Journal of the Korean Statistical Society (2000), **29**: *3, pp 361–373*

An Algorithm for One-Sided Generalized Least Squares Estimation and Its Application [†]

Chul Gyu Park [1]

ABSTRACT

A simple and efficient algorithm is introduced for generalized least squares estimation under nonnegativity constraints in the components of the parameter vector. This algorithm gives the exact solution to the estimation problem within a finite number of pivot operations. Besides an illustrative example, an empirical study is conducted for investigating the performance of the proposed algorithm. This study indicates that most of problems are solved in a few iterations, and the number of iterations required for optimal solution increases linearly to the size of the problem. Finally, we will discuss the applicability of the proposed algorithm extensively to the estimation problem having a more general set of linear inequality constraints.

Keywords: Least squares projection; Mahalanobis distance; Linear constraints; Kuhn-Tucker conditions; Pivot operation, Exact solution.

1. INTRODUCTION

Estimation of unknown parameters has been one of important issues in statistics. Very often, some knowledge or information is incorporated in the estimation problem by imposing appropriate restrictions on those unknown parameters. Many examples of this type of problems are found in Robertson et n.(1988) that provides an overview of past studies particularly related with order restrictions.

Even though some of estimation problems have elegant closed-form solutions as in Barlow and Brunk (1972), Dykstra and Robertson(1982), Dykstra and Lee(1991) and Park(1998), many other problems require expensive numerical work or complicated computation procedures for their solutions. Kudo(1963), Dykstra(1983), Sasabuchi et n.(1983) and El Barmi and Dykstra(1996) suggested iterative algorithms that reduce much of computational work in solving the problems in that category.

[†]This work was supported by Korea Research Foundation Grant (KRF-1999-015-DP0068)
[1]Department of Applied Statistics, The University of Suwon, Suwon, 445-743, Korea

In this paper, we consider a Mahalanobis distance type of objective function for minimizing under the constraint that all the components of the unknown parameter vector are nonnegative. That is, our problem is

$$\mathrm{Min}_{\mathbf{u}\in K} \quad (\mathbf{x}-\mathbf{u})^T\mathbf{W}(\mathbf{x}-\mathbf{u}) \tag{1.1}$$

where $K = \{\mathbf{u} \in R^k : \ u_i \geq 0, \ i = 1, 2, \cdots, k\}$ and $\mathbf{W}$ is a known $k \times k$ positive definite matrix with its elements w_{ij}'s. As discussed in Robertson et n.(1988), a vector $\mathbf{u}^* \in K$ solves the problem (1.1) if

(i) $(\mathbf{x}-\mathbf{u}^*)^T\mathbf{W}\ \mathbf{u}^* = 0,$ and

(ii) $(\mathbf{x}-\mathbf{u}^*)^T\mathbf{W}\ \mathbf{u}\ \leq 0$ for all $\mathbf{u} \in K$.

However, no closed-form solution exists for this problem. Obviously, if $\mathbf{x}$ is in K, $\mathbf{u}^* = \mathbf{x}$. Otherwise the solution $\mathbf{u}^*$ must lie on the boundary of K which is composed of $2^k - 1$ distinct faces. Thus, the solution is the projection of $\mathbf{x}$ onto the linear subspace generated by the face that contains $\mathbf{u}^*$. But, it is very hard to determine which of those faces contains $\mathbf{u}^*$. Kudo(1963) introduces a heuristic method to find such subspace by observing the derivative of the objective function. However, this method is not an algorithm in a strict sense because it appeals to intuitive judge in the selection of subspaces upon which to project. In a different perspective, Dykstra(1983) and Wollan and Dykstra(1985) suggest infinite iterative algorithms which are guaranteed to converge correctly. Even though they converge very fast for some problems, they can not give the exact optimal solution within finite steps.

We propose here a simple and efficient algorithm which is designed to give the exact solution within a finite number of pivot operations. As we see in Section 4, this approach can also be applied to a wide set of least squares problems having general linear inequality constraints.

2. CHARACTERIZATION OF THE OPTIMAL SOLUTION

We will derive the optimality conditions for the problem (1.1) which provide a theoretical basis of our algorithm. First, note that the problem is equivalent to minimizing

$$\frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}(x_i - u_i)(x_j - u_j)w_{ij}$$

subject to $u_i \geq 0,\ i = 1, 2, \cdots, k$. For future notational convenience, let $\mathbf{u} = (u_1, u_2, \cdots, u_k, u_{k+1}, \cdots, u_{2\times k})^T$ denote the parameter vector augmented by Lagrangian multipliers, $u_{k+1}, u_{k+2}, \cdots, u_{2\times k}$. Then, the objective function involving those Lagrangian multipliers becomes $Q(\mathbf{u}) = \frac{1}{2}\sum_{i=1}^{k}\sum_{j=1}^{k}(x_i - u_i)(x_j - u_j)w_{ij} - \sum_{l=1}^{k} u_l u_{k+l}$. By setting $\frac{\partial Q}{\partial u_i} = 0,\ i = 1, 2, \cdots, k$, we obtain the Kuhn-Tucker optimality conditions(pp 314 - 315 of Luenberger(1984)) for the minimization problem as follows:

$$-\sum_{j=1}^{k} w_{ij}u_i + u_{k+i} = -\sum_{j=1}^{k} w_{ij}x_j,\ i = 1, 2, \cdots, k, \tag{2.1}$$

and

$$u_i u_{k+i} = 0,\ u_i \geq 0,\ \text{and } u_{k+i} \geq 0,\ i = 1, 2, \cdots, k. \tag{2.2}$$

Let $b_i = -\sum_{j=1}^{k} w_{ij}x_j, i = 1, 2, \cdots, k$, and consider a $k \times 2k$ matrix $\mathbf{C}$ whose elements are given by

$$c_{ij} = \begin{cases} -w_{ij} & \text{if } j = 1, 2, \cdots, k,\ i = 1, 2, \cdots, k \\ 1 & \text{if } j = k + i,\ i = 1, 2, \cdots, k \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

Then, the equations in (2.1) can be restated as

$$\mathbf{Cu} = \mathbf{b} \tag{2.4}$$

where $\mathbf{b} = (b_1, b_2, \cdots, b_k)^T$.

Let $\mathbf{c}_j$ denote the jth column vector of $\mathbf{C}$ such that $\mathbf{c}_j = (c_{1j}, c_{2j}, \cdots, c_{kj})^T$, $j = 1, 2, \cdots, 2k$. Consider a set of indices defined as $\mathbf{L} = \{1, 2, \cdots, 2k\}$, and partition $\mathbf{L}$ into two sets, say $\mathbf{B}$ and $\mathbf{N}$, such that $\mathbf{B} = \{i_1, i_2, \cdots, i_k;\ i_l \text{ is either } l \text{ or } k + l \text{ for } l = 1, 2, \cdots, k\}$ and $\mathbf{N} = \mathbf{L} \setminus \mathbf{B}$. Then, there are 2^k possible ways of forming the index set $\mathbf{B}$. Let $\mathbf{C_B}$ be the $k \times k$ matrix whose column vectors are $\mathbf{c}_j,\ j \in \mathbf{B}$. Let $\mathbf{u_B}$ be the k-dimensional vector with its components, $u_j,\ j \in \mathbf{B}$. We define $\mathbf{C_N}$ and $\mathbf{u_N}$ similarly. Based upon these notations, we can rewrite the equation (2.4) as

$$\mathbf{u_B} + \mathbf{C_B}^{-1}\mathbf{C_N}\mathbf{u_N} = \mathbf{C_B}^{-1}\mathbf{b}. \tag{2.5}$$

Here, we call $\mathbf{u_B}$ and $\mathbf{u_N}$ the vectors of basic and nonbasic variables, respectively. The corresponding matrix $\mathbf{C_B}$ is called the basis matrix for the linear equation (2.5).

If $\mathbf{C_B}^{-1}\mathbf{b} \geq \mathbf{0}$ where the inequality holds componentwise, the optimal solution satisfying Kuhn-Tucker conditions (2.1) and (2.2) will be

$$\mathbf{u_B} = \mathbf{C_B}^{-1}\mathbf{b} \text{ and } \mathbf{u_N} = 0. \tag{2.6}$$

The unique existence of such an optimal index set $\mathbf{B}$ is obvious from the fact that our problem has a unique solution. For the exact optimal solution, we may rely on branching technique which considers all possible sets of $\mathbf{B}$. However, this method is awfully inefficient because there are 2^k possible cases of $\mathbf{B}$ each of which requires to check the optimality conditions by computing $\mathbf{C_B}^{-1}\mathbf{b}$. In the following section we will propose an efficient algorithm for finding the optimal solution.

3. THE PROPOSED ALGORITHM

3.1. The Algorithm

The new algorithm we propose here is primarily composed of pivot operations that are used in linear programming. Readers not familiar with this topic may refer to Bazara and Jarvis(1977) for details. However, due to the first part of constraints in (2.2), the choice of basic variable at each iteration of the algorithm must be made between u_i and u_{k+i} for all i. In addition, the criteria for entering and leaving basic variables are quite different from those of usual linear programming. As will be investigated later, this algorithm finds the exact solution to the problem (1.1) within a finite number of iterations. Following is the detailed procedure of the proposed algorithm.

Step 0 : Construct a $k \times 2k$ matrix $\mathbf{C}$ as given in (2.3) and compute $b_i = -\sum_{j=1}^{k} w_{ij}x_j$, $i = 1, 2, \cdots, k$. Also, let $i_l = k + l$, $l = 1, 2, \cdots, k$ and form an index set $\mathbf{B} = \{i_1, i_2, \cdots, i_k\}$ to specify the initial basic variables.

Step 1 : Find r such that $b_r = \min_{1 \leq i \leq k} b_i$. If there are more than one such r, choose it arbitrarily. If $b_r \geq 0$, stop the algorithm. Otherwise go to Step 2.

Step 2 : Determine the entering and leaving basic variables. That is, if $r \notin \mathbf{B}$(or $r \in \mathbf{B}$), replace $k+r$(or r) in $\mathbf{B}$ by r(or $k+r$) by setting $i_r = r$(or $k+r$).

Step 3 : Update $\mathbf{C}$ and $\mathbf{b}$ by computing the new elements as

$$c_{ij}^{new} = \begin{cases} c_{ij}/c_{ii_r} & \text{if } i = r \\ c_{ij} - c_{rj}c_{ii_r}/c_{ri_r} & \text{if } i \neq r \end{cases} \tag{3.1}$$

and

$$b_i^{new} = \begin{cases} b_i/c_{ri_r} & \text{if } i = r \\ b_i - b_r c_{ii_r}/c_{ri_r} & \text{if } i \neq r. \end{cases} \tag{3.2}$$

Step 4 : Repeat Steps $1-3$ until the termination rule in Step 1 is met.

Note that we conduct in Step 3 a pivot operation with pivot element c_{ri_r}. The choice of r with minimum-valued $b_r(<0)$ in Step 1 is to ensure rapid arrival at the optimal solution. The following theorem specifies the optimal solution obtained by our algorithm.

Theorem 3.1. *Let* $\mathbf{b}^*$ *and* $\mathbf{B}^*$ *denote the updated ones of* $\mathbf{b}$ *and* $\mathbf{B}$ *respectively in the final iteration. Then, the optimal solution,* $\mathbf{u}^*$*, to the problem (1.1) is given by*

$$u_i^* = \begin{cases} b_i^* & \text{if } i \in \mathbf{B}^* \\ 0 & \text{if } i \notin \mathbf{B}^*. \end{cases} \tag{3.3}$$

Proof: Let ${\mathbf{C}^*}_{bfB}$ be the $k \times k$ matrix composed of column vectors of $\mathbf{C}$ corresponding to the basic variables specified in $\mathbf{B}^*$. Then, the pivot operations up to the final iteration leads the equation (2.4) to

$$\mathbf{u}_{\mathbf{B}^*} + \mathbf{C}_{\mathbf{B}^*}^{-1}\mathbf{C}_{\mathbf{N}^*}\mathbf{u}_{\mathbf{N}^*} = \mathbf{C}_{\mathbf{B}^*}^{-1}\mathbf{b}$$

where $\mathbf{N}^* = \mathbf{L} \setminus \mathbf{B}^*$. Since $\mathbf{b}^* = \mathbf{C}_{\mathbf{B}^*}^{-1}\mathbf{b}$, the solution in (3.3) is equivalent to ${\mathbf{u}_{\mathbf{B}^*}}^* = \mathbf{C}_{\mathbf{B}^*}^{-1}\mathbf{b}$ and ${\mathbf{u}_{\mathbf{N}^*}}^* = \mathbf{0}$, and thus satisfies (2.1). Moreover, those basic variables have been chosen such that only one of i and $k+i$ is contained in $\mathbf{B}^*$ for all $i = 1, 2, \cdots, k$. This fact implies that conditions in (2.2) are satisfied. Therefore, the solution in (3.3) satisfies all the optimality conditions. □

For illustrating how to apply our algorithm, we take an example that was considered in Kudo(1963). In this example, the negative log-likelihood function with a multivariate normal observation $\mathbf{x}^T = (-10.0, -1.0, 10.0, 0.3)$ is minimized under the constraint that all the components of mean vector are nonnegative. In table 1, the covariance matrix and its inverse(denoted by Σ and W respectively) are listed. In Table 2, the coefficient matrix and the right hand side of equation (2.4) are computed first. We also set $\mathbf{B} = \{5, 6, 7, 8\}$ to specify the initial basic variables. In Step 1, we get $r = 3$ with $b_r = -11.8880$. In the next step, we replace $k+r = 4+3 = 7$ in $\mathbf{B}$ by $r = 3$ because $3 \notin \mathbf{B}$. Thus, the new index set is $\mathbf{B} = \{5, 6, 3, 8\}$. Now, we update $\mathbf{C}$ and $\mathbf{b}$ by pivot operation with pivot element c_{33} and get the new tableau. Following the same procedures, we get the final

Table 1. Covariance Matrix and its Inverse

Covariance($\mathbf{\Sigma}$)				Inverse($\mathbf{W} = \mathbf{\Sigma}^{-1}$)			
1.0000	.2000	.2000	-.1000	1.0656	-.1665	-.1596	.0040
.2000	1.0400	.2400	-.4200	-.1665	1.1844	-.1620	.3800
.2000	.2400	1.0800	-.2000	-.1596	-.1620	1.0100	.1000
-.1000	-.4200	-.2000	1.1800	.0040	.3800	.1000	1.0000

tableau in which all the right hand sides are nonnegative and the corresponding index set is $\mathbf{B} = \{5, 2, 3, 8\}$. Since $1 \notin \mathbf{B}$, $2 \in \mathbf{B}$, $3 \in \mathbf{B}$, and $4 \notin \mathbf{B}$, the solution is given by $u_1^* = 0$, $u_2^* = 0.7607$, $u_3^* = 11.8923$, and $u_4^* = 0$.

3.2. Performance of The Algorithm

We will investigate the number of iterations required for obtaining the optimal solution. First recall that there are 2^k possible ways of forming the index set, $\mathbf{B}$, that specifies the basic variables. It is guaranteed from the uniqueness of the solution that one of those sets is optimal. Thus, if a new index set $\mathbf{B}$ is produced at each iteration, our algorithm will eventually meet the optimal one. The following proposition tells us an aspect of our algorithm.

Proposition: Let $\mathbf{C}^\circ$ and $\mathbf{B}^\circ$ denote the updated ones of $\mathbf{C}$ and $\mathbf{B}$ at any iteration. Then, we have for any $i = 1, 2, \cdots, k$

$$c_{ii}^\circ = 1 \text{ and } c_{i,k+i}^\circ < 0 \quad \text{if } \ i \in \mathbf{B}^\circ$$

and

$$c_{ii}^\circ < 0 \text{ and } c_{i,k+i}^\circ = 1 \quad \text{if } \ i \notin \mathbf{B}^\circ.$$

Proof: Assume without loss of generality that $\mathbf{B}^\circ = \{1, 2, \cdots, r, k+r+1, k+r+2, \cdots, 2k\}$. Partition the $k \times 2k$ matrix $\mathbf{C}$ such that

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{I}_1 & \mathbf{0} \\ \mathbf{C}_{12}^T & \mathbf{C}_{22} & \mathbf{0} & \mathbf{I}_2 \end{bmatrix}$$

where $\mathbf{C}_{11}$, $\mathbf{C}_{12}$ and $\mathbf{C}_{22}$ are $r \times r$, $r \times (k-r)$ and $(k-r) \times (k-r)$ matrices and $\mathbf{I}_1$ and $\mathbf{I}_2$ are $r \times r$ and $(k-r) \times (k-r)$ identity matrices, respectively. Then, the matrix $\mathbf{C}^\circ$ updated according to the pivot operations until the current iteration

Table 2. Pivot Operations Applied to Kudo's Example

$\mathbf{c}_1$	$\mathbf{c}_2$	$\mathbf{c}_3$	$\mathbf{c}_4$	$\mathbf{c}_5$	$\mathbf{c}_6$	$\mathbf{c}_7$	$\mathbf{c}_8$	b
				$\mathbf{B} = \{5, 6, 7, 8\}$				
-1.0656	.1665	.1596	-.0040	1.0000	.0000	.0000	.0000	12.0845
.1665	-1.1844	.1620	-.3800	.0000	1.0000	.0000	.0000	1.0256
.1596	.1620	-1.0100	-.1000	.0000	.0000	1.0000	.0000	-11.8880
-.0040	-.3800	-.1000	-1.0000	.0000	.0000	.0000	1.0000	-.8800
				$\mathbf{B} = \{5, 6, 3, 8\}$				
-1.0404	.1921	.0000	-.0198	1.0000	.0000	.1580	.0000	10.2059
.1921	-1.1584	.0000	-.3960	.0000	1.0000	.1604	.0000	-.8812
-.1580	-.1604	1.0000	.0990	.0000	.0000	-.9901	.0000	11.7703
-.0198	-.3960	.0000	-.9901	.0000	.0000	-.0990	1.0000	.2970
				$\mathbf{B} = \{5, 2, 3, 8\}$				
-1.0085	.0000	.0000	-.0855	1.0000	.1658	.1846	.0000	10.0598
-.1658	1.0000	.0000	.3419	.0000	-.8632	-.1385	.0000	.7607
-.1846	.0000	1.0000	.1538	.0000	-.1385	-1.0123	.0000	11.8923
-.0855	.0000	.0000	-.8547	.0000	-.3419	-.1538	1.0000	.5983

can be expressed as

$$\mathbf{C}^\circ = \begin{bmatrix} \mathbf{I}_1 & \mathbf{C}_{11}{}^{-1}\mathbf{C}_{12} & \mathbf{C}_{11}{}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{22} - \mathbf{C}_{12}^T\mathbf{C}_{11}{}^{-1}\mathbf{C}_{12} & -\mathbf{C}_{12}^T\mathbf{C}_{11}{}^{-1} & \mathbf{I}_2 \end{bmatrix}.$$

Thus, it is obvious that $c_{ii}^\circ = 1$ if $i \in \mathbf{B}^\circ$ and $c_{i,k+i}^\circ = 1$ if $i \notin \mathbf{B}^\circ$. Now, it suffices to prove that all the diagonal elements of $\mathbf{C}_{11}{}^{-1}$ and $\mathbf{C}_{22} - \mathbf{C}_{12}^T\mathbf{C}_{11}{}^{-1}\mathbf{C}_{12}$ are negative-valued. Recall from (2.3) that

$$\mathbf{W} = -\begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^T & \mathbf{C}_{22} \end{bmatrix}.$$

Since $\mathbf{W}$ is positive definite, $-\mathbf{C}_{11}$ is also positive definite. Moreover, the inverse of $\mathbf{W}$ is given by

$$\mathbf{W}^{-1} = \begin{bmatrix} -(\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}{}^{-1}\mathbf{C}_{12}{}^T)^{-1} & \mathbf{C}_{11}{}^{-1}\mathbf{C}_{12}(\mathbf{C}_{22} - \mathbf{C}_{12}^T\mathbf{C}_{11}{}^{-1}\mathbf{C}_{12})^{-1} \\ \mathbf{C}_{22}{}^{-1}\mathbf{C}_{12}^T(\mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{12}^T)^{-1} & -(\mathbf{C}_{22} - \mathbf{C}_{12}^T\mathbf{C}_{11}{}^{-1}\mathbf{C}_{12})^{-1} \end{bmatrix}.$$

Since $\mathbf{W}^{-1}$ is positive definite, so is $-[\mathbf{C}_{22} - \mathbf{C}_{12}^T\mathbf{C}_{11}{}^{-1}\mathbf{C}_{12}]^{-1}$. This implies that $-[\mathbf{C}_{22} - \mathbf{C}_{12}^T\mathbf{C}_{11}{}^{-1}\mathbf{C}_{12}]$ is positive definite. Thus, all the diagonal elements of

$\mathbf{C}_{11}{}^{-1}$ and $\mathbf{C}_{22} - \mathbf{C}_{12}^T \mathbf{C}_{11}{}^{-1} \mathbf{C}_{12}$ have negative values. This completes the proof. □

There is an important fact that should be pointed out regarding this proposition. Since the pivot element (c_{rr} or $c_{r,k+r}$) is always negative, the right hand side(b_r) of the equation corresponding to the pivot element becomes positive after the pivot operation in Step 3. Hence, in the next iteration, the entering basic variable is determined between u_i and u_{k+i} for some i other than r unless the termination rule has been met. Thus, the new index set $\mathbf{B}$ is different from the previous one, which implies that our algorithm does not alternate between c_{rr} and $c_{r,k+r}$ in choosing pivot element. Although it can not be shown rigorously, our experience reveals that the proposed algorithm does not produce the same index set that has ever appeared. Since there are at most 2^k possible sets for $\mathbf{B}$, we will eventually encounter the optimal index set within a finite number of pivot operations.

Since there are 2^k possible ways of forming the basis matrix, the number of iterations required for the optimal solution might be expected to increase exponentially as we increase the size(k) of the problem. But, a significant number of iterations are reduced by selecting pivot row r such that $b_r = \min_{1 \le i \le k} b_i$ in Step 1. This fact can be observed from the distribution of the number of iterations in Table 3 constructed by solving 10,000 randomly generated problems for each of $k = 3, 4, \cdots, 10$. Positive definite matrix, $\mathbf{W}$, involved in the minimization problem is constructed by generating a $k \times k$ matrix $\mathbf{A}$ of full rank and then setting $\mathbf{W} = \mathbf{A}\mathbf{A}^T$. Each component of $\mathbf{x}$ is randomly generated from an interval [-10, 10]. As we see in Table 3, only a few iterations are needed for obtaining the exact solution for our minimization problem. For smaller $k(\le 5)$, about eighty five percents of the problems are solved within k iterations. More desirably for larger k, the the maximum number of iterations increases linearly(rather than exponentially) to the size of the problem. This computational efficiency is the most appealing aspect of our algorithm among others such as algorithmic simplicity and easiness in its implementation.

Table 3. Distribution of the Number of Iterations Needed for Optimal Solution (Based on 10,000 Randomly Generated Problems for Each Case)

	Number of iterations									
k	1	2	3	4	5	6	7	8	9	10
3	.1233	.3619	.3545	.1449	.0148	.0005	.0001	-	-	-
4	.0641	.2460	.3499	.2408	.0824	.0150	.0015	.0003	-	-
5	.0337	.1540	.2979	.2845	.1607	.0546	.0124	.0018	.0004	-
6	.0152	.0945	.2241	.2846	.2168	.1149	.0378	.0090	.0026	.0005
7	.0081	.0526	.1638	.2529	.2532	.1628	.0679	.0264	.0093	.0024
8	.0036	.0319	.1061	.2118	.2502	.2009	.1145	.0494	.0209	.0070
9	.0018	.0153	.0666	.1621	.2234	.2319	.1606	.0808	.0347	.0157
10	.0008	.0092	.0412	.1148	.2010	.2310	.1903	.1072	.0602	.0252
11	.0008	.0048	.0258	.0824	.1530	.2073	.2060	.1600	.0872	.0399
12	.0002	.0034	.0164	.0529	.1209	.1884	.2041	.1833	.1201	.0594
13	-	.0018	.0100	.0376	.0845	.1461	.2002	.1925	.1534	.0862
14	-	.0009	.0043	.0242	.0612	.1161	.1713	.1961	.1687	.1197
15	-	.0005	.0031	.0142	.0402	.0930	.1435	.1864	.1821	.1444

	Number of iterations									
k	11	12	13	14	15	16	17	18	19	20
3	-	-	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-	-	-
5	-	-	-	-	-	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	-
7	.0006	-	-	-	-	-	-	-	-	-
8	.0029	.0006	.0002	-	-	-	-	-	-	-
9	.0048	.0018	.0004	.0001	-	-	-	-	-	-
10	.0126	.0041	.0017	.0007	-	-	-	-	-	-
11	.0192	.0076	.0036	.0020	.0003	.0001	-	-	-	-
12	.0269	.0144	.0057	.0020	.0010	.0006	.0002	.0001	-	-
13	.0460	.0236	.0102	.0045	.0023	.0008	.0002	.0001	-	-
14	.0728	.0345	.0146	.0080	.0044	.0014	.0012	.0004	.0001	.0001
15	.0925	.0526	.0257	.0108	.0058	.0032	.0012	.0005	.0002	.0001

4. APPLICATIONS

4.1. Simple Order Restriction

A problem that has received considerable attention is that of finding the solution of $\mathbf{u}$ that minimizes

$$Q(\mathbf{u}) = \frac{1}{2}(\mathbf{x}-\mathbf{u})^T\mathbf{W}(\mathbf{x}-\mathbf{u})$$

under the constraints

$$u_1 \le u_2 \le \cdots \le u_k.$$

As is well known, no closed form exists for the solution to this problem except for the case with very particular structure of $\mathbf{W}$. Such examples are discussed in Brunk(1955), Bartholomew(1959, 1961), and are comprehensively reviewed in Robertson et n.(1987).

However, we can find the exact solution to the problem even in the case of arbitrary positive definite matrix $\mathbf{W}$ by applying our algorithm. To see how to apply our algorithm, consider the Kuhn-Tucker optimality conditions for the minimization problem which are derived with some modifications as follows:

$$-\sum_{j=1}^{k}[(x_j-u_j)\sum_{l=1}^{i} w_{lj}] + v_{(k-1)+i} = 0 \quad i=1,2,\cdots,k-1, \tag{4.1}$$

$$-\sum_{j=1}^{k}[(x_j-u_j)\sum_{l=1}^{k} w_{lj}] = 0, \tag{4.2}$$

$$v_{(k-1)+i}(u_i-u_{i+1}) = 0,\ u_i-u_{i+1} \le 0,\ \text{and}\ v_{(k-1)+i} \ge 0,\ \ i=1,2,\cdots,k-1 \tag{4.3}$$

where $v_{(k-1)+i}$, $i=1,2,\cdots,k-1$ are Lagrangian multipliers.

Now, we reparameterize by setting $v_0 = u_1$ and $v_i = u_{i+1}-u_i$, $i=1,2,\cdots,k-1$. Also, let $\alpha_{ij} = \sum_{l=1}^{i} w_{lj}$ and $\beta_{ij} = \sum_{l=j+1}^{k}\alpha_{il}$. Then, equations (4.1) and (4.2) are expressed as

$$\sum_{j=0}^{k-1}\beta_{ij}v_j + v_{(k-1)+i} = \sum_{j=1}^{k}\alpha_{ij}x_j, \quad i=1,2,\cdots,k-1, \tag{4.4}$$

$$\sum_{j=0}^{k-1}\beta_{kj}v_j = \sum_{j=1}^{k}\alpha_{kj}x_j. \tag{4.5}$$

From (4.5), we get

$$v_0 = \frac{1}{\beta_{k0}}[\sum_{j=1}^{k} \alpha_{kj} x_j - \sum_{j=1}^{k-1} \beta_{kj} v_j]. \tag{4.6}$$

Putting (4.6) into (4.4), the conditions (4.1) - (4.3) rewritten as

$$\sum_{j=1}^{k-1} (\beta_{ij} - \frac{\beta_{i0}\beta_{kj}}{\beta_{k0}}) v_j + v_{(k-1)+i} = \sum_{j=1}^{k} (\alpha_{ij} - \frac{\beta_{i0}\alpha_{kj}}{\beta_{k0}}) x_j, \quad i = 1, 2, \cdots, k-1, \tag{4.7}$$

$$v_i v_{(k-1)+i} = 0, \; v_i \geq 0, \text{ and } v_{(k-1)+i} \geq 0, \quad i = 1, 2, \cdots, k-1. \tag{4.8}$$

Since the expressions (4.7) and (4.8) are of the same fashions as (2.1) and (2.2), respectively, our algorithm is now applicable with appropriate construction of **C** and **b** at the initialization step.

4.2. Linear Inequality Constraints

Quadratic problems having a more general set of linear constraints can be solved by applying our algorithm. Suppose we want to minimize $Q(\mathbf{u})$ subject to

$$\sum_{j=1}^{k} a_{ij} v_j \leq 0 \quad \text{for } i = 1, 2, \cdots, r.$$

We may apply the infinite iterative algorithm suggested in Dykstra(1983). But, we have more interests in obtaining the exact solution within finite steps. Let **A** be the $r \times k$ matrix whose elements are a_{ij}, $i = 1, 2, \cdots, r$, $j = 1, 2, \cdots, k$. Assume that **A** has a full row rank. As discussed in Section 8.6 of Luenberger(1969), the dual problem is to minimize

$$S(\mathbf{v}) = \frac{1}{2}\mathbf{v}^T \mathbf{P} \mathbf{v} - \mathbf{x}^T \mathbf{A}^T \mathbf{v}$$

subject to $v_i \geq 0$, $i = 1, 2, \cdots, r$ where $\mathbf{P} = \mathbf{A}\mathbf{W}^{-1}\mathbf{A}^T$.

Let $\mathbf{y} = \mathbf{P}^{-1}\mathbf{A}\mathbf{x}$. Then, the dual problem is equivalent to minimizing

$$R(\mathbf{v}) = \frac{1}{2}(\mathbf{y} - \mathbf{v})^T \mathbf{P} (\mathbf{y} - \mathbf{v})$$

under the constraints $v_i \geq 0$, $i = 1, 2, \cdots, r$. Since this problem is of the canonical form as given in (1.1), the dual solution, say $\mathbf{v}^*$, can be obtained by applying our algorithm. From the duality arguments, the optimal solution to the original (primal) problem is given by $\mathbf{u}^* = \mathbf{x} - \mathbf{W}^{-1}\mathbf{A}^T\mathbf{v}^*$.

References

Barlow, R. E., and Brunk, H. D. (1972), "The isotonic regression problem and its dual," *Journal of the American Statistical Association*, 67, 140-147.

Bartholomew, D. J. (1959), "A test of homogeneity for ordered alternatives," *Biometrika*, 46, 36-48.

Bartholomew, D. J. (1961), "A test of homogeneity of means under restricted alternatives" (with discussions), *Journal of the Royal Statistical Society, Series B*, 23, 238-281.

Bazara, M. S. and Jarvis, J. J. (1977), *Linear Programming and Network Flows*, Wiley: New York.

Brunk, H. D. (1955), "Maximum likelihood estimates of monotone parameters," *Annals of Mathematical Statistics*, 26, 607-616.

Dykstra R. L. (1983), "An algorithm for restricted least squares regression," *Journal of the American Statistical Association*, 78, 837-842.

Dykstra R. L., and Lee, C. I. C. (1991), "Multinomial estimation procedures for isotonic cones," *Statistics and Probability Letters*, 11, 155-160.

Dykstra, R. L., and Robertson, T. (1982), "Order Restricted Statistical Tests on Multinomial and Poisson Parameters: The Starshaped Restriction," *Annals of Statistics*, 10, 1246-1252.

El Barmi, H., and Dykstra, R. L. (1996), "Restricted product multinomial and product Poisson maximum likelihood estimation based upon Fenchel duality," *Statistics and Probability Letters*, 29, 117-123.

Kudo, A. (1963), "A multivariate analogue of the one-sided test," *Biometrika*, 50, 403-418.

Luenberger, D. G. (1969), *Optimization by Vector Space Methods*, Wiley, New York.

Luenberger, D. G. (1984), *Linear and Nonlinear Programming*, Addison Wesley, Reading.

Park, C. G. (1998), "Least squares estimation of two functions under order restriction in isotonicity," *Statistics and Probability Letters*, 37, 97-100.

Robertson, T., Wright, F. T. & Dykstra, R. L. (1988), *Order Restricted Statistical Inference.* Wiley, New York.

Sasabuchi, S., Inutsuka, M, and Kulatunga, D. D. S. (1983), "A multivariate version of isotonic regression," *Biometrika*, 70, 465-472.

Wollan, P. C., and Dykstra, R. L. (1986), "Minimizing linear inequality constrained Mahalanobis distances," *Applied Statistics*, 36, 234-240.