedge\mathbf{Y}*\|^{2}$$

$$+ 2(\mathbf{X}+\mathbf{U}-\mathbf{Y}*,\mathbf{Y}*-\mathbf{X}*\wedge\mathbf{Y}*)$$

$$= \|\mathbf{X}+\mathbf{U}-\mathbf{X}*\wedge\mathbf{Y}*\|^{2}$$

and it follows

$$\| \mathbf{X} - \mathbf{X}^* \|^2 = \| \mathbf{X} + \mathbf{U} - \mathbf{X}^* \|^2 - g(\mathbf{X}^*)$$

$$\geq \| \mathbf{X} + \mathbf{U} - \mathbf{X}^* \wedge \mathbf{Y}^* \|^2 - g(\mathbf{X}^* \wedge \mathbf{Y}^*)$$

$$= \| \mathbf{X} - \mathbf{X}^* \wedge \mathbf{Y}^* \|^2.$$

By the fact that $||X-X^*|| \le ||X-Z||$ for each $Z \in M$ and $Z = X^*$ if the equality holds, we have $X^* \land Y^* = X^*$ and hence $X^* \le Y^*$. This completes the proof. []

<u>Corollary 5.3.1.</u> If $\rho \ge 0$ and $a \le VX \le b$, then $V^{-1}a \le P(X \mid M, \rho) \le V^{-1}b$ where a and b are 2×1 column vectors.

An upper bound and a lower bound of $P(X | M, \rho)$ for $\rho \ge 0$ can be determined by the above corollary. Let Y and Z be $2 \ge k$ and $2 \ge (n-k)$ matrices defined by Y(j) = X(j), $j = 1, \ldots, k$ and Z(j) = X(k+j), $j = 1, \ldots, n-k$ respectively. Let $M_k := \{U \in R^{2 \ge k} : u_{ij} \le u_{ij+1}, i = 1, 2, j = 1, \ldots, k-1\}$ for each positive integer k. If there exists a 2×1 vector c such that $P(Y | M_k, \rho) \leq c$ and $P(X | M_{n-k}, \rho) \geq c$, then by (5.2), (5.4) and (5.5), $P(X | M, \rho)$ can be obtained by considering the first k components and the last n-k components independently for each ρ between -1 and 1. From (2.2) $P(X | -M, \rho) = -P(-X | M, \rho)$, so if $\rho \geq 0$ and $a \leq VX \leq b$ then $V^{-1}a \leq P(X | -M, \rho) \leq V^{-1}b$. But it is not true in general that $X \geq 0$ implies $P(X | M, \rho) \geq 0$ for $\rho \geq 0$.

<u>Theorem 5.4.</u> If $\rho \leq 0$ and $a \leq X \leq b$, then $a \leq P(X \mid M, \rho) \leq b$ where a and b are 2×1 vectors.

<u>Proof.</u> We shall show that if $X \ge 0$ then $P(X | M, \rho) \ge 0$ and $P(X | -M, \rho) \ge 0$. It will then follow that $X \ge c$ implies $P(X-c | M, \rho) \ge 0$ and $P(X-c | -M, \rho) \ge 0$ and hence $P(X | M, \rho) \ge c$ and $P(X | -M, \rho) \ge c$. By the identity $P(X | M, \rho) = -P(-X | -M, \rho)$, if $X \le b$, then $-X \ge -b$ and hence $-P(X | M, \rho) = P(-X | -M, \rho) \ge -b$. Therefore, if $a \le X \le b$ then $a \le P(X | M, \rho) \le b$ provided that $\rho \le 0$.

Let a and β be non-negative real numbers and let x and y be real numbers. If $x \leq 0$, then

$$\begin{aligned} (a_{-x})^{2} + (\beta_{-y})^{2} - 2\rho(a_{-x})(\beta_{-y}) &= (1-\rho^{2})(a_{-x})^{2} + (\beta_{-y-a\rho+\rho_{x}})^{2} \\ &\geq (1-\rho^{2})a^{2} + (\beta_{-y-a\rho+\rho_{x}})^{2} \\ &= (1-\rho^{2})a^{2} + (\beta_{-y+\rho_{x}})^{2} + \rho^{2}a^{2} \\ &- 2\rho a(\beta_{-y+\rho_{x}}) \\ &= a^{2} + (\beta_{-y+\rho_{x}})^{2} - 2\rho a(\beta_{-y+\rho_{x}}). \end{aligned}$$

If also $y - \rho x < 0$, then

$$a^{2} + (\beta - y + \rho \mathbf{x})^{2} - 2\rho a(\beta - y + \rho \mathbf{x}) \ge a^{2} + \beta^{2} - 2\rho a\beta$$

Therefore

$$(a-x, \beta-y)V(a-x, \beta-y)^{t} \ge (a, \beta-(y-\rho_{x})\sqrt{0})V(a, \beta-(y-\rho_{x})\sqrt{0})^{t}$$
.

Similarly, if $y \leq 0$ then

$$(a-x,\beta-y)V(a-x,\beta-y)^{t} \ge (a-(x-\rho y)_{V}0,\beta)V(a-(x-\rho Y)_{V}0,\beta)^{t}$$
.

Let $Z \in M$ and let k be the largest index between 1 and n such that Z(k) is not non-negative. Let Y be defined by Y(j) = Z(j) for each j > k and

$$y_{1j} = \begin{cases} (z_{1j} - \rho z_{2j}) \vee 0 & \text{if } z_{1k} \ge 0 \\ 0 & \text{if } z_{1k} < 0 \end{cases}$$
$$y_{2j} = \begin{cases} 0 & \text{if } z_{1k} \ge 0 \\ (z_{2j} - \rho z_{1j}) \vee 0 & \text{if } z_{1k} < 0 \end{cases}$$

for $j = 1, \ldots, k$. It is trivial that $Y \ge 0$ and $Y \in M$.

If $z_{1k} < 0$, then $z_{1j} < 0$ for j = 1, ..., k. By the inequality described above, if $X \ge 0$ then

$$(\mathbf{X}(j)-\mathbf{Z}(j))^{\mathsf{t}}\mathbf{V}(\mathbf{X}(j)-\mathbf{Z}(j)) \geq (\mathbf{Z}(j)-\mathbf{Y}(j))^{\mathsf{t}}\mathbf{V}(\mathbf{X}(j)-\mathbf{Y}(j))$$

j = 1, ..., n. Therefore $||X-Z|| \ge ||X-Y||$. If $z_{1k} \ge 0$, then $z_{2k} < 0$ and similarly $||X-Z|| \ge ||Z-Y||$. It follows that $P(X|M,\rho) \ge 0$. By a symmetric argument in the sense of reversal we have $P(X|-M,\rho) \ge 0$. []

<u>Theorem 5.5.</u> If $\rho \ge 0$, $x_{1k-1} \ge x_{1k}$ and $x_{1k-1} - x_{1k} \ge \rho(x_{2k-1} - x_{2k})$ then $x_{1k}^* = x_{1k-1}^*$ where $X^* = P(X \mid M, \rho)$. Similarly if we interchange indices 1 and 2.

<u>Proof</u>. Suppose it were not true, i.e., $x_{lk}^* > x_{lk-l}^*$. From (5.4) and (5.5), we have

$$\Sigma_{h=k-1}^{n} (x_{1j} - \rho x_{2j}) \leq \Sigma_{h=k-1}^{n} (x_{1j}^{*} - \rho x_{2j}^{*})$$

$$\Sigma_{h=k}^{n} (x_{1j} - \rho x_{2j}) = \Sigma_{h=k}^{n} (x_{1j}^{*} - \rho x_{2j}^{*})$$

and

$$\Sigma_{h=k+1}^{n}(x_{1j}^{-\rho x}-\rho x_{2j}^{-\rho x}) \leq \Sigma_{h=k+1}^{n}(x_{1j}^{*}-\rho x_{2j}^{*})$$
 if $k+1 \leq n$.

Therefore $x_{1k-1} - \rho x_{2k-1} \le x_{1k-1}^* - \rho x_{2k-1}^*$ and $x_{1k} - \rho x_{2k} \ge x_{1k}^* - \rho x_{2k}^*$.

Consider the case $x_{2k}^* = x_{2k-1}^*$. We have

$$x_{1k} - \rho x_{2k} \ge x_{1k}^* - \rho x_{2k}^*$$

> $x_{1k-1}^* - \rho x_{2k-1}^*$
 $\ge x_{1k-1} - \rho x_{2k-1}^*$

and hence $\rho(\mathbf{x}_{2k-1}-\mathbf{x}_{2k}) > \mathbf{x}_{1k-1} - \mathbf{x}_{1k}$. This contradicts the assumption that $\mathbf{x}_{1k-1} - \mathbf{x}_{1k} \ge \rho(\mathbf{x}_{2k-1}-\mathbf{x}_{2k})$.

Consider the alternative $x_{2k}^* > x_{2k-1}^*$. Then we have $x_{2k-1} - \rho x_{1k-1} \le x_{2k-1}^* - \rho x_{1k-1}^*$ and $x_{2k} - \rho x_{1k} \ge x_{2k}^* - \rho x_{1k}^*$. Since $x_{1k-1} - \rho x_{2k-1} \le x_{1k-1}^* - \rho x_{2k-1}^*$ and $x_{1k} - \rho x_{2k} \ge x_{1k}^* - \rho x_{2k}^*$, it follows that

$$VX(k-1) \leq VX^{*}(k-1)$$
 and $VX^{*}(k) \leq VX(k)$

and hence $x_{lk} \ge x_{lk}^* > x_{lk-1}^* \ge x_{lk-1}$. This contradicts the assumption that $x_{lk-1} \ge x_{lk}$.

If $\rho \ge 0$, the index (1,k) or (2,k) with $k \ge 1$ satisfying both inequalities described in the above theorem is said to be <u>pivotal</u>. An algorithm based on the pivotal indices will be given in the next section.

 $\frac{\text{Theorem 5.6.}}{\text{If } \rho \leq 0 \text{ and } x_{1k-1} - x_{1k} \geq \rho(x_{2k-1} - x_{2k})$ then $x_{1k}^* = x_{1k-1}^*$ where $X^* = P(X | M, \rho)$. Similarly if we interchange indices 1 and 2.

<u>Proof</u>. By the same argument as in the proof of Theorem 5.5, if $x_{1k}^* > x_{1k-1}^*$, then we have $x_{1k-1} - \rho x_{2k-1} \le x_{1k-1}^* - \rho x_{2k-1}^*$ and $x_{1k} - \rho x_{2k} \ge x_{1k}^* - \rho x_{2k}^*$. Therefore,

$$x_{1k} - \rho x_{2k} \ge x_{1k}^* - \rho x_{2k}^*$$

> $x_{1k-1}^* - \rho x_{2k-1}^*$
 $\ge x_{1k-1} - \rho x_{2k-1}^*$

and hence $\rho(x_{2k-1}-x_{2k}) > x_{1k-1} - x_{1k}$. This contradicts the assumption.

If $\rho \leq 0$, the index (1,k) or (2,k) with k > 1 satisfying the inequality described in the above theorem is said to be <u>pivotal</u>. Let $\rho \geq 0$. If $\mathbf{x}_{1k-1} \geq \mathbf{x}_{1k}$ or $\mathbf{x}_{2k-1} \geq \mathbf{x}_{2k}$, then at least one of (1,k) and (2,k) is pivotal. But if $\mathbf{x}_{1k-1} \geq \mathbf{x}_{1k}$ and $\mathbf{x}_{2k-1} \geq \mathbf{x}_{2k}$, it is not necessary that both (1,k) and (2,k) are pivotal. Therefore, whenever there is a violation, i.e., $\mathbf{x}_{1k-1} \geq \mathbf{x}_{1k}$, there is a pivotal element. Let $\rho \leq 0$. If $\mathbf{x}_{1k-1} \geq \mathbf{x}_{1k}$ or $\mathbf{x}_{2k-1} \geq \mathbf{x}_{2k}$, it is not necessary that (1,k) or (2,k) is pivotal. But is $\mathbf{x}_{1k-1} \geq \mathbf{x}_{1k}$ and $\mathbf{x}_{2k-1} \geq \mathbf{x}_{2k}$, then both (1,k) and (2,k) are pivotal. In some cases, we may have violations but there is no pivotal element.

V.3 Simplified Projection

The bivariate isotonic regression of X is the generalized projection of X to the closed convex cone M. By Theorem 2.13, $P(X|M,\rho)$ can be obtained as the projection to a linear space which is generated by a subset of $\{1_{ij}: i = 1, 2; j = 1, ..., n\}$. Therefore, how to obtain the projection to such a linear space in our present structure is our primary work.

Let Γ be a subindex set of $\{(i, j): i = 1, 2; j = 1, ..., n\}$, let $\Gamma_i = \{j: (i, j) \in \Gamma\}$, i = 1, 2 and let S_{Γ} be the linear space generated by $\{1_{ij}: (i, j) \in \Gamma\}$, i.e.,

$$S_{\Gamma} = \{Y: Y = \Sigma_{j \in \Gamma_{1}} \beta_{1j} \beta_{1j} + \Sigma_{j \in \Gamma_{2}} \beta_{2j} \beta_{2j} \beta_{j}, \beta_{ij} \text{ real}\}.$$

For convenience, in the following three paragraphs we denote $X = (X_1, X_2)$ as a 2n-component row vector and similarly for Y, Z and l_{ij} 's. Let T be a 2n x 2n positive definite matrix which is the tensor product of V and I_n , i.e., $t_{ii} = 1$, $t_{in+i} = -\rho$, $t_{ij} = t_{in+j} = 0$ for i, j = 1, ..., n with $j \neq i$ and $t_{n+i, i} = -\rho$, $t_{n+i, n+i} = 1$, $t_{n+i, n+j} = t_{n+i, j} = 0$ for i, j = 1, ..., n with $j \neq i$. Let Γ_1 have p elements, Γ_2 have q elements and let A be the $(p+q) \ge 2n$ matrix which is composed of l_{ij} , $(i, j) \in \Gamma$, such that the first p rows are $l_{1j_1}, l_{1j_2}, ..., l_{1j_p}$ with $j_1 < j_2 < \cdots < j_p$ and the last q rows are $l_{2k_1}, l_{2k_2}, \cdots, l_{2k_q}$ with $k_1 < k_2 < \cdots < k_q$. Every row vector Y in S_{Γ} is of the form Y = UA where U is a p+q-component row vector.

Let us define an inner product by

$$(\mathbf{X}, \mathbf{Z}) = \mathbf{XT}\mathbf{Z}^{\mathsf{t}}$$

for each pair of row vectors X and Z. The quantity $||X-Z||^2 = (X-Z, X-Z)$ is the same as f(Z) in (5.3). The projection of X to the linear space S_{Γ} is the product of X and the 2n x 2n matrix P where $P = TA^t (ATA^t)^{-1}A$. To verify the above result, we shall show (X-XP, Y) = 0 for each $Y \in S_{\Gamma}$. Let U be the p+q row vector such that UA = Y. It is trivial that the matrix A has rank p+q and ATA^t is invertible. Therefore

$$(X - XP, Y) = (X - XP)TY^{t}$$

= $XTA^{t}U^{t} - XTA^{t}(ATA^{t})^{-1}ATA^{t}U^{t}$
= $XTA^{t}U^{t} - XTA^{t}U^{t}$
= 0.

The $(p+q) \ge 2n$ matrix A can be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2 \end{bmatrix}$$

where C_1 is a pxn matrix with rank p and C_2 is a qxn matrix with rank q. The $(p+q) \times (p+q)$ matrix ATA^t is

$$\mathbf{ATA}^{t} = \begin{bmatrix} C_{1}C_{1}^{t} & -\rho C_{1}C_{2}^{t} \\ -\rho C_{2}C_{1}^{t} & C_{2}C_{2}^{t} \end{bmatrix}.$$

Let U_1 and U_2 be p-component and q-component row vectors respectively suth that

$$[U_1 \ U_2] = XTA^t(ATA^t)^{-1}.$$

Then $[U_1C_1 \ U_2C_2] = XP$. Since $[U_1 \ U_2]ATA^t = XTA^t$, we have

$$\begin{bmatrix} U_{1} & U_{2} \end{bmatrix} \begin{bmatrix} I_{p} & -\rho C_{1} C_{2}^{t} \\ -\rho C_{2} C_{1}^{t} (C_{1} C_{1}^{t})^{-1} & C_{2} C_{2}^{t} \end{bmatrix} = XTA^{t} \begin{bmatrix} (C_{1} C_{1}^{t})^{-1} & 0 \\ 0 & I_{q} \end{bmatrix}$$

and

$$\begin{bmatrix} U_{1} & U_{2} \end{bmatrix} \begin{bmatrix} I_{p} & 0 \\ -\rho C_{2} C_{1}^{t} (C_{1} C_{1}^{t})^{-1} & C_{2} C_{2}^{t} -\rho^{2} C_{2} C_{1}^{t} (C_{1} C_{1}^{t})^{-1} C_{1} C_{2}^{t} \end{bmatrix}$$

= $XT A^{t} \begin{bmatrix} (C_{1} C_{1}^{t})^{-1} & 0 \\ 0 & I_{q} \end{bmatrix} \begin{bmatrix} I_{p} & \rho C_{1} C_{2}^{t} \\ 0 & I_{q} \end{bmatrix}$
= $[(X_{1} - \rho X_{2})C_{1}^{t} & (X_{2} - \rho X_{1})C_{2}^{t}] \begin{bmatrix} (C_{1} C_{1}^{t})^{-1} & \rho (C_{1} C_{1}^{t})^{-1} C_{1} C_{2}^{t} \\ 0 & I_{q} \end{bmatrix}$

Let $B_1 = C_1^t (C_1 C_1^t)^{-1} C_1$. Then
(5.6)
$$U_{2}(C_{2}C_{2}^{t}-\rho^{2}C_{2}B_{1}C_{2}^{t}) = (X_{2}-\rho X_{1})C_{2}^{t}+\rho(X_{1}-\rho X_{2})B_{1}C_{2}^{t}$$

and

$$U_{1} - \rho U_{2}C_{2}C_{1}^{t}(C_{1}C_{1}^{t})^{-1} = (X_{1} - \rho X_{2})C_{1}^{t}(C_{1}C_{1}^{t})^{-1}$$

It follows that

(5.7)
$$(XP)_2 = U_2C_2$$

and

(5.8)
$$(XP)_1 = (X_1 - \rho X_2 + \rho (XP)_2)B_1$$

Similarly if we interchange indices 1 and 2.

Let
$$\Gamma_1 = \{j_1, j_2, \dots, j_p\}$$
 and let $Y = (y_1, y_2, \dots, y_n)$ be a
row vector. Then $YC_1^t = (a_1, \dots, a_p)$ such that $a_k = \sum_{h=j_k}^n y_h$,
 $k = 1, \dots, p$, and $YB_1 = (\beta_1, \dots, \beta_n)$ such that
 $\beta_j = \sum_{h=j_k}^{j_{k+1}-1} y_h / (j_{k+1}-j_k)$, $k = 0, 1, \dots, p+1$ where $j_k \leq j < j_{k+1}$,
 $j_0 = 1$ and $j_{p+1} = n+1$. Similarly if we interchange indices 1
and 2.

Suppose $k \in \Gamma_1 \cap \Gamma_2$ and $k \ge 1$. Let us define

$$S_{1} = \{Y: Y = \Sigma_{i=1}^{2} \Sigma_{j \in \Gamma_{i}} \beta_{ij} (1_{ij} - 1_{ik}), \beta_{ij} \text{ real} \}$$
$$j \leq k$$

and

$$S_{2} = \{Y: Y = \Sigma_{i=1}^{2} \Sigma_{j \in \Gamma_{i}} \beta_{ij} \beta_{ij}, \beta_{ij} \text{ real} \}.$$
$$j \ge k$$

Then S_1 and S_2 are orthogonal and $S_1 + S_2 = S_{\Gamma}$. It follows that

(5.9)
$$P(X|S_{\Gamma}, \rho) = P(X|S_{1}, \rho) + P(X|S_{2}, \rho).$$

In other words, $P(X|S_{\Gamma}, \rho)$ can be obtained by considering the first k-1 components and the last n-k+1 components independently.

Suppose $\Gamma_2 = \{1\}$ and $1 \in \Gamma_1$. Then C_2 is a row vector with each entry having value one, $B_1 C_2^t = C_2^t$, $C_2 C_2^t = n$ and U_2 is a real number. Thus (5.6) is $an = \sum_{j=1}^n x_{2j}$ and $X_2^* = aC_2$, i.e., $x_{21}^* = x_{22}^* = \dots = x_{2n}^* = \sum_{j=1}^n x_{2j}/n$.

Let $X = \sum_{i=1}^{2} \sum_{j=1}^{n} a_{ij} l_{ij}^{1}$, $X^* = \sum_{i=1}^{2} \sum_{j=1}^{n} a_{ij}^* l_{ij}^{1}$ and let $\Lambda = \{(i, j): a_{ij}^* > 0\} \cup \{(1, 1), (2, 1)\}$ where $X^* = P(X | M, \rho)$. Theorem 5.5 shows that if $\rho \ge 0$, $a_{1k} \le 0$ and $a_{1k} \le \rho a_{2k}^{2}$ for some k > 1, then $a_{1k}^* = 0$, $(1, k) \notin \Lambda$ and (1, k) is pivotal. Similarly if we interchange indices 1 and 2. Whenever $X \notin M$, there is at least one pivotal element. Theorem 5.6 shows that if $\rho < 0$ and $a_{1k} \le \rho a_{2k}^{2}$ for some k > 1, then $a_{1k}^* = 0$, $(1, k) \notin \Lambda$ and (1, k)is pivotal. Similarly if we interchange indices 1 and 2. If $X \notin M$, we may or may not have pivotal elements.

Let Λ^0 be the collection of all non-pivotal elements. Then $\Lambda^0 \supset \Lambda$. Since $P(X|M,\rho) = P(X|S_{\Lambda},\rho)$ as indicated in Theorem 2.13 and (5.2), by the smoothing property we have
$$\begin{split} & P(X \mid M, \rho) = P(P(X \mid S_{\Lambda^0}, \rho) \mid S_{\Lambda}, \rho). \quad \text{Let} \quad X^1 = P(X \mid S_{\Lambda^0}, \rho). \quad \text{If} \\ & X^1 \in M, \quad \text{then by Theorem 2.8 and Theorem 2.11 we have} \\ & X^1 = P(X \mid M, \rho). \quad \text{Otherwise, write} \quad X^1 = \sum_{i=1}^2 \sum_{j=1}^n \alpha_{ij}^1 1_{ij} \quad \text{and let} \quad \Lambda^1 \\ & \text{be the set of non-pivotal elements in} \quad \Lambda \quad \text{with respect to} \quad X^1. \quad \text{Since} \\ & P(X^1 \mid M_{\Lambda}, \rho) = X^*, \quad \Lambda^0 \supset \Lambda^1 \supset \Lambda. \quad \text{If} \quad \rho \ge 0, \quad \text{then} \quad \Lambda^0 \neq \Lambda^1. \\ & \text{Applying the above procedure inductively, we shall terminate at a} \\ & \text{positive integer} \quad k \quad \text{such that} \quad X^k \in M \quad \text{and} \quad X^k = X^*. \quad \text{The projection} \quad P(X^h \mid S_{\Lambda^h}, \rho) \quad \text{for} \quad h = 0, 1, \dots, k-1 \quad \text{can be obtained by (5.6),} \\ & (5.7), \quad (5.8) \text{ and } (5.9). \quad \text{Such an algorithm for} \quad \rho \ge 0 \quad \text{is called the} \\ & \text{Simplified Projection (cf. Appendix V).} \end{split}$$

The monotone decreasing sequence $\{\Lambda^h\}$ can be replaced by a monotone decreasing sequence $\{\Gamma^h\}$ such that $\Gamma^h \supseteq \Lambda^h$. The purpose for the presence of Γ^h is that we may use (5.9) more efficiently and hence the order of the simultaneous linear equations (5.6) can be reduced significantly. Such a device may be helpful when we use desk calculators.

If $\rho < 0$, we may possibly obtain the bivariate isotonic regression of X by the Simplified Projection algorithm. However, if there exists an k such that $X^k \notin M$ and $\Lambda^{k-1} = \Lambda^k$, then the Simplified Projection algorithm will fail to yield the bivariate isotonic regression of X. At this stage, we may choose any one of the $2 \times n$ matrices X, X^1, X^2, \dots, X^k as our given data. For each h

between one and k, the problem of obtaining $P(\mathbf{X}^{h} | M_{\Lambda h}, \rho)$ is the same problem of obtaining $P(\mathbf{X} | M, \rho)$ for $P(\mathbf{X}^{h} |_{\Lambda h}, \rho) = P(\mathbf{X} | M, \rho).$

<u>Example 5.1.</u> The data X given below is a portion of that from Bhattacharyya and Kotz (1966).

$$\mathbf{X} = \begin{bmatrix} 25 & 13 & 2 & 15 & 14 & 21 \\ \\ 57 & 36 & 77 & 89 & 76 & 62 \end{bmatrix}$$

Let $\rho = -0.1$. By Theorem 5.6, the set of non-pivotal indices Λ^0 is $\{(1,1), (1,4), (1,6), (2,1), (2,3), (2,4)\}$. Since $4 \in \Lambda_1^0 \cap \Lambda_2^0$ where $\Lambda_1^0 = \{j: (i,j) \in \Lambda^0\}$, we may consider the first three elements and the last three elements independently as shown in (5.9). Let $\Gamma = \{(i,j): (i,j) \in \Lambda^0, j \le 3\}$. Then $\Gamma_1 = \{1\}$ and $1 \in \Gamma_2$. So $x_{11}^1 = x_{12}^1 = x_{13}^1 = (25+13+2)/3 = 13.333$. $X_2 - \rho X_1 + \rho (XP)_2 = (58.167, 35.967, 75.867)$ for the first three elements. By (5.8) we take the average for the first two elements of $X_2 - \rho X_1 + \rho (XP)_2$ and the first three elements of X^1 are

Similarly, we have the last three elements of X^{l} . Therefore

$$\mathbf{x}^{1} = \begin{bmatrix} 13.333 & 13.333 & 13.333 & 15.183 & 15.183 & 19.633 \\ 47.067 & 47.067 & 75.867 & 75.667 & 75.667 & 75.667 \end{bmatrix}$$

At this stage, Λ^1 is $\{(1,1), (1,4), (1,6), (2,1), (2,3)\}$ and the equation (5.6) is

$$\begin{bmatrix} 5.94 & 3.96 \\ 3.96 & 3.9667 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 393.03 \\ 300.03 \end{bmatrix}$$

The solution to the above linear equations is $\beta_1 = 47.066$ and $\beta_2 = 28.651$. By (5.7) and (5.8), we have

$$\mathbf{x}^{2} = \begin{bmatrix} 13.338 & 13.338 & 13.338 & 15.178 & 15.178 & 19.628 \\ 47.066 & 47.066 & 75.717 & 75.717 & 75.717 & 75.717 \end{bmatrix}$$

where $X^2 = P(X^1 | S_{\Lambda^1}, \rho)$. Since $X^2 \in M$, X^2 is $P(X | M, \rho)$ because

$$\begin{split} \mathbf{P}(\mathbf{X} \mid \mathbf{M}, \boldsymbol{\rho}) &= \mathbf{P}(\mathbf{X} \mid \mathbf{M}_{\Lambda^1}, \boldsymbol{\rho}) = \mathbf{P}(\mathbf{P}(\mathbf{P}(\mathbf{X} \mid \mathbf{S}_{\Lambda^0}, \boldsymbol{\rho}) \mid \mathbf{S}_{\Lambda^1}, \boldsymbol{\rho}) \mid \mathbf{M}_{\Lambda^1}, \boldsymbol{\rho}) \\ &= \mathbf{P}(\mathbf{P}(\mathbf{X}^1 \mid \mathbf{S}_{\Lambda^1}, \boldsymbol{\rho}) \mid \mathbf{M}_{\Lambda^1}, \boldsymbol{\rho}) = \mathbf{P}(\mathbf{X}^2 \mid \mathbf{M}_{\Lambda^1}, \boldsymbol{\rho}) = \mathbf{X}^2. \end{split}$$

The projection $P(X^1 | M_{\Lambda^1}, \rho)$ is X^2 , but generally $P(X^1 | M, \rho)$ is not X^2 . Consider X^1 . We have

$$x^{1}(1) = x^{1}(2) = \begin{bmatrix} 13.333 \\ 47.067 \end{bmatrix} \le x^{1}(j) \quad j = 3, 4, 5, 6.$$

By Theorem 5.4, $P(X^1 | M, \rho)$ can be obtained by considering the first two elements of X^1 and the last four elements of X^1 independently. Consider the last four elements of X^1 . By the Pool-Adjacent-Violators algorithm, we have $P(X_2^1 | M_2)$ is the constant 75.717. The 4-component row vector $X_1^1 - \rho X_2^1$ is monotone increasing. By Theorem 5.2, $Y_2 = P(X_2^1 | M_2)$ and $Y_1 = X_1^1 + \rho(Y_2 - Y_2^1)$ where $Y = P(X^1 | M, \rho)$. Therefore,

$$P(\mathbf{X}^{1} | \mathbf{M}, \rho) = \begin{bmatrix} 13.333 \ 13.333 \ 13.348 \ 15.178 \ 15.178 \ 19.628 \\ 47.067 \ 47.067 \ 75.717 \ 75.717 \ 75.717 \ 75.717 \ 75.717 \end{bmatrix}$$

V.4 Approximation

Let X be a given $2 \ge n$ matrix and let $X^* = P(X \mid M, \rho)$. The process introduced here is that we may obtain a region A(j)such that $X^*(j) \in A(j)$, j = 1, ..., n by the Pool-Adjacent-Violators algorithm. But the process can be only applied to the case $\rho \ge 0$.

Let $a_1 = \min\{x_{1j} - \rho x_{2j}; j = 1, ..., n\},\$ $a_2 = \max\{x_{1j} - \rho x_{2j}; j = 1, ..., n\},\ \beta_1 = \min\{x_{2j} - \rho x_{1j}; j = 1, ..., n\},\$ $\beta_2 = \max\{x_{2j} - \rho x_{1j}; j = 1, ..., n\}$ and let 2 x n matrices X, Y, T and U be defined by

(5.10)
$$\begin{cases} z_{1j} - \rho z_{2j} = x_{1j} - \rho x_{2j} \\ z_{2j} - \rho z_{1j} = \beta_{1} \end{cases}$$
$$\begin{cases} y_{1j} - \rho y_{2j} = x_{1j} - \rho x_{2j} \\ y_{2j} - \rho y_{1j} = \beta_{2} \end{cases}$$
$$\begin{cases} t_{1j} - \rho t_{2j} = \alpha_{1} \\ t_{2j} - \rho t_{1j} = x_{2j} - \rho x_{1j} \end{cases}$$
and

$$u_{1j} - \rho u_{2j} = a_2$$

 $u_{2j} - \rho u_{1j} = x_{2j} - \rho x_{2j}$

for j = 1, ..., n. By the structure, $Z_2 - \rho Z_1, Y_2 - \rho Y_1, T_1 - \rho T_2$ and $U_1^{-\rho}U_2^{-\rho}$ are constant and hence monotone increasing. Theorem 5.1 shows that if $\rho \ge 0$, then their bivariate isotonic regressions Z*, Y*, T* and U* can be obtained by the Pool-Adjacent-Violators algorithm.

For each j, $VZ(j) \leq VX(j) \leq VY(j)$ and $VT(j) \leq VX(j) \leq VU(j)$. If $\rho \geq 0$, then by Theorem 5.3 we have

 $Z^{*}(j) \leq X^{*}(j) \leq Y^{*}(j)$

and

$$T*(j) \le X*(j) \le U*(j)$$

where X^* is the bivariate isotonic regression of X. Therefore

$$(5.11) Z*(j) \lor T*(j) \le X*(j) \le Y*(j) \land U*(j).$$

When $\rho = 0$, Theorem 5.1 shows that $Z_1^* = Y_1^* = P(X_1 | M_1)$, $Z_2^* = \beta_1$, $Y_2^* = \beta_2$, $T_1^* = \alpha_1$, $U_1^* = \alpha_2$ and $T_2^* = U_2^* = P(X_2 | M_2)$. It follows that $Z^*(j) \lor T^*(j) = Y^*(j) \land U^*(j)$ for $j = 1, \ldots, n$. If $\rho > 0$ and it is small enough, then the region A(j) determined by (5.11) is small for each $j = 1, \ldots, n$. In such a situation, a good approximation can be obtained.

(5.3) shows that $x_{1n} - \rho x_{2n} \leq x_{1n}^* - \rho x_{2n}^*$ and $x_{2n} - \rho x_{1n} \leq x_{2n}^* - \rho x_{1n}^*$. Combining (5.2) and (5.3), we have $x_{11} - \rho x_{21} \geq x_{11}^* - \rho x_{21}^*$ and $x_{21} - \rho x_{11} \geq x_{21}^* - \rho x_{11}^*$. Therefore, no matter whether $\rho \geq 0$ or $\rho < 0$, $VX(1) \geq VX*(1)$ and $VX(n) \leq VX*(n)$. For each i, $x_{11}^* \leq x_{1n}^*$ and $\sum_{j=1}^n x_{1j} = \sum_{j=1}^n x_{1j}^*$, it follows that $x_{11}^* \leq \overline{x}_1 \leq x_{1n}^*$ where $\overline{x}_1 = \sum_{j=1}^n x_{1j}/n$.

Example 5.2. Let X be the 2×5 matrix given below and let $\rho = 1/2$.

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 2 & 2 & 4 \\ 0 & 1 & 3 & 2 & 1 \end{bmatrix}$$

Let X^* be the bivariate isotonic regression of X. We would guess that $X^*(1) = X(1)$. Consider the last four elements of X. If the regions A(2), A(3), A(4) and A(5) determined by (5.11) for the last four elements are such that $a \in A(j)$ implies $a \ge X(1)$, j = 2, 3, 4, 5, then our conjecture X * (1) = X(1) has been verified.

Let Z, Y, T and U be the 2×4 matrices defined by (5.10) with respect to the last four components of X. They are

$$Z = 1/3 \begin{bmatrix} 0 & 0 & 2 & 12 \\ -3 & -3 & -2 & 3 \end{bmatrix}$$
$$Y = 1/3 \begin{bmatrix} 6 & 6 & 8 & 18 \\ 9 & 9 & 10 & 15 \end{bmatrix}$$
$$T = 1/3 \begin{bmatrix} 3 & 6 & 4 & 0 \\ 3 & 9 & 5 & -3 \end{bmatrix}$$
$$U = 1/3 \begin{bmatrix} 15 & 18 & 16 & 12 \\ 5 & 15 & 11 & 3 \end{bmatrix}.$$

and

Let Z^* , Y^* , T^* and U^* be the bivariate isotonic regressions of X, Y, T and U respectively. Then $Z^* = Z$, $Y^* = Y$,

$$T * = 1/9 \begin{bmatrix} 9 & 10 & 10 & 10 \\ 9 & 11 & 11 & 11 \end{bmatrix}$$
$$U * = 1/9 \begin{bmatrix} 45 & 46 & 46 & 46 \\ 27 & 29 & 29 & 29 \end{bmatrix}.$$

+

and

Therefore,
$$A(2) = \{(\alpha, \beta)^{t} : 1 \le \alpha \le 2, 1 \le \beta \le 3\},\$$

 $A(3) = \{(\alpha, \beta)^{t} : 10/9 \le \alpha \le 2, 11/9 \le \beta \le 3\},\$
 $A(4) = \{(\alpha, \beta)^{t} : 10/9 \le \alpha \le 8/3, 11/9 \le \beta \le 29/9\}$ and
 $A(5) = \{(\alpha, \beta)^{t} : 4 \le \alpha \le 46/9, 11/9 \le \beta \le 29/9\}.$ It follows that

\$

X*(1) = X(1).

For the remaining four elements, we have $VX*(2) \leq VX(2)$. In other words $X*(2) \leq X(2)$. Since $X*(2) \in A(2)$ and $X(2) \leq a$ for each $a \in A(2)$, we have X*(2) = X(2). In the remaining three elements of X, $X_1 - \rho X_2$ is monotone increasing. By Theorem 5.1, we can obtain the bivariate isotonic regression for the last three elements. Therefore,

$$\mathbf{X}^* = \begin{bmatrix} 1 & 1 & 3/2 & 2 & 9/2 \\ & & & & \\ 0 & 1 & 2 & 2 & 2 \end{bmatrix} . \quad []$$

BIBLIOGRAPHY

- Alexander, M.J. An Algorithm for Obtaining Maximumlikelihood Estimates of a Set of Partially Ordered Parameters. Rocketdyne Technical Paper. 1970.
- Ash, R. B. Real Analysis and Probability. Academic Press, New York. 1972. 476 p.
- Ayer, M., H.D. Brunk, G.M. Ewing, W.T. Reid and E. Silverman. An Empirical Distribution Function for Sampling with Incomplete Information. The Annals of Mathematical Statistics. 26:641-647. 1955.
- Barlow, R.E. and H.D. Brunk. The Isotonic Regression Problem and Its Dual. Journal of the American Statistical Association. 67:140-147. 1972.
- Barlow, R.E., D.J. Bartholomew, J.M. Bremner and H.D. Brunk. Statistical Inference Under Order Restrictions. Wiley, New York. 1972. 388 p.
- 6. Berberian, S.K. Measure and Integration. Chelsea, New York. 1965. 312 p.
- Bhattacharyya, G.K. and J.H. Klotz. The Bivariate Trend of Lake Mendota. Technical Report No. 98, Department of Statistics, University of Wisconsin, Madison, Wisconsin. 1966.
- 8. Blackwell, D. Conditional Expectation and Unbiased Estimation. The Annals of Mathematical Statistics. 18:105-110. 1947.
- Breiman, L. Probability. Addison-Wesley, Reading, Massachusetts. 1968. 421 p.
- Brunk, H.D. Maximum Likelihood Estimates of Monotone Parameters. The Annals of Mathematical Statistics. 26:607-616. 1955.
- Brunk, H.D. Best Fit to a Random Variable by a Random Variable Measurable with Respect to a σ-Lattice. Pacific Journal of Mathematics. 11:785-802. 1961.

- Brunk, H.D. On an Extension of the Concept Conditional Expectation. Proceeding of the American Mathematical Society. 14:298-304. 1963.
- Brunk, H.D. Conditional Expectation Given a σ-Lattice and Applications. The Annals of Mathematical Statistics. 36:1339-1350. 1965.
- Brunk, H.D., G.M. Ewing and W.R. Utz. Minimizing Integrals in Certain Classes of Monotone Functions. Pacific Journal of Mathematics. 7:833-847. 1957.
- Brunk, H.D. and S. Johansen. A Generalized Radon-Nikodym Derivative. Pacific Journal of Mathematics. 34:585-617. 1970.
- Darst, R.B. The Lebesgue Decomposition, Randon-Nikodym Derivative, Conditional Expectation, and Martingale Convergence for Lattices of Sets. Pacific Journal of Mathematics. 35:581-600. 1970.
- Dixon, W.J. and F.J. Massey, Jr. Introduction to Statistical Analysis. McGraw-Hill, New York. 1957. 638 p.
- Draper, N.R. and H. Smith. Applied Regression Analysis. Wiley, New York. 1966. 407 p.
- Hadley, G. Linear Programming. Addison-Wesley, Palo Alto. 1962. 520 p.
- Halmos, P.R. Measure Theory. Van Nostrand, Princeton, N.J. 1956. 304 p.
- Halmos, P.R. Finite-Dimensional Vector Spaces. Van Nostrand, Priceton, New Jersey. 1958. 199 p.
- 22. Halmos, P.R. Naive Set Theory. Van Nostrand, New York. 1960. 104 p.
- Halmos, P.R. and L.J. Savage. Application of the Radon-Nikodym Theorem to the Theory of Sufficient Statistics. The Annals of Mathematical Statistics. 20:225-241. 1949.
- 24. Hartigan, J.A. Representation of Similarity Matrices by Trees. Journal of the American Statistical Association. 62:1140-1158. 1967.

- 25. Johansen, S. The Descriptive Approach to the Derivative of a Set Function with Respect to a σ -Lattice. Pacific Journal of Mathematics. 21:49-58. 1967.
- John, P. W. M. Statistical Design and Analysis of Experiments. Macmillan, New York. 1971. 356 p.
- Mangasarian, O.L. Non-linear Programming. McGraw-Hill, N.Y. 1969. 220 p.
- Neveau, J. Mathematical Foundations of the Calculus of Probability. Holden-Day, San Francisco. 1965. 223 p.
- Pervin, W.J. Foundations of General Topology. Academic Press, New York. 1964. 209 p.
- Rao, C.R. Linear Statistical Inference and Its Applications. Wiley, New York. 1965. 522 p.
- Reid, W.T. A Simple Optional Control Problem Involving Approximation by Monotone Functions. Journal of Optimization Theory and Applications. 2:365-377. 1968.
- Robertson, T. A Smoothing Property for Conditional Expectations Given σ-Lattices. The American Mathematical Monthly. 75:515-518. 1968.
- Rockafellar, R.T. Convex Analysis. Princeton University Press. 1970. 451 p.
- Rudin, W. Real and Complex Analysis. McGraw-Hill, New York. 1966. 412 p.
- Scheffe, H. The Analysis of Variance. Wiley, New York. 1959. 477 p.
- 36. Searle, S.R. Linear Models. Wiley, New York. 1971. 532 p.
- Taylor, A.E. Introduction to Functional Analysis. Wiley, New York. 1958. 423 p.
- Thompson, W.A., Jr. The Problem of Negative Estimates of Variance Components. The Annals of Mathematical Statistics. 33:273-289. 1962.

APPENDICES

APPENDIX I

_	i	1	2	3	4	5	6	7	8	9	10
March	i	48°	48°	45°	44°	5 2°	48°	43°	44°	48°	56°
	10+i	51°	55°	46°	47°	5 3°	68°	5 8°	56°	5 8°	60°
	20+i	60°	65°	60°	57°	63°	62°	68°	5 2°	55°	61°
	30+i	5 2°									
- April	i	49°	5 3°	50°	5 2°	61°	55°	55°	62°	55°	46°
	10+i	5 3°	51°	56°	65°	69°	56°	59°	63°	5 8°	55°
	20+i	60°	65°	56°	54°	60°	61°	59°	59°	59°	70°

(Daily Maximum) Temperatures Measured (in Fahrenheit) at Oregon State University in 1974 (cf. Section I. 2)

APPENDIX II

Combined Explosive Rates (cf. Section I.2)

 r_j explosive rate of ten samples dropped at the jth height. $r_1 = 0.30$, $r_2 = 0.20$, $r_3 = 0.70$, $r_4 = 0.80$ and $r_5 = 0.50$. Q(i,j) explosive rate if 10x|j-i+1| samples were dropped at the jth height which were dropped at heights between the ith height and the jth height.

$$q_{ij} q_{ij} = q_{ji}, q_{ij} = (r_i + ... + r_j)/(j-i+1)$$
 if $i \le j$.

		- ij -	((-, j, -,	
j	1	2	3	4	5
1	0.30	0.25-	0.40-	0.50 -	0.50-
2	0.25+	0.20	0.45-	0.57-	0.55-
3	0.40+	0.45+	0.70	0.75-	0.67-
4	0.50+	0.57+	0.75+	0.80	0 65 -
5	0.50+	0.55+	0.67+	0.65+	0.50
	j 1 2 3 4 5	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$

Table of $q_{..}$'s (and Q(i, j)'s)

 $Q(i,j) \ge q_{ij}$ if $i \le j$, $Q(i,j) \le q_{ij}$ otherwise.

APPENDIX III

MLE of Variance Components (cf. Section I.2)

 $y_{ijk} = \mu + a_i + b_j + c_{ij} + e_{ijk} \quad i = 1, \dots, r; \ j = 1, \dots, s;$ $k = 1, \dots, t \quad \text{where} \quad \{a_i\}, \ \{b_j\}, \ \{c_{ij}\} \quad \text{and} \quad \{e_{ijk}\} \quad \text{are mutually}$ independent sets of normal variates with variances $\sigma_A^2, \ \sigma_B^2, \ \sigma_{AB}^2$ and σ^2 and with means zero. Let y be the rst-component column vector which is composed of $\{y_{ijk}\}$. Let

L = likelihood function

$$= (2\pi)^{-\frac{1}{2}\mathbf{r}\mathbf{st}} (\det(\Delta))^{-\frac{1}{2}} \exp\{-\frac{1}{2} (y-\mu)^{t} \Delta^{-1} (y-\mu)\}$$

and

$$\lambda = -2 \log(L)$$

= $(y-\mu)^{t} \Delta^{-1}(y-\mu) + \log \det(\Delta) + \text{constant}$

where $\Delta = Cov(y, y)$. Let z be the rst-component column vector which is composed of

$$y... = a. + b. + c.. + e...$$

$$y_{i} \cdot - y \cdot \cdot = a_{i} - a_{i} + c_{i} - c_{i} + e_{i} - e_{i} \quad i = 1, \dots, r-1$$

$$y_{j} - y_{i} = b_{j} - b_{i} + c_{j} - c_{i} + e_{j} - e_{i} \quad j = 1, \dots, s-1$$

$$y_{ij} - y_{i} - y_{j} - y_{i} = c_{ij} - c_{i} - c_{j} + c_{i} + e_{ij} - e_{i} - e_{j} + e_{i}$$

$$i = 1, \dots, r-1; \quad j = 1, \dots, s-1$$

$$y_{ijk}^{-y}_{ij}$$
 = $e_{ijk}^{-} e_{ij}$.
i = 1,...,r; j = 1,...,s; k = 1,...,t-1

where a, b, c_i, c_j, c_i, e_i, e_j, e_{ij}, e_{ij}, y_{i...}, y

$$y_{ijk} = y_{...} + (y_{i..} - y_{...}) + (y_{.j} - y_{...}) + (y_{ij} - y_{i..} - y_{.j} + y_{...})$$

+ $(y_{ijk} - y_{ij.})$

if i < r, j < s, t < k. For i = r, the $y_{r..} - y_{...}$ term is replaced by $\sum_{i=1}^{r-1} - (y_{i..} - y_{...})$ and the $y_{rj.} - y_{r..} - y_{.j.} + y_{...}$ term is replaced by $\sum_{i=1}^{r-1} - (y_{ij.} - y_{i..} - y_{.j.} + y_{...})$. Similarly for the cases j = s and t = k. Therefore

$$\lambda = \sum_{a} f_{a} (\log \theta_{a} + MS_{a} / \theta_{a}) + constant$$

where f_a is the degree of freedom associated with MS_a and θ_a is the expectation of MS_a for each a = A, B, AB, e.

(I) max L
subject to
$$\theta_{e} \leq \theta_{AB} \leq \theta_{A}$$
, $\theta_{AB} \leq \theta_{B}$
(II) min λ
subject to $\theta_{e} \leq \theta_{AB} \leq \theta_{A}$, $\theta_{AB} \leq \theta_{B}$
(III) min $\Sigma_{a}f_{a}(\log \theta_{a} - \log MS_{a} - 1 + MS_{a}/\theta_{a})$
subject to $\theta_{e} \leq \theta_{AB} \leq \theta_{A}$, $\theta_{AB} \leq \theta_{B}$

The problems (I), (II) and (III) are equivalent. Let $\Phi(x) = -\log x$. Then Φ is convex. Let

$$\Delta(\mathrm{MS}_{a}, \theta_{a}) = \log \theta_{a} - \log(\mathrm{MS}_{a}) + (\mathrm{MS} - \theta_{a})/\theta_{a}.$$

By Theorem 1.10 of Barlow and coworkers (1972), (III) is equivalent to (IV).

(IV)
$$\min \Sigma_{a} f_{a} (MS_{a} - \theta_{a})^{2}$$

subject to $\theta_{e} \leq \theta_{AB} \leq \theta_{A}, \ \theta_{AB} \leq \theta_{B}$

APPENDIX IV

Bounded Isotonic Regression (cf. Example 3.6)

min $f(Z) = \sum_{i=1}^{10} (z_i - x_i)^2$										
subject to $g_i(Z) = 10 + 1.5(i-1) - z_i \le 0$ $i = 1,, 10$										
$g_{10+i}(Z) = z_i - 13 - 1.5(i-1) \le 0$ $i = 1,, 10$										
and $g_{20+i}(Z) = z_i - z_{i+1} \le 0, i = 1,, 9$										
i	1	2	3	4	5	6	7	8	9	10
×	25	13	2	15	14	21	9	33	25	15
z _{li}	10	11.5	13	14.5	16	17.5	19	20.5	22	23.5
^z 2i	13	14.5	16	17.5	19	20.5	22	23.5	25	26.5
× x _i	13.33	13.33	13.33	14.5	14.5	15	15	24.33	24.33	24.33
x.	13	13	13	15	16	19	19	23.5	23.5	23.5
λ _i	0	0	22	0	4	0	16	0	0	14
λ _{10+i}	24	0	0	0	0	0	0	19	0	0
^λ 20+i	0	0	0	0	0	4	0	0	3	

 λ_{i} , i = 1,...,29 are Lagrangian multipliers

$$\nabla f(\hat{\mathbf{X}}) = (-24 \quad 0 \quad 22 \quad 0 \quad 4 \quad -4 \quad 20 \quad -19 \quad -3 \quad 17)$$

$$\Sigma_{i=1}^{10} \lambda_i \nabla g_i(\hat{\mathbf{X}}) = (\quad 0 \quad 0 \quad -22 \quad 0 \quad -4 \quad 0 \quad -16 \quad 0 \quad 0 \quad -14)$$

$$\Sigma_{i=11}^{20} \lambda_i \nabla g_i(\hat{\mathbf{X}}) = (\quad 24 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 19 \quad 0 \quad 0)$$

$$\Sigma_{i=21}^{29} \lambda_i \nabla g_i(\hat{\mathbf{X}}) = (\quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 4 \quad -4 \quad 0 \quad 3 \quad -3)$$

Kuhn-Tucker condition

$$g_{i}(\hat{\mathbf{X}}) \leq 0, \quad i = 1, \dots, 29$$
$$\lambda_{i}g_{i}(\hat{\mathbf{X}}) = 0, \quad i = 1, \dots, 29$$
$$\nabla f(\hat{\mathbf{X}}) + \Sigma_{i=1}^{29} \lambda_{i} \nabla g_{i}(\hat{\mathbf{X}}) = 0$$

.



Isotonic Regression Blocking

 $\hat{x}_1, \hat{x}_5, \hat{x}_7$ and \hat{x}_8 are determined by Lemma 3.1 × -- X value * -- X* value, X* is the isotonic regression of X 0 -- \hat{X} value, \hat{X} is the bounded isotonic regression of X

APPENDIX V Simplified Projection Algorithm (cf. Section V.3) PROGRAM LEET SIMPLIFIED PROJECTION ALGORITHM DIMENSION X(2,230) PX(2,230) Y1(233) Y2(233) V(2) INTEGER PP(2,231) I1(201) I2(201) I0(2) P,3 RFAD(60,14) N, R, V, ID FORMAT(I5,3F1) 6,2A2) IF(FOF(60)) CALL EXIT PEAD(60,13) (Y(1,3), J=1,N) PFAD(60,13) (Y(1,3), J=1,N) FORMAT(I3F3,3) N1=V+1 13 14 FURMAT(1)F3.3) N1=N+1 D0 2P I=1,2 BP(I,N1)=1H3 IF(ID(T).FD.2HIN)GO TO 25 R==5 19 V(T) =- V(I) 25 00 29 J=1,N PX(I,J)=V(I)*X(T,J) BP(I,J)=1H3 29 CONTINUE 0000 THIS PROGRAM IS FOR P > 3. DETERMINE PLYOTAL INDICES. IF (P.LT. 7.)GO TO 96 IMC= 0 D0 47 J=2, N D1=FX(1, J) - PX(1, J-1) P2=PX(2, J) - PX(2, J-1) I1(J)=I2(J)=9 IF(3P(1, J) - E0. 1HP)GO TO 43 IF(01.GT.3..09.01.GT.F*D2)GO TO 43 BP(1, J)=1HP IMC=1 IF(2P(2, J) - E 34 47 CONTINUE CCCC IF NO INCREASING IN NUMBER OF PIVOTAL INDICES, PX IS THE OPTIMAL SOLUTION. IF(INC.50.3)60 F0 87 P=0=0 I1(1)=I2(1)=NL=1 D0 85 NR=2,N1 J=NP=NL+1 IF(3P(1,NR).E0.1HP)60 T0 5° I1(J)=1 P=0+1 60 Y1 (J)=PX(1,NR)-R*PX(2,NP) IF(BP(2,NR).E0.1HP)G0 TO 54 I?(J)=1 10 (J) - 1 Q= Q+1 64 Y2 (J)=PX (2,NR) - R*PX (1,NR) IF (I1(J) - F3.0.0R+I2(J) - F3.0)G1 T0 85 Y1 (1)=PX (1,NL) - R*PX (2,NL) Y2 (1)=PX (2,NL) - R*PX (1,NL) NO=NR-NL 000 DECOMPOSITION TECHNIQUE. IF (P.LT.C) GO TO 75 GALL PROJ(ND,R,I1,I2,Y1,Y2,C) GO TO 76 75 CALL PROJ(ND,R,I2,I1,Y2,Y1,C) 76 DO BI J=1,ND K=J+NL-1 PX (1,K)=Y1(J) PX (2,K)=Y2(J) 80 CONTINUE

```
NL=NR

P=0=C

D0 85 I=2.N

11(I)=I2(I)=0

85 C0NIINUF

G0 T0 34

37 D0 91 I=1.2

V(I)=1./V(I)

D1 C0NIINUE

W0 ITE(61, 33)(J,X(1,J),PX(1,J),X(2,J),PX(2,J),J=1,N)

93 F0RMAT(f1f,T12,fX(1,J)f,T24,fPX(1,J)f,T41,fX(2,J)f,T53,fPX(2,J)f,/

1(f1f,I3,2F13.6,3X,2F13.6))

G0 T0 13

96 WRITE(61,97)

37 F0RMAT(f1IMPROPER R VALUEf)

G0 T0 13

96 WRITE(61,97)

37 F0RMAT(f1IMPROPER R VALUEf)

G0 T0 13

96 WRITE(61,97)

37 F0RMAT(f1IMPROPER R VALUEf)

60 T0 13

96 WRITE(61,97)

97 F0RMAT(f1IMPROPER R VALUEf)

60 T0 13

96 WRITE(61,97)

97 F0RMAT(f1IMPROPER R VALUEf)

60 T0 13

F0RMAT(f1IMPROPER R VALUEf)

F0 T0 13

F0 T0 13
```

.

```
SUBROUTINE PROJ(N9,R, J1, J2, 31, 82, M)

DIMENSION J1(201), J2(201), 81(200), 82(200), 8(100), C(100, 100),

1CJ(200), Z(10),

ND1=N0+1
                    IL=1
                   00 119 1=2,N01
IF(J1(I).EQ.0)G0 TU 119
      IF (J1(I) .EQ.0)GO T

IR=I-1

AV=0.

D0 113 J=IL.IP

AV=AV+91(J)

113 CONTINUE

AV=(AV+P)/FLOAT(K)

D0 117 J=IL,IP

B2(J)=32(J)+AV

117 CONTINUE

IL=I

119 CONTINUE

K=0
                   K=0

0 127 I=1,N0

IF(J2(I).F0.0)G0 TO 127

K=K+1
     K=K+1
B(K)=C.
DO 127 J=I.ND
P(K)=3(K)+32(J)
127 CONTINUE
000
                                                         MXM SIMULTANEOUS LINEAR EQUATIONS.
              (5.5)
                                       CZ=3
    (5.5)
L=9
RSQ=P*?
70 164 JJ=1.N0
IF(J2(JJ).E9.J)60 TO 164
D0 137 I=1.JJ
CJ(I)=9.
137 CONTINUE
D0 140 I=JJ.ND
CJ(I)=1.
140 CONTINUE
IL=1
D0 155 I=2.N01
IF(J1(I).E9.7)60 TO 155
IR=T-1
K=I-IL
AV=C.
CO 140 J=IL.IP
      K=I-IL
AV=C.
DO 149 J=IL, IP
AV=AV+CJ(J)
149 CONTINUE
AV=(AV*RSQ)/FLOAT(K)
DO 153 J=IL, IR
CJ(J)=CJ(J)-AV
153 CONTINUE
IL=I
155 CONTINUE
K=3
L=1+1
                    E=E+1
      L=L+1

DO 164 T=1,40

IF(J2(I).E1.0)G0 TO 164

K=K+1

C(K,L)=C.

DO 164 J=I.40

G(K,L)=C(K,L)+CJ(J)

164 CONTINUE
             C IS AN MXM SYMMETRIC NONSINGULAR MATRIX WITH POSITIVE VALUE IN
EACH ENTRY SUCH THAT EACH DIAGONAL ENTRY HAS LARGER VALUE THAN
THOSE ON THE SAME COLUMN.
00000
       M1=M-1

DJ 179 J=1,41

JP=J+1

IF (JP.GT.4)50 TO 179

DU 179 I=J9,4

F=C(I,J)/C(J,J)

B(I)=3(I)=5*3(J)

DO 179 K=J9,4

C(I,K)=C(I,K)-F*C(J,K)

179 CONTINUE

7(4)=3(M)/C(M.M)
                     2(M)=3(M)/C(M,M)
```

1

```
D0 189 I=2,4

K=4+1-T

SB=C

K1=K+1

D0 197 J=K1,4

SB=SB+C(K,J)+Z(J)

147 CONTINUE

Z(K)=(R(K)-S3)/C(K,K)

149 CONTINUE

C Z IS THE SOLUTION CF (5.6).

7 IS THE INCREMENT OF 32 GIVEN BELOW.

C 32 AND 31 ARE THE SOLUTIONS IN (5.7) AND (5.8) RESPECTIVELY.

SUM=J=3

D0 201 I=1,ND

IF(J2(I)+E3+3)GD TO 2J0

J=J+1

SUM=ZUM+Z(J)

201 CONTINUE

IL=1

D0 215 I=2,ND1

IF(J1(I)+E3+3) SO TO 215

IR=I-1

K=I-1

AV=AV+B1(J)+3*B2(J)

210 CONTINUE

AV=AV+B1(J)+3*B2(J)

211 CONTINUE

AV=AV+FLOAT(<)

D0 214 J=IL,IR

B1(J)=AV

214 CONTINUE

RETUPN

FND
```

.

...

١

ļ

INDEX

affine space 15 basic function 42 basic set 41 bivariate isotonic regression 124 boundary point 12 bounded from above 34 bounded from below 34 bounded isotonic regression 79 chain 60 closed 12 comparable 33 complete lattice 85 component 35 conditional expectation given a σ -lattice 100 cone 19 connected 35 convergence in L, 96 convex 12 convex hull 39 covariance 97 dual cone 22 equivalence class 36 essential infimum 104 essential supremum 104 extreme vector 26 field 85 finer 88 finitely generated cone 25 generalized projection 13 generated σ -field 85 generated σ -lattice 85 greatest lower bound 34 Hilbert space 12 immediate predecessor 36 immediate successor 36 immediately comparable pair 36 indicator 41

induced complete lattice 86 induced ordering 86 induced σ -field 93 induced σ -lattice 93 interior point 26 isotonic function 36 isotonic regression 37 Jensen's inequality 110 lattice 85 least upper bound 34 level set 41 linear ordered σ -lattice 92 linear ordering 33 lower bound 34 lower set 41 martingale 18 maximal element 33 Maximum Lower Set algorithm 65 minimal element 34 Minimum Upper Set algorithm 62 minimum violator 62 Minimum Violator algorithm 62 monotone class 85 monotonicity 103 multivariate isotonic regression 121 norm 12 ordered set 34 partial ordering 33 partition 66 pivotal element 62, 134, 135 Pool-Adjacent-Violators algorithm 57 probability space 93 projection 12 Projection of Minimum Violators algorithm 53

```
Projection of Violators
 algorithm 57
quasi-ordering 33
random variable 93
reached 35
relative boundary point 27
relative interior point 27
restricted conditional
 expectation 111
restricted isotonic regression 37
reversal of an ordered set 35
reversal of an ordering 35
reversed tree 35
root 60
separable 35
\sigma-field 85
\sigma-lattice 85
\Sigma-measurable 93
simple random variable 94
Simplified Projection algorithm 141
smoothing property 17, 109
submartingale 118
tree 34
tree structure 34
unimodal 79
unrelated 35
upper bound 34
upper set 41
upper set determined by \omega 41
upper set in C 67
variance 97
violator 53,62
weight function 36
```