Journal of the Korean
Statistical Society
Vol. 22, No. 2, 1993

An Efficient Computing Method of the Orthogonal Projection Matrix for the Balanced Factorial Design

Byung-Chun Kim[1] and Jong-Tae Park[2]

ABSTRACT

It is well known that design matrix X for any factorial design can be represented by a product $X = TX_o$ where T is replication matrix and X_o is the corresponding balanced design matrix. Since X_o consists of regular arrangement of 0's and 1's, we can easily find the spectral decomposition of $X_o'X_o$. Also using this we propose an efficient algorithm for computing the orthogonal projection matrix for a balanced factorial design.

KEYWORDS: Orthogonal projection matrix, Moore-Penrose inverse, spectral decomposition.

1. INTRODUCTION

When a linear model $\mathbf{y} = X\beta + \mathbf{e}$ is given, the total sum of squares $\mathbf{y}'\mathbf{y}$ can be decomposed into

$$\mathbf{y}'\mathbf{y} = \mathbf{y}'P_X\mathbf{y} + \mathbf{y}'(I - P_X)\mathbf{y} \tag{1.1}$$

where $P_X = X(X'X)^-X'$ and $(X'X)^-$ is a generalized inverse of $X'X$. $\mathbf{y}'P_X\mathbf{y}$ denotes the model sum of squares and $\mathbf{y}'(I - P_X)\mathbf{y}$ denotes the error sum of squares.

[1] Department of Mathematics, KAIST, Taejon, 305-701, Korea
[2] Department of Mathematics, KAIST, Taejon, 305-701, Korea

The other sums of squares can be expressed as the quadratic form of $\mathbf{y}$. The matrix $X(X'X)^{-}X'$, denoted by P_X, is called the orthogonal projection matrix and plays an important role in analyzing the linear model. Also P_X can be expressed as $P_X = XX^+$ where X^+ is the Moore-Penrose inverse of a matrix X because P_X is unique for any choice of generalized inverse of $X'X$. In this point of view, it is important to find the Moore-Penrose inverse of design matrix. In the statistical areas of analysis of variance and regression, the characteristic of this matrix is the basis for much of the modern development. Many statisticians [e.g., Kempthrone(1980), Kennedy and Gentle(1980), and Lowerre(1982)] contributed to deriving the Moore-Penrose inverse X^+ for the special statistical models. More recently, Kim and Lee(1986) have found the explicit form of the Moore-Penrose inverse of the design matrix for one- and two-way classification models. Kim and Sunwoo(1989) found the explicit form of the Moore-Penrose inverse X^+ of the design matrix of the model $\mathbf{y} = X\beta + \mathbf{e}$ for the balanced model with no interactions using the relationship between the Moore-Penrose inverse and the minimum norm least squares solution. Also Kim and Sunwoo(1990) derived the iterative method for computing the Moore-Penrose inverse of design matrix for a balanced factorial design with interactions and from this the orthogonal projection matrix is computed. But X^+ is not easy to compute, especially when X has many columns, and by this reason, the type of X^+ used in ANOVA is not known up to now. To avoid this difficulty, we shall find $(X'X)^+$ instead of X^+ in obtaining P_X whether there are interactions or not, because $X'X$ is symmetric positive semi-definite and thus it is easy to compute $(X'X)^+$ from the spectral decomposition of $X_o'X_o$ for the corresponding balanced model.

In this paper we will suggest an efficient procedure for computing P_X for a balanced model using the nonzero eigenvalues and eigenvectors of $X_o'X_o$. In section 2 we describe the explicit form of the nonzero eigenvalues of $X_o'X_o$ and some matrix that is decisive in obtaining P_X in cases with and without interaction effects, and using this derive the explicit form of the projection matrix for a balanced model. In addition, some examples will be given. In section 3, as application, a simple method for computing F-statistic for testing a general linear hypothesis is proposed.

2. THE ORTHOGONAL PROJECTION MATRIX FOR THE BALANCED MODEL

In this article we consider the model

$$\mathbf{y} = X\beta + \mathbf{e} \tag{2.1}$$

where

$\mathbf{y}$: an $N \times 1$ vector of observations

β: a $p \times 1$ vector of unknown parameters
X: an $N \times p$ design matrix consisting of 0's and 1's
N: the total number of observations ($= nm$)
n: the number of replications in each cell
m: the total number of cells.

For model (2.1) the corresponding balanced model is defined as the model that contains exactly one observation in each cell and design matrix of the model is denoted by X_o. It is well known that $X = TX_o$ holds, where T is replication matrix of the form $T = I_m \otimes \mathbf{1}_n$ using Kronecker product. It is clear that $T'T = nI_m$. Hence it suffices to compute P_{X_o} of X_o because of $P_X = X(X'X)^+X' = TX_o(nX_o'X_o)^+X_o'T' = (1/n)TP_{X_o}T'$.

Put $r = \text{rank}(X_o)$. Let $X_o'X_o = Q_r\Lambda_rQ_r'$ be a spectral decomposition of $X_o'X_o$ where Λ_r is a diagonal matrix of positive eigenvalues of $X_o'X_o$ and Q_r is a $p \times r$ matrix whose columns are orthonormalized eigenvectors corresponding to positive eigenvalues. From this $(X_o'X_o)^+ = Q_r\Lambda_r^{-1}Q_r'$ holds. Now let $Z = X_oQ_r\Lambda_r^{-1/2}$. Then

$$P_{X_o} = ZZ' \tag{2.2}$$

is an obvious result once a spectral decomposition is applied to $X_o'X_o$.

From (2.2) the orthogonal projection matrix of a balanced model can be easily computed using Z. Kim and Park(1992) derived the explicit form of Λ_r and Z for multi-way factorial design in two cases: with and without interaction effects. The nonzero eigenvalues λ_i's of $X_o'X_o$ and the form of Z for k-way factorial design depend on only the number of levels of main effects. The eigenvalues and Z are as follows. The numbers in brackets are multiplicities of the corresponding eigenvalues.

(i) Case without interaction effects: Let m_j be the number of levels of jth main effects and let $Z = [\mathbf{z}_0\ Z_1 \cdots Z_k]$ where $Z_j = [\mathbf{z}_1 \cdots \mathbf{z}_{m_j-1}]$, $j = 1, ..., k$. Then for $i(j) = 1, ..., m_j - 1$

$$\lambda_0 = m + \sum_{j=1}^{k} \frac{m}{m_j}\ [1], \quad \lambda_j = \frac{m}{m_j}\ [m_j - 1],$$

$$\mathbf{z}_0 = \frac{1}{\sqrt{m}}\mathbf{1}_m$$

$$\mathbf{z}_{i(j)} = \frac{\sqrt{m_j}}{\sqrt{m}\sqrt{i(j)[i(j)+1]}} \times$$

$$\left[\mathbf{1}_{m_1} \otimes \cdots \otimes \left(-i(j)\mathbf{e}_{i(j)+1} + \sum_{k(j)=1}^{i(j)} \mathbf{e}_{k(j)}\right)_{m_j} \otimes \cdots \otimes \mathbf{1}_{m_k}\right]$$

where $\mathbf{1}_m$ is an $m \times 1$ vector of unities and $(-i(j)\mathbf{e}_{i(j)+1} + \sum_{k(j)=1}^{i(j)} \mathbf{e}_{k(j)})_{m_j}$ is an $m_j \times 1$ vector having 1 as first through i(j)th components, -i(j) as (i(j)+1)th component and zero elsewhere.

(ii) Case with c interaction effects: Let S_j be the index set of main factors which are contained in jth interaction effect and $d_j = \prod_{s\in S_j} m_s$. Also let $Z = [\mathbf{z}_0\ Z_1 \cdots Z_k\ Z_{k+1} \cdots Z_{k+c}]$, where $Z_i = [\mathbf{z}_1\ \cdots\ \mathbf{z}_{m_i-1}]$ and Z_{k+j} is the matrix whose columns are $\mathbf{z}_{l(s)_j}$'s. Then for $l(i) = 1, ..., m_i - 1$, and $l(s) = 1, ..., m_s - 1$ for each $s \in S_j$

$$\lambda_0 = m + \sum_{i=1}^{k} \frac{m}{m_i} + \sum_{j=1}^{c} \frac{m}{d_j}\ [1], \quad \lambda_i = \frac{m}{m_i} + \sum_{j=1}^{c} h_j\ [m_i - 1],$$

$$\lambda_{k+j} = \frac{m}{d_j} + \sum_{n=j+1}^{c} t_n\ [\prod_{s\in S_j}(m_s - 1)],$$

$$\mathbf{z}_0 = \frac{1}{\sqrt{m}}\mathbf{1}_m$$

$$\mathbf{z}_{l(i)} = \frac{\sqrt{m_i}}{\sqrt{m}\sqrt{l(i)[l(i)+1]}} \times \left[\mathbf{1}_{m_1} \otimes \cdots \otimes \left(-l(i)\mathbf{e}_{l(i)+1} + \sum_{k(i)=1}^{l(i)} \mathbf{e}_{k(i)}\right)_{m_i} \otimes \cdots \otimes \mathbf{1}_{m_k}\right]$$

$$\mathbf{z}_{l(s)_j} = \frac{\sqrt{d_j}}{\sqrt{m}\sqrt{\sum_{s\in S_j} l(s)(l(s)+1)}} \left[\otimes_{i=1}^{k} B_i\right]$$

where

$$h_j = \begin{cases} m/d_j & \text{if } i \in S_j \\ 0 & \text{if } i \notin S_j \end{cases}, \quad t_n = \begin{cases} m/d_n & \text{if } S_j \subset S_n \\ 0 & \text{if } S_j \not\subset S_n \end{cases},$$

$$B_i = \begin{cases} -l(i)\mathbf{e}_{l(i)+1} + \sum_{k(i)=1}^{l(i)} \mathbf{e}_{k(i)} & \text{if } i \in S_j \\ \mathbf{1}_{m_i} & \text{if } i \notin S_j \end{cases}$$

Next, because of $ZZ' = P_{X_o}$ and $Z'Z = I_r$, Z is the matrix whose columns are orthonormalized eigenvectors corresponding to eigenvalue 1 of P_{X_o}. This fact is available to extend the orthogonal projection matrix for an unbalanced model(Kim and Park(1992)). Also in view of (2.2), it is interesting to note that the decompositions of the projectors hold as shown in the following corollaries.

Corollary 2.1. Let $X_o = [\mathbf{1}_m\ X_1 \cdots X_k]$ be the corresponding balanced design matrix for k-way factorial design without interactions, where X_i is the incidence matrix for ith main effect. Then

$$P_{X_o} = P_{X_1} + \cdots + P_{X_k} - (k-1)P_M$$

where $P_M = \mathbf{1}_m(\mathbf{1}_m'\mathbf{1}_m)^{-1}\mathbf{1}_m'$.

Proof. From the fact that X_i's can be expressed as Kronecker products of identity matrix and column vectors of unities(Rogers(1984)), that is, $X_i = \mathbf{1}_{m_1} \otimes \cdots \otimes I_{m_i} \otimes \cdots \otimes \mathbf{1}_{m_k}$, we have

$$P_{X_i} = X_i(X_i'X_i)^+X_i' = \frac{m_i}{m}X_iX_i'.$$

By (2.2)

$$P_{X_o} = ZZ' = \mathbf{z}_0\mathbf{z}_0' + \sum_{j=1}^{k} Z_jZ_j'.$$

Then

$$\mathbf{z}_0\mathbf{z}_0' = \frac{1}{m}\mathbf{1}_m\mathbf{1}_m' = \mathbf{1}_m(\mathbf{1}_m'\mathbf{1}_m)^{-1}\mathbf{1}_m' \equiv P_M$$

and for $j = 1, 2, ..., k$

$$\begin{aligned} Z_jZ_j' &= \sum_{i(j)=1}^{m_j-1} \mathbf{z}_{i(j)}\mathbf{z}_{i(j)}' \\ &= \frac{m_j}{m}\sum_{i(j)=1}^{m_j-1}\left[J_{m_1} \otimes \cdots \otimes \frac{1}{i(j)[i(j)+1]}H_{i(j)} \otimes \cdots \otimes J_{m_k}\right] \\ &= \frac{m_j}{m}\left[J_{m_1} \otimes \cdots \otimes \sum_{i(j)=1}^{m_j-1}\frac{1}{i(j)[i(j)+1]}H_{i(j)} \otimes \cdots \otimes J_{m_k}\right] \end{aligned}$$

where $H_{i(j)} = [1 \ \cdots \ 1 \ -i(j) \ 0 \ \cdots \ 0]'[1 \ \cdots \ 1 \ -i(j) \ 0 \ \cdots \ 0]$ and J_{m_j} is an $m_j \times m_j$ matrix of unities. Let $H = \sum_{i(j)=1}^{m_j-1} \frac{1}{i(j)[i(j)+1]}H_{i(j)}$ and h_{ij} be the (i,j)th element of H. Then we have

$$h_{ij} = \begin{cases} 1 - \frac{1}{m_j} & \text{if } i = j \\ -\frac{1}{m_j} & \text{if } i \neq j \end{cases}.$$

Hence we have $H = I_{m_j} - \frac{1}{m_j}J_{m_j}$ and $Z_jZ_j' = \frac{m_j}{m}X_jX_j' - \frac{1}{m}J_m = P_{X_j} - P_M$. This completes the proof.

Corollary 2.2. Let $X_o = [\mathbf{1}_m \; U_o \; V_o]$ be the corresponding balanced design matrix for the model with k main effects and c interaction effects, where U_o and V_o are the incidence matrices for the main and interaction effects, respectively. Then

$$P_{X_o} = P_U + P_{V_o} - P_U P_{V_o}$$

where $U = [\mathbf{1}_m \; U_o]$.

Proof. Since $X_o = [U \; V_o]$ we get

$$X_o' X_o = \begin{bmatrix} U' \\ V_o' \end{bmatrix} [U \quad V_o] = \begin{bmatrix} U'U & U'V_o \\ V_o'U & V_o'V_o \end{bmatrix}$$

and

$$P_{X_o} = [U \quad V_o] \begin{bmatrix} U'U & U'V_o \\ V_o'U & V_o'V_o \end{bmatrix}^{-} \begin{bmatrix} U' \\ V_o' \end{bmatrix}.$$

See Marsaglia and Styan(1974) or Searle(1982) on a generalized inverse of the partitioned matrix. Also it can be easily shown that $P_U P_{V_o}$ is symmetric. The proof follows from the two results.

We introduce some examples to find the orthogonal projection matrix P_{X_o}.

Example 1. Consider a three-way model

$$\mathbf{y} = \mathbf{1}_m \mu + X_1 \alpha + X_2 \beta + X_3 \gamma + \mathbf{e}$$

with $m_1 = m_2 = m_3 = 2$. For the model we would have

$$X_o = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Then appropriate matrices are as follows:

$$\Lambda_r = \begin{bmatrix} 20 & & & \\ & 4 & & \\ & & 4 & \\ & & & 4 \end{bmatrix}, \quad Z = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}.$$

Hence from (2.2) P_{X_o} becomes

$$P_{X_o} = \frac{1}{4} \begin{bmatrix} 2 & 1 & 1 & 0 & 1 & 0 & 0 & -1 \\ 1 & 2 & 0 & 1 & 0 & 1 & -1 & 0 \\ 1 & 0 & 2 & 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 1 & 2 & -1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 & 2 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 & 1 & 2 & 0 & 1 \\ 0 & -1 & 1 & 0 & 1 & 0 & 2 & 1 \\ -1 & 0 & 0 & 1 & 0 & 1 & 1 & 2 \end{bmatrix}.$$

Example 2. Consider the three-way model with $\alpha\beta$ interaction effect

$$\mathbf{y} = \mathbf{1}_m\mu + X_1\alpha + X_2\beta + X_3\gamma + V(\alpha\beta) + \mathbf{e}$$

having $m_1 = m_2 = m_3 = 2$. For the model we would get $S_1 = \{1,2\}$, $d_1 = m_1 m_2 = 4$ and

$$X_o = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Also as in example 1 appropriate matrices are given by

$$\Lambda_r = \begin{bmatrix} 22 & & & & \\ & 6 & & & \\ & & 6 & & \\ & & & 4 & \\ & & & & 2 \end{bmatrix}, \quad Z = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 & 1 \end{bmatrix}.$$

By (2.2) P_{X_o} becomes

$$P_{X_o} = \frac{1}{8} \begin{bmatrix} 5 & 3 & 1 & -1 & 1 & -1 & 1 & -1 \\ 3 & 5 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 5 & 3 & 1 & -1 & 1 & -1 \\ -1 & 1 & 3 & 5 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 5 & 3 & 1 & -1 \\ -1 & 1 & -1 & 1 & 3 & 5 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 5 & 3 \\ -1 & 1 & -1 & 1 & -1 & 1 & 3 & 5 \end{bmatrix}.$$

3. AN APPLICATION

It is important to compute various sums of squares that are necessary for analyzing the model. In this point of view, the orthogonal projection matrix is useful. As an application, using this P_X we describe the computing procedure for F-statistic for testing a general linear hypothesis. Now the general hypothesis is taken to be of the form

$$H : L'\beta = \mathbf{q} \tag{3.1}$$

where L' is of full row rank. Under the normality assumption of $\mathbf{y}$, that is, $\mathbf{y} \sim N(X\beta, \sigma^2 I)$, the F-statistic for testing $H : L'\beta = \mathbf{q}$ is

$$F(H) = \frac{(L'\hat{\beta} - \mathbf{q})'(L'GL)^{-1}(L'\hat{\beta} - \mathbf{q})}{\text{rank}(L)\hat{\sigma}^2} \tag{3.2}$$

where $\hat{\beta}$ is a solution of the normal equations, G is a generalized inverse of $X'X$, and

$$\hat{\sigma}^2 = \frac{\text{SSE}}{N - \text{rank}(X)} = \frac{\mathbf{y}'(I - P_X)\mathbf{y}}{N - \text{rank}(X)}. \tag{3.3}$$

Note that $\hat{\beta}$ is a solution of the normal equations among the possible solutions. So we can take a generalized inverse of $X'X$ as $(X'X)^{+}$. Hence for a balanced model, $L'GL$ and $L'\hat{\beta}$ can be easily computed as follows:

$$L'GL = M'P_X M, \tag{3.4}$$
$$L'\hat{\beta} = M'P_X \mathbf{y} \tag{3.5}$$

where $L' = M'X$ for some M.

4. CONCLUSIONS

From (2.2) we can easily obtain the orthogonal projection matrix for a given balanced model by computing Z. Also since the explicit form of Z depends on only the number of levels of main effects, it is simple to find Z whether there are interactions or not. Hence we can reduce much computation which is necessary for obtaining P_X. This Z is also useful to obtain the explicit form of the orthogonal projection matrix for an unbalanced model. Besides, this result can be applied to compute not only the minimum norm generalized least squares solution of the normal equation, but also the best linear unbiased estimator of $X\beta$ in the general model $\mathrm{E}(\mathbf{y}) = X\beta$, $\mathrm{Cov}(\mathbf{y}) = \sigma^2 H$ for a symmetric positive definite matrix.

REFERENCES

(1) Kennedy, W.J. and Gentle, J.E. (1980). *Statistical computing.* Marcel Dekker, Inc.

(2) Kim, B.C. and Lee, J.T. (1986). The Moore-Penrose inverse for the classification models. *Journal of the Korean Statistical Society*, **15**, 46-61.

(3) Kim, B.C. and Park, J.T. (1992). The orthogonal projection matrix for the unbalanced factorial design. *Journal of the Japanese Society of Computational Statistics*, **5**, 33-42.

(4) Kim, B.C. and Sunwoo, H.S. (1989). The Moore-Penrose inverse for the balanced ANOVA models. *Journal of the Japanese Society of Computational Statistics*, **2**, 1-7.

(5) Kim, B.C. and Sunwoo, H.S. (1990). The M-P inverse for the balanced factorial design with interactions. *Journal of the Japanese Society of Computational Statistics*, **3**, 25-34.

(6) Lowerre, J.M. (1982). An introduction to modern matrix methods and statistics. *The American Statistician*, **36**, 113-115.

(7) Marsaglia, G. and Styan, G.P.H. (1974). Rank conditions for generalized inverses of partitioned matrices. *Sankhyā, Indian Journal of Statistics*, **36**, 437-442.

(8) Rogers, G.S. (1984). Kronecker products in ANOVA – A first step. *The American Statistician*, **38**, 197-202.

(9) Searle, S.R. (1982). *Matrix algebra useful for statistics.* New York, John Wiley.