<**Research Paper**>

Journal of the Korean Statistical Society
Vol. 19, No. 1, 1990

The Algorithm of Sweep-the-Negatives and its Applications to Order Restricted Inference

Byong-Dok Kim*, Jae-Rong Choi** and Akio Kudô***

ABSTRACT

Consider the extreme problem: $\min(\mu - y)'(\mu - y)$ subject to $A\mu \geq 0$, where A is an $n \times p$ matrix, which often occurs in solving the maximum likelihood estimator with ordered restrictions in parameter space. In case the matrix AA' has all non-positive off-diagonal elements, some propositions in this paper guarantee that the above extreme solutions are achieved at most at n sweep out steps in Gaussian eliminations. Some typical examples of Sweep-the-Negatives method are given.

1. Introduction

Statistical inference under the ordered restrictions frequently occurs in data analysis, especially in biological treatments, psychological works, and social sciences. The order restricted inference is said to have been started by Bartholomew(1959 a, b). These are the first papers in this area which widely drew attentions of statistical community. Kudô(1963) considered the problem of statistical inference under the condition that all the components of the mean vector are positive, give an n-variate normal distribution with the known variance matrix, by making use of the properties of convex cone. The above works and their applications are described in Barlow et al.(1972).

* Department of Mathematics, Chang-won University, Chang-won 641-773, Korea
** Department of Mathematics, Dong-A University, Pusan 604-714, Korea
*** Department of Mathematics, Kyushu University 33, Fukuoka 812, Japan

Later Kudô and Choi(1975) have generalized the previous result to the case when the variance matrix is singular. In the process of developing in statistical inference under the ordered restrictions, it is remarkable that the derivations of test functions and optimum estimators are very difficult. Barlow et al.(1972) introduced the Pool-Adjacent-Violators Algorithm(PAVA) which solves the maximum likelihood estimator (M.L.E.) of parameters in case of simple or simple tree order. Recently Dykstra et al.(1986) and Robertson et al. (1988) have proposed some methods for the estimations of parameters under the ordered restrictions.

The PAVA finds its justification in its graphical interpretation. It uses the greatest convex minorant(GCM) of the cumulative sum diagram (CSD). On the other hand, the Sweep -the-Negatives algorithm (Choi, 1976a) bases on the characterization of the location of the projection of a point to a convex polyhedral cone given by Kudô(1963) and further by Kudô and Choi (1975).

The purpose of this paper is to illustrate explicitly how the algorithm developed in Choi(1976a, b) works and to demonstrate it in more details. We are engaged in geometric or linear algebraic approach to formulate the statistical inference with order restrictions. Our main proposition with respect to convex polyhedral cone prepares the method for solving the estimation problems by applying Gaussian elimination (Kennedy and Gentle, 1980). Furthermore we obtain a simple algorithm solving the estimation problem in case the matrix AA' has all non-positive off-diagonal elements. Our algorithm can be widely applied to simple or simple tree order, upper starshaped problem (Shaked, 1979), and duality.

2. Geometrical Properties of the Polyhedral Convex Cone

Let $a_1, a_2, \cdots, a_n$ be a set of p-vectors satisfying the condition that any subset of size less than or equal to p from n forms an independent set of vectors. Let $A=(a_1, a_2, \cdots, a_n)'$ be an $n \times p$ matrix and let $N=\{1, 2, \cdots, n\}$ be an index set. Let

$$C_N=\{z \mid z \in R^p, (a_i, z) \geq 0, i \in N\}$$

be a closed convex polyhedral cone determined by $a_1, a_2, \cdots, a_n$. C_N is denoted by C for simplicity. Given a point $y \in R^p$, the distance $d(y, C)$ between the point y and the set C is defined by $d(y, C)=\inf \| y-z \| = \inf \sqrt{(y-z, y-z)}$, where infimum is taken for all the values of z in C.

If $y \notin C$, there is a point $y_0 \in C$ such that $d(y, C) = d(y, y_0)$. If $y \in C$, it is evident that $d(y, C) = 0$ and $y_0 = y$. The point y_0 is called the projection of y on the convex cone C. The closedness of the cone confirms the existence of the projection y_0 and the convexity also confirms the uniqueness of y_0. Let M denote a subset of N, and let m be the size of the set M. Now let F_M be a subset of C, defined by

$$F_M = \begin{cases} \{z \mid z \in R^p, (a_i, z) = 0, i \in M; (a_i, z) > 0, i \in N - M\}, & m < \min(p, n) \\ \{z \mid z \in R^p, (a_i, z) = 0, i \in N\}, & m \geq \min(p, n) \end{cases} \tag{2.1}$$

which may be an empty set in some cases.

Remark 2.1 In case $m = n < p$, F_M is a $(p-n)$-dimensional hyperplane and in case $p \leq m \leq n$, $F_M = \{0\}$.

Hereafter we take $M = \{1, 2, \cdots, m\}$, $m \leq p$, without loss of generality. Furthermore let us define the following:

S_M: the linear subspace spanned by $\{a_i \mid i \in M\}$.

$B_M = \{z \mid z \in R^p, (a_i, z) = 0, i \in M\}$: the linear space orthogonal to S_M.

$a^{(1)}_M, a^{(2)}_M, \cdots, a^{(m)}_M$: vectors contained in S_M, such that $(a^{(i)}_M, a_j) = \delta_{ij}$ for all $i, j \in M$, where δ_{ij} is Kronecker's delta.

$D_M = \{z \mid z \in R^p, (a^{(i)}_M, z) \leq 0, i \in M\}$.

$\hat{y}_M$: the projection of y to B_M.

$C^*_M = \{z^* \mid z^* \in R^p, (z^*, z) \leq 0, z \in C_M\}$: the dual cone of C_M.

In what follows, we describe some propositions with respect to the convex cone, using the above definitions and notations. The proofs are omitted.

Proposition 2.1 In case $p \leq n$, $D_M \cap S_M = C^*_M$. Particularly $D_N = C^*_N$ in case $m = p = n$.

Proposition 2.2

$$C_N = F_N \cup \Big(\sum_{\phi \subset M \subset N} F_M \Big), \tag{2.2}$$

where the summation runs over all the subsets M of N with $m < \min(n, p)$.

Remark 2.2 The closed convex cone is partitioned into disjoint sum of the convex cones, and the problem is, for given y, to dertermine which one of these cones the projection of y belongs to.

Proposition 2.3

$$\left(\begin{pmatrix} a'_1 \\ \vdots \\ a'_m \end{pmatrix} (a_1,\cdots,a_m)\right)^{-1} \begin{pmatrix} a'_1 \\ \vdots \\ a'_m \end{pmatrix} = \begin{pmatrix} a_M^{(1)\prime} \\ \vdots \\ a_M^{(m)\prime} \end{pmatrix}. \tag{2.3}$$

We will give a necessary and sufficient condition that the projection of y lies in a disjoint face of the cone F_M. Corresponding to M, let us partition the matrix A into the following form,

$$A=(a_1,\ a_2,\cdots,a_m;\ a_{m+1},\cdots,a_n)'=(A'_1,\ A'_2)'. \tag{2.4}$$

Proposition 2.4 The necessary and sufficient condition that for a given point y, the projection y_0 of y on the convex polyhedral cone C, is contained in F_M is given by

$$d_1=(A_1\ A'_1)^{-1}\ A_1 y\leq 0, \tag{2.5}$$

and

$$d_2=A_2\hat{y}_M>0, \tag{2.6}$$

where

$$\hat{y}_M=[I-A'_1(A_1\ A'_1)^{-1}\ A_1]y.$$

And in this case, $y_0=\hat{y}_M$.

Remark 2.3 In case m=0, that is, $M=\phi$, the condition (2.5) is vacuous and the inequality (2.6) is reduced to $(a_i,\ y)>0,\ i\in N$, and in case M=N, the condition (2.6) is vacuous and only (2.5) remains.

Remark 2.4 In case m=p, $y_0=0$, because any subset of $\{a_1,\cdots,a_n\}$ with size p spans the entire space R^p.

Remark 2.5 According to (2.3), (2.5) equals to $(a^{(i)}_M,\ y)\leq 0$ for all $i\in M$.

Proof. In case n=p and $\{a_1,\ a_2,\cdots,a_n\}$ are independent, the proof of this proposition was given for $0\leq m<n$ by Kudô(1963); in case $n\geq p,\ 0\leq m<n$, that was given by Kudô and

Choi(1975); in case of $m=n=p$ that was given in Choi(1975); and also we shall easily prove that in case $m \leq n < p$ as in Kudô and Choi(1975).

3. The M.L.E. under the Ordered Restrictions

Proposition 2.4 enables us to provide an algorithm of a special convex programming necessary for calculating estimate under the ordered restrictions. Our algorithm consists of checking out the condition (2.5) and (2.6) in Proposition 2.4 on the process of Gaussian elimination with a pivot matrix $A_1A'_1$, where $A_1=(a_1, a_2, \cdots, a_m)'$ for all subsets $M \subset N$.

Now we state a method to get the M.L.E. of the mean value under the ordered restrictions. Let a random vector $Y=(Y_1, Y_2, \cdots, Y_p)'$ have a multivariate normal distribution with the mean vector μ and the covariance matrix I_p:

$$Y \sim N_p(\mu, I_p). \tag{3.1}$$

Considering the transformation $X=AY$, we have $X \sim N_n(\theta, D)$, where $\theta=A\mu$ and $D=AA'$. The M.L.E. under the condition that $\theta_1=\theta_2=\cdots=\theta_m=0$ $(m<p)$ is clearly nothing but the projection of y on B_M.

For a given y, finding the M.L.E. $\hat{\mu}$ of μ under the condition $A\mu \geq 0$ is equivalent to finding the solution of

$$\min_{A\mu \geq 0}(\mu-y)'(\mu-y). \tag{3.2}$$

If D is non-singular $(n=p$ in this case), then (3.2) is equivalent to

$$\min_{\theta \geq 0}(\theta-x)' D^{-1}(\theta-x) \tag{3.3}$$

for a given x and a positive definite matrix D. The solution $\hat{\mu}$ of the minimizing problem (3.2), regardless of D being non-singular or singular, is guaranteed to exist and it satisfies the conditions (2.5) and (2.6) for a subset M, because the existence and the uniqueness are guaranteed by the closedness and convexity of the cone.

We consider the partitions of x and D corresponding to (2.4):

$$x=\begin{pmatrix} x_{(1)} \\ x_{(2)} \end{pmatrix} = \begin{pmatrix} A_1y \\ A_2y \end{pmatrix}, \quad D=\begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}(A'_1, A'_2), \tag{3.4}$$

where $A_1=(a_1, a_2,\cdots,a_m)'$ and $A_2=(a_{m+1},\cdots,a_n)'$.

Now consider the following matrix

$$\begin{pmatrix} D_{11} & D_{12} & x_{(1)} \\ D_{21} & D_{22} & x_{(2)} \\ A'_1 & A'_2 & y \end{pmatrix}. \tag{3.5}$$

Sweeping out (3.5) taking D_{11} as the pivot means post multiplying the following matrix to the matrix (3.5):

$$\begin{pmatrix} D^{-1}_{11} & 0 & 0 \\ -D_{21}D^{-1}_{11} & I & 0 \\ -A'_1D^{-1}_{11} & 0 & I \end{pmatrix}.$$

And after simple calculation, we have

$$\begin{pmatrix} I_m & D^{-1}_{11}D_{12} & D^{-1}_{11}x_{(1)} \\ 0 & D_{22}-D_{21}D^{-1}_{11}D_{12} & x_{(2)}-D_{21}D^{-1}_{11}x_{(1)} \\ 0 & A'_2-A'_1D^{-1}_{11}D_{12} & y-A'_1D^{-1}_{11}x_{(1)} \end{pmatrix} = \begin{pmatrix} I_m & * & d_1 \\ 0 & * & d_2 \\ 0 & * & d_3 \end{pmatrix}. \tag{3.6}$$

Proposition 2.4 indicates that the projection y_0 is located on F_M if and only if $d_1 \leq 0$ and $d_2 > 0$ in (3.6). At the same time, it is easily shown that d_3 equals to $\hat{\mu}$, the solution of (3.2), and also

$$\begin{pmatrix} 0 \\ x_{(2)}-D_{21}D^{-1}_{11}x_{(1)} \end{pmatrix} = \begin{pmatrix} 0 \\ d_2 \end{pmatrix}$$

equals to $\hat{\theta}$, the solution of (3.3). The index subset M satisfying the conditions $d_1 \leq 0$ and $d_2 > 0$ is unique, because the projection is unique and the cone is partitioned disjointly by Proposition 2.2.

Consequently, in order to find the M.L.E. of mean vectors, we have to compute Gaussian elimination of (3.5) for all the combinations of the index subset M and examine the signs of the upper parts of the last column of (3.6). For some large number n, however, it is meaningless to compute Gaussian elimination. In the next section we describe a special case.

4. Sweep-the-Negatives

In the algorithm mentioned in the previous section, we have to apply a systematic trial and error method in order to find out the subset M of N, where $i \in M$ implies $\hat{\theta}_i = 0$. Hence we might have as many as 2^n-1 of Gaussian elimination before arriving at the solution in a general case.

Choi(1976a) gave us a simple algorithm under the condition that all the off-diagonal elements of D are non-positive. We call this method Sweep-the-Negatives algorithm.

In order to demonstrate the algorithm of Sweep-the-Negatives, we introduce two lemmas with respect to some positive semidefinite matrix, and making use of these lemmas, we prove Proposition 4.3 which gives us an efficient Gaussian elimination method for our problem.

Lemma 4.1 Let $D = (d_{ij})$ be an $n \times n$ positive definite matrix with $d_{ij} \leq 0$ $(i \neq j)$ and let D^{-1} be the inverse of D. Then the elements of D^{-1} are all non-negative.

Proof. We prove this by the inductive form. When $n=1$, the lemma is obvious. Suppose that the lemma holds when the order of D is n-1. We consider the case when the order is n $(n \geq 2)$. Let D be partitioned as

$$D = \begin{pmatrix} D_1 & d_0 \\ d'_0 & d_{nn} \end{pmatrix}. \tag{4.1}$$

where D_1 is an $(n-1) \times (n-1)$ positive definite matrix satisfying the inductive assumption and d_0 is a column (n-1)-vector with $d_0 \leq 0$ and also $d_{nn} > 0$. From (4.1), we have $|D| = (d_{nn} - d'_0 D^{-1}_1 d_0)\,|D_1|$. Since the matrices D and D_1 are positive definite, their determinants are positive. Therefore we get $\varepsilon = d_{nn} - d'_0 D^{-1}_1 d_0 = |D| / |D_1| > 0$. After some computation, we have

$$D^{-1} = \begin{pmatrix} D^{-1}_1 + (1/\varepsilon)(D^{-1}_1 d_0)(D^{-1}_1 d_0)' & (-1/\varepsilon) D^{-1}_1 d_0 \\ (-1/\varepsilon)(D^{-1}_1 d_0)' & 1/\varepsilon \end{pmatrix}.$$

Since all the elements of D^{-1}_1 are non-negative from the inductive assumption and

$D^{-1}_{1}d_0 \leq 0$, we can easily show that all elements of D^{-1} are non-negative.

Lemma 4.2 Let D be an $n \times n$ positive semidefinite matrix which is partitioned as

$$D = \begin{pmatrix} D_{11}(m, m) & D_{12}(m, n-m) \\ D_{21}(n-m, m) & D_{22}(n-m, n-m) \end{pmatrix}. \quad (4.2)$$

And let $D_{22.1} = D_{22} - D_{21}D^{-1}_{11}D_{12}$. Then $D_{22.1}$ is a positive semidefinite matrix.

Proof. We can easily show that the matrix $P'DP$ is positive semidefinite if and only if D is a positive semidefinite matrix for a non-singular matrix P. Putting

$$P = \begin{pmatrix} I_m & -D^{-1}_{11}D_{12} \\ 0 & I_{n-m} \end{pmatrix},$$

we have

$$P'DP = \begin{pmatrix} D_{11} & 0 \\ 0 & D_{22.1} \end{pmatrix},$$

which is positive semidefinite. Hence $D_{22.1}$ is a positive semidefinite matrix.

For simplicity, denote a positive definite(semidefinite) matrix with all non-positive off-diagonal elements by a p.(s.)d. matrix with all n.p.o.d. elements.

Proposition 4.3 Let D in (4.2) be a p.s.d. matrix with all n.p.o.d. elements. Then the result of sweeping out, taking D_{11} as the pivot matrix,

$$\begin{pmatrix} I_m & D^{-1}_{11}D_{12} \\ 0 & D_{22.1} \end{pmatrix}$$

satisfies the following:

(a) $D^{-1}_{11}D_{12}$ has all non-positive elements.

(b) $D_{22.1}$ is again a p.s.d. matrix with all n.p.o.d. elements.

Proof. Since D_{12} has all non-positive elements and D^{-1}_{11} has all non-negative elements by Lemma 4.1, (a) holds.

Now we prove (b). $D_{22.1}$ is a positive semidefinite matrix by Lemma 4.2. Since D_{22} has all non-positive off-diagonal elements and all the elements of $D_{21}D^{-1}_{11}D_{12}$ are non-negative, $D_{22.1}$ is a matrix with all non-positive off-diagonal elements.

This proposition gives us the following:

(a) Recall the matrix (3.5) made from (3.4). Assume $x_{(1)} \leq 0$ and $x_{(2)} > 0$. We can assume this, because by applying simultaneous permutations of rows and columns, we can make the matrix (3.5) satisfy this assumption. If there are no non-positive elements, then x itself is the solution.

(b) If D in (3.5) is a p.s.d. matrix with all n.p.o.d. elements, then in the resulting matrix (3.6), we have $D^{-1}_{11} x_{(1)} \leq 0$. In other word, the non-positive components remain non-positive after the sweep out operation taking D_{11} as the pivot.

(c) The sweep out operation taking D_{11} as the pivot, of course, may create new non-positive elements in the last column of the matrix (3.6). We can apply the same operation, after making suitable permutations on rows and columns simultaneously. The Proposition 4.3 guarantees that the rows having the non-positive last component remain to have the same sign.

(d) We can repeat the same process until we arrive at the situation where no more rows with the non-positive last component exist except the set of rows whose diagonal elements have been used as the pivotal element.

(e) Apply the permutations on rows and columns simultaneously so that they are in the original order. Replace the non-positive elements by zeros, and this is the solution.

Remark 4.1 Since sweeping out with respect to the zeros in the last column does not change the entities in it, we may sweep out with respect to only the negative elements in applications.

Consequently, the algorithm of Sweep-the-Negatives can be stated in the following sentences:

"If all the elements in the last column of the matrix(D, x) are all non-negative, x itself is the solution. If the last column has one or more negative elements, select any one among them, say the i-th one, and apply the sweep out operation taking the (i, i)-th element as the pivot, and repeat this process until when all of the diagonal elements corresponding to the negative elements in the last column have become 1's. Replace the negative elements in the last column by zeros, and this is the solution".

An important feature of our algorithm is that we can select any starting point and any path to arrive at the unique solution. The PAVA has the same feature.

Now we will illustrate a numerical example.

Example 4.1 Consider the problems (3.2) and (3.3) for given $y=(2, 1, 5, 31)'$, and the restriction matrix $A=(a_1, a_2, a_3, a_4)'$, where $a_1=(1, 0, 0, 0)'$, $a_2=(-1, 1, 0, 0)'$, $a_3=(-2, -3, 1, 0)'$, and $a_4=(-1, -1, -5, 1)'$.

We have $x=Ay=(2, -1, -2, 3)'$ and the matrix $D=(d_1, d_2, d_3, d_4)'$, where $d_1=(1, -1, -2, -1)'$, $d_2=(-1, 2, -1, 0)'$, $d_3=(-2, -1, 14, 0)'$, and $d_4=(-1, 0, 0, 28)'$, which is a p.d. matrix with all n.p.o.d. elements. The matrix corresponding to (3.5) is

$$\begin{pmatrix} D & x \\ A' & y \end{pmatrix} = \begin{pmatrix} 1 & -1 & -2 & -1 & 2 \\ -1 & 2 & -1 & 0 & -1 \\ -2 & -1 & 14 & 0 & -2 \\ -1 & 0 & 0 & 28 & 3 \\ 1 & -1 & -2 & -1 & 2 \\ 0 & 1 & -3 & -1 & 1 \\ 0 & 0 & 1 & -5 & 5 \\ 0 & 0 & 0 & 1 & 31 \end{pmatrix}. \qquad (4.3)$$

Sweeping out the matrix (4.3) two times, taking (2, 2) and (3, 3) elements as the pivots respectively, we have the following matrix corresponding to (3.6)

$$\begin{pmatrix} 1/27 & 0 & 0 & -1 & 28/27 \\ -16/27 & 1 & 0 & 0 & -16/27 \\ -5/27 & 0 & 1 & 0 & -5/27 \\ -1 & 0 & 0 & 28 & 3 \\ 1/27 & 0 & 0 & -1 & 28/27 \\ 1/27 & 0 & 0 & -1 & 28/27 \\ 5/27 & 0 & 0 & -5 & 140/27 \\ 0 & 0 & 0 & 1 & 31 \end{pmatrix}.$$

The solution of (3, 3) and (3, 2) are as follows:

$$\hat{\theta}= \begin{pmatrix} 0 \\ d_2 \end{pmatrix} =(28/27, 0, 0, 3)', \qquad \hat{\mu}=d_3=(28/27, 28/27, 140/27, 31)'.$$

Remark 4.2 When D is known to be non-singular, we need not work on a large matrix in (4.3). Instead we can work on the smaller matrix (D, x) only.

5. Applications to Order Restricted Inference

The algorithm of Sweep-the-Negatives is useful as the condition is fairly mild. Some typical examples are the cases that the parameter restrictions are as follows:

(a) $t\mu_i+(1-t)\mu_{i+1}\leq\mu_{i+2}$ with $(1-t)\omega_i-t\omega_{i+1}\geq 0$, where ω_i is the weight of μ_i, $\omega_i>0$, $0\leq t<1$ and $i\geq 1$.

(b) $\mu_1\leq t\mu_2\leq t^2\mu_3\leq\cdots\leq t^{i-1}\mu_i$ with $t>0$, $\omega_i>0$.

(c) Isotonic regression (the special case when $t=0$ in (a) or $t=1$ in (b)).

(d) Upper starshaped restriction with non-decreasing weight.

In detail, we discuss the cases (c), (d) and the duality.

5-1. Simple Order

The simplest partial order, deeply investigated and frequently applied in practice, is $y_1\leq y_2\leq\cdots\leq y_n$, and it is called "Simple order" in Barlow et al.(1972). In this case the isotonic problem can be reformulated as follows.

Given a set $\{y_1, y_2,\cdots,y_n\}$ and weights$(\omega_1, \omega_2,\cdots,\omega_n)$, find out the minimum of $\Sigma_{i=1}^n\,\omega_i(\mu_i-y_i)^2$ under the conditions $\mu_1\leq\mu_2\leq\cdots\leq\mu_n$. Let $y=(y_1, y_2,\cdots,y_n)'$, $\mu=(\mu_1, \mu_2,\cdots,\mu_n)'$ and $W=\mathrm{diag}(\omega_1, \omega_2,\cdots,\omega_n)$, the diagonal matrix with diagonal elements $\omega_1, \omega_2,\cdots,\omega_n$, then the problem is to find out the minimum of $(\mu-y)'\,W(\mu-y)$ under the same conditions.

Let us consider the transformation $x=Ay$ $(\theta=A\mu)$ where

$$A=\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{pmatrix}. \tag{5.1}$$

The solution of the above isotonic regression is equivalent to

$$\min_{A\mu\geq 0}(\mu-y)'\,W(\mu-y)=\min_{\tilde{A}\tilde{\mu}\geq 0}(\tilde{\mu}-\tilde{y})'(\tilde{\mu}-\tilde{y}),$$

where $\tilde{A}=AW^{-1/2}$, $\tilde{\mu}=W^{1/2}\mu$, and $\tilde{y}=W^{1/2}y$. In this case, we have the matrix corresponding to (3.5) as

$$\begin{pmatrix} D & x \\ \tilde{A}' & \tilde{y} \end{pmatrix}, \tag{5.2}$$

where $x=\tilde{A}\tilde{y}=Ay$ and $D=\tilde{A}\tilde{A}'=AW^{-1}A'$ is equal to

$$\begin{pmatrix} \omega^{-1}_1+\omega^{-1}_2 & -\omega^{-1}_2 & 0 & \cdots & 0 & 0 & 0 \\ -\omega^{-1}_2 & \omega^{-1}_2+\omega^{-1}_3 & -\omega^{-1}_3 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\omega^{-1}_{n-2} & \omega^{-1}_{n-2}+\omega^{-1}_{n-1} & -\omega^{-1}_{n-1} \\ 0 & 0 & 0 & \cdots & 0 & -\omega^{-1}_{n-1} & \omega^{-1}_{n-1}+\omega^{-1}_n \end{pmatrix}.$$

Note that D is a p.d. matrix with all n.p.o.d. elements. Therefore we can solve the isotonic regression in case simple order by applying our algorithm instead of the PAVA. Now we will illustrate a numerical example.

Example 5.1 Let $Y=(Y_1, Y_2, Y_3, Y_4)'$ have a multivariate normal distribution with mean vector $\mu=(\mu_1, \mu_2, \mu_3, \mu_4)'$ and the covariance matrix $W^{-1}=\text{diag}(1, 2, 3, 1)$. In order to obtain the isotonic regression $\hat{\mu}$ of μ, given $y=(2, 3, 0, 3)'$, under the ordered restrictions $\mu_1\leq\mu_2\leq\mu_3\leq\mu_4$, we ought to find

$$\min_{\mu_1\leq\cdots\leq\mu_4} (\mu-y)'\,W(\mu-y) = \min_{\mu_1\leq\cdots\leq\mu_4} \left\{(\mu_1-2)^2+\frac{1}{2}(\mu_1-3)^2+\frac{1}{3}\mu_3^2+(\mu_4-3)^2\right\}. \tag{5.3}$$

In this case, the matrix (5.2) is as follows:

$$\begin{pmatrix} D & x \\ \tilde{A}' & \tilde{y} \end{pmatrix} = \begin{pmatrix} 3 & -2 & 0 & 1 \\ -2 & 5 & -3 & -3 \\ 0 & -3 & 4 & 3 \\ -1 & 0 & 0 & 2 \\ \sqrt{2} & -\sqrt{2} & 0 & 3/\sqrt{2} \\ 0 & \sqrt{3} & -\sqrt{3} & 0 \\ 0 & 0 & 1 & 3 \end{pmatrix} \tag{5.4}$$

Sweeping out(5.4) two times, taking (2, 2) and (1, 1) elements as the pivots, we can show that the lower part of the last column in the resulting matrix is $\tilde{\tilde{\mu}}=(21/11,\ 21\sqrt{2}/22,\ 7\sqrt{3}/11,\ 3)'$. And thus we have the isotonic regression of (5.3), $\hat{\mu}=(21/11, 21/11, 21/11, 3)'$, considering the identity $\hat{\mu}=W^{-1/2}\tilde{\tilde{\mu}}$.

It is clear that the algorithm of Sweep-the-Negatives is quite parallel to the PAVA. Indeed, we can verify the parallelism by the following manner.

Given $(y_1,\cdots,y_n)$ and $(\omega_1,\cdots,\omega_n)$, we apply the PAVA algorithm. Suppose $(y_l,\ y_{l+1})$ is a violating pair. After pooling them, we have

$$(y_1,\cdots,y_{l-1},\ y_{l,l+1},\ y_{l+2},\cdots,y_n)$$

where

$$y_{l,l+1}=\frac{\omega_l y_l+\omega_{l+1}y_{l+1}}{\omega_l+\omega_{l+1}}\ . \tag{5.5}$$

Note that $y_l>y_{l+1}$ implies $x_l<0$ in the algorithm of Sweep-the-Negatives. Sweep out the matrix (5.2), taking the (i, i)-th element as the pivot, then we have the same result as (5.5).

5-2. Upper Starshaped Restriction

Shaked(1979) derived the M.L.E. of normal mean vector subject to the starshaped restriction. Starshaped vectors arise naturally in certain situations where finite populations are amalgamated.

A vector $\mu=(\mu_1,\cdots,\mu_p)'$ is said to be upper starshaped, provided

$$\mu_{k+1}\geq\bar{u}_k\geq 0, \qquad k=0,\ 1,\ 2,\cdots,p-1,$$

where $\bar{\mu}_k(k=1,\ 2,\cdots,p-1)$ is the weighted average of $\mu_1,\ \mu_2,\cdots,\mu_k$ with weights $\omega_1,\ \omega_2,\cdots,\omega_k$ and $\bar{\mu}_0=0$. The upper starshaped ordering might be termed "Increasing on the Average". Note that if $0\leq\mu_k\leq\mu_{k+1}$, $k=1,\ 2,\cdots,p-1$, then μ is upper starshaped.

Let us state a proposition without proof.

Proposition 5.1 Let $W=\mathrm{diag}(\omega_1,\ \omega_2,\cdots,\omega_n)$ and $A=(a_{ij})$ be an $n\times n$ matrix with

$$a_{ij} = \begin{cases} 0 & i<j \\ 1 & i=j \\ -\omega_j/\Sigma_{k=1}^{i-1}\omega_k, \quad 0<\omega_k \leq \omega_{k+1} & i>j. \end{cases} \tag{5.6}$$

Then $AW^{-1}A'$ is a p.d. matrix with all n.p.o.d. elements.

Now consider the case (3.1), and again the minimizing problem (3.2). In addition, let the unknown mean μ be an upper starshaped vector with non-decreasing weights. When we take the transformation $x=Ay$ where $A=(a_{ij})$ satisfies the condition (5.6), Proposition 5.1 shows that the algorithm of Sweep-the-Negatives can be applied to the upper star-shaped restriction with non-decreasing weights. Actually applying our algorithm to the numerical example in Shaked(1979), we can obtain the same result.

6. Duality

Consider again the problem (3.2) for given vector y and matrix A, where AA′ is a p.d. matrix with all n.o.p.d. elements. Now we make use of Sweep-the-Negatives for the solution of (3.2) or (3.3) in case D^{-1} is a p.d. matrix with all n.p.o.d. elements, considering their dualities. The primal problems are equivalent to

$$\min_{\mu \in C} (\mu-y)'(\mu-y), \tag{6.1}$$

where $C=\{\mu \mid A\mu \geq 0\}$, and

$$\min_{\theta \geq 0} (\theta-x)' D^{-1}(\theta-x). \tag{6.2}$$

We consider the case M=N, and denote $a^{(i)}_N$ by $a^{(i)}$ for simplicity. Let A be a non-singular matrix, and let $A^{(n)}=(a^{(1)},\cdots,a^{(n)})'$. From (2.3), we have $A^{(n)}=(AA')^{-1}A=(A')^{-1}$.

Owing to Proposition 2.1, we get

$$C^*=\{\mu^* \mid (a^{(i)}, \mu^*)\leq 0, i\in N\}=\{\mu^* \mid -(A')^{-1} \mu^*\geq 0\}.$$

The dual problem of (6.1) is to find μ^* which attains the optimal solution $\hat{\mu}^*$ such that

$$-\min_{\mu^* \in C^*} (\mu^*-y)'(\mu^*-y)=-(\hat{\mu}^*-y)'(\hat{\mu}^*-y).$$

$\hat{\mu}^*$ is related to the optimal solution $\hat{\mu}$ of (6.1) by the identity $\hat{\mu}=y-\hat{\mu}^*$. In the like manner, we have the dual of (6.2) as

$$-\min_{\theta^*\geq 0} (\theta^*+D^{-1}x)' D(\theta^*+D^{-1}x), \tag{6.3}$$

irrespective of constant term and the optimal solution $\hat{\theta}^*$ is related to the optimal solution $\hat{\theta}$ of (6.2) by the identity $\hat{\theta}=x+D\hat{\theta}^*$.

The above discussions for duality give us an advantageous use of Sweep-the-Negatives in the case that D^{-1} is a p.d. matrix with all n.p.o.d. elements.

Example 6.1 Let $Y=(Y_1, Y_2, Y_3, Y_4)'$ be normally distributed with mean μ and unit variance. Now let $X=AY$, where the rows of A are $a'_1=(1, 0, 0, 0)$, $a'_2=(1/2, 1, 0, 0)$, $a'_3=(1/2, 1/5, 1, 0)$, and $a'_4=(2/5, 4/25, 3/10, 1)$.

Let us consider the M.L.E. of mean θ under the ordered restriction that $\theta=A\mu\geq 0$ for given sample mean $y=(-10, 25, 10, -7)'$ or $x=(-10, 20, 10, -4)'$. We can easily formulate the minimizing problems (6.1) and (6.2) taking A. From $D=AA'$, we have $D^{-1}=(d^1, d^2, d^3, d^4)'$, where $d^1=(29/20, -2/5, -17/50, -1/5)'$, $d^2=(-2/5, 21/20, -17/100, -1/10)'$, $d^3=(-17/50, -17/100, 109/100, -3/10)'$, and $d^4=(-1/5, -1/10, -3/10, 1)'$, and $x^*=-D^{-1}x=(251/10, -237/10, -121/10, 7)'$.

As D^{-1} is a p.d. matrix with all n.p.o.d. elements, we can solve the duality (6.3) rather than (6.2). Sweeping out only the matrix (D^{-1}, x^*) two times, we obtain$(10, -25, -15, 0)'$ as the last column. As the remainders of the last column, not swept out, are all non-negative, the solution of the dual problem is $\hat{\theta}^*=(10, 0, 0, 0)'$. By the identity $\hat{\theta}=x+D\hat{\theta}^*$, we have $\hat{\theta}=(0, 25, 15, 0)'$. Furthermore we are sure that $\hat{\mu}=(0, 25, 10, -7)'$ and $\hat{\mu}^*=(-10, 0, 0, 0)'$ from one of the identities $\hat{\mu}=A^{-1}\hat{\theta}$, $\hat{\mu}^*=-A'\hat{\theta}^*$ and $\hat{\mu}=y-\hat{\mu}^*$.

The following proposition guarantees that a typical example of the duality is the case that the restriction matrix A is given as (6.4).

Proposition 6.1 Let $A=(a_{ij})$ be an $n\times n$ matrix with

$$a_{ij}=\begin{cases} 0 & i<j \\ 1 & i=j \\ a\ (0\leq a\leq 1) & i>j. \end{cases} \tag{6.4}$$

Then $(AA')^{-1}$ is a p.d. matrix with all n.p.o.d. elements.

Another typical example is the case that the covariance matrix D is an equicorrelation matrix with positive correlation coefficients. In fact, if $D=(d_{ij})$ is an $n \times n$ matrix, where $d_{ii}=1$ and $d_{ij}=r$ $(0 \le r < 1)$ $(i \ne j)$, then we obtain $D^{-1}=(d^{ij})$ with

$$d^{ij}= \begin{cases} \{1+(n-2)r\}/[(1-r)\{1+(n-1)r\}] & i=j \\ -r/[(1-r)\{1+(n-1)r\}] & i \ne j, \end{cases}$$

which is a p.d. matrix with all n.p.o.d. elements.

Acknowledgement

The authors would like to give deep appreciation to the referees for their kind arrangement and comments in improving this work.

References

1. Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D.(1972). *Statistical Inference under Order Restrictions, The Theory and Application of Isotonic Regression,* New York.: John Wiley and Sons.
2. Bartholomew, D.J.(1959a). A Test of Homogeneity for Ordered Alternatives, *Biometrika,* Vol. 46, 36-48.
3. Bartholomew, D.J.(1959b). A Test of Homogeneity for Ordered Alternatives, II, *Biometrika,* Vol. 46, 328-335.
4. Choi, J.R.(1975). An Equality Involving Orthant Probabilities, *Communications in Statistics,* Vol. 4, 1167-1175.
5. Choi, J.R.(1976a). The Law of Sweep-the-Negatives in the Estimation under Order Restrictions, *Memoirs of the Faculty of Science, Kyushu University,* A, Vol. 30, 135-143.
6. Choi, J.R.(1976b). Applications of Duality to the Law of Sweep-the-Negatives, *Memoirs of the Faculty of Science, Kyushu University,* A, Vol. 30, 183-189.
7. Dykstra, R.L. and Robertson, T.(1982). Order Restricted Statistical Tests on Multinomial and Poisson Parameters: The Starshaped Restriction, *The Annals of Statistics,* Vol. 10, 1246-1252.
8. Dykstra, R.L., Robertson, T. and Wright, F.T.(eds.)(1986). *Advances in Order Restricted Statistical Inference,* Proceedings, Iowa City, 1985, Springer-Verlag.

9. Kennedy, W.J. and Gentle, J.E.(1980). *Statistical Computing,* New York.: Marcel Dekker, Inc.
10. Kudô, A.(1963). A Multivariate Analogue of the One-sided Test. *Biometrika,* Vol. 50, 403-418.
11. Kudô, A. and Choi, J.R.(1975). A Generalized Multivariate Analogue of the One-sided Test. *Memoirs of the Faculty of Science, Kyushu University,* A, Vol. 29, 303-328.
12. Kudô, A. and Kim, B.D.(1987). On Algorithms of Some Isotonic Regression Problems, Satellite Meeting on Mathematical Statistics and Probability for the 46th Session of ISI, September 3-5, 1987, at Kyoto University, Japan.
13. Robertson, T., Wright, F.T. and Dykstra, R.L.(1988). *Order Restricted Statistical Inference,* New York.: John Wiley.
14. Shaked, M.(1979). Estimation of Starshaped Sequences of Poisson and Normal Means, *The Annals of Statistics,* Vol. 7, 729-741.