VARIANCE ANALYSES FOR KERNEL REGRESSORS WITH NESTED REPRODUCING KERNEL HILBERT SPACES

Akira Tanaka, Hideyuki Imai*

Division of Computer Science
Hokkaido University
Kita-14, Nishi-9, Kita-ku,
Sapporo, 060-0814 Japan.

Koji Takamiya

Faculty of Economics
Niigata University
8050 2-no-cho, Ikarashi, Nishi-ku,
Niigata, 950-2181 Japan.

ABSTRACT

Learning based on kernel machines is widely known as a powerful tool for various fields of information science including signal processing such as function estimation from finite sampling points. One of central topics of kernel machines is model selection, especially selection of a kernel or its parameters. In our previous works, we investigated the generalization error of a model space itself corresponding to a selected kernel in kernel regressors. In this paper, we discuss the generalization error in a model space corresponding to a selected kernel in kernel regressors; and prove that the variance of a learning result is reduced when we adopt a kernel corresponding to a larger reproducing kernel Hilbert space.

Index Terms— kernel machines, model selection, generalization error, orthogonal projection, variance

1. INTRODUCTION

Learning based on kernel machines [1], represented by the support vector machine [2] and the kernel ridge regression [3, 4], is widely known as a powerful tool for various fields of information science such as pattern recognition, regression estimation, and density estimation. Moreover, these techniques are also important in the fields of signal processing, represented by sampling theorems aiming to reconstruct signals from finite sampling points. In general, an appropriate model selection, aiming to minimize generalization error, is required in order to obtain a desirable learning result by kernel machines. Although many methods for model selection have been proposed (see [5] for instance), theoretical analyses of generalization error are still important since they may be useful for construction or analysis of model selection methods. Generalization error can be decomposed into two components. One is the error of a model space itself, that is, the distance between an unknown true function and the model space, and the other is an error in the model space, that is, the distance between a learning result and the orthogonal projection of the unknown true function onto the model space. In our previous work [6], we investigated the former one and proved that a kernel corresponding to the smallest reproducing kernel Hilbert space (RKHS), including an unknown true function, gives the best model space among a class of nested RKHS's with an invariant metric. Moreover, we also clarified in [7] that an invariant metric is a crucial condition for this property. On the other hand, the behavior of the latter generalization error with respect to a kernel or its parameters is not discussed sufficiently, while that with respect to training data set was sufficiently investigated (see [8] for instance).

In this paper, we consider a class of kernels corresponding to a nested class of RKHS's with an invariant metric as the same with [6] and analyze the latter generalization error for those RKHS's. On the basis of our analyses, we prove that the latter generalization error can be reduced when we adopt a kernel corresponding to a larger RKHS.

2. OVERVIEW OF THE THEORY OF REPRODUCING KERNEL HILBERT SPACES

In this section, we give an overview of the theory of reproducing kernel Hilbert spaces [9, 10].

Definition 1 *[9] Let $\mathbf{R}^n$ be an n-dimensional real vector space and let $\mathcal{H}$ be a class of functions defined on $\mathcal{D} \subset \mathbf{R}^n$, forming a Hilbert space of real-valued functions. The function $K(\boldsymbol{x}, \tilde{\boldsymbol{x}}),\ (\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \mathcal{D})$ is called a reproducing kernel of $\mathcal{H}$, if*

1. *For every fixed $\tilde{\boldsymbol{x}} \in \mathcal{D}$, $K(\cdot, \tilde{\boldsymbol{x}})$ is a function belonging to $\mathcal{H}$.*
2. *For every fixed $\tilde{\boldsymbol{x}} \in \mathcal{D}$ and every fixed $f(\cdot) \in \mathcal{H}$,*

$$f(\tilde{\boldsymbol{x}}) = \langle f(\cdot), K(\cdot, \tilde{\boldsymbol{x}}) \rangle_{\mathcal{H}}, \tag{1}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ denotes the inner product of the Hilbert space $\mathcal{H}$.

*This work was partially supported by Grant-in-Aid No.21700001 for Young Scientists (B) from the Ministry of Education, Culture, Sports and Technology of Japan.

The Hilbert space that has a reproducing kernel K is called a reproducing kernel Hilbert space (RKHS), denoted by $\mathcal{H}_K$. The reproducing property Eq.(1) enables us to treat a value of a function at a point in $\mathcal{D}$ while we can not deal with a value of a function in a general Hilbert space such as L^2. Note that reproducing kernels are positive definite [9]:

$$\sum_{i,j=1}^{N} c_i c_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0, \tag{2}$$

for any N, $c_1, \ldots, c_N \in \mathbf{R}$, and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathcal{D}$. In addition, $K(\boldsymbol{x}, \tilde{\boldsymbol{x}}) = K(\tilde{\boldsymbol{x}}, \boldsymbol{x})$ for any $\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \mathcal{D}$ is followed [9]. If a reproducing kernel $K(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ exists, it is unique [9]. Conversely, every positive definite function $K(\boldsymbol{x}, \tilde{\boldsymbol{x}})$ has the unique corresponding RKHS [9].

Next, we introduce the Schatten product [11] that is a convenient tool to reveal the reproducing property of kernels.

Definition 2 *[11] Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces. The Schatten product of $g \in \mathcal{H}_2$ and $h \in \mathcal{H}_1$ is defined by*

$$(g \otimes h) f = \langle f, h \rangle_{\mathcal{H}_1} g, \quad f \in \mathcal{H}_1. \tag{3}$$

Note that $(g \otimes h)$ is a linear operator from $\mathcal{H}_1$ onto $\mathcal{H}_2$. It is easy to show that the following relations hold for $h, v \in \mathcal{H}_1$, $g, u \in \mathcal{H}_2$.

$$(h \otimes g)^* = (g \otimes h), \quad (h \otimes g)(u \otimes v) = \langle u, g \rangle_{\mathcal{H}_2} (h \otimes v), \tag{4}$$

where the superscript * denotes the adjoint operator.

3. FORMULATION OF LEARNING PROBLEMS

Let $\{(y_k, \boldsymbol{x}_k) \mid k \in \{1, \ldots, \ell\}\}$ be a given training data set with an output value $y_k \in \mathbf{R}$ and the corresponding input vector $\boldsymbol{x}_k \in \mathbf{R}^n$, satisfying

$$y_k = f(\boldsymbol{x}_k) + n_k, \tag{5}$$

where $f(\cdot)$ denotes an unknown true function and n_k denotes a zero-mean additive noise. In pattern recognition problems, y_k denotes a class label, and in regression estimation problems, it denotes a value of the function $f(\cdot)$ at the point $\boldsymbol{x}_k$ with additive noise. The aim of machine learning is to estimate the unknown true function $f(\cdot)$ by using the given training data set and statistical properties of the noise.

In this paper, we assume that the unknown true function $f(\cdot)$ belongs to the RKHS $\mathcal{H}_K$ corresponding to a certain kernel K. If $f \in \mathcal{H}_K$, then Eq.(5) is rewritten as

$$y_i = \langle f(\cdot), K(\cdot, \boldsymbol{x}_i) \rangle_{\mathcal{H}_K} + n_i, \tag{6}$$

on the basis of the reproducing property of kernels. Let $\boldsymbol{y} = [y_1, \ldots, y_\ell]'$ and $\boldsymbol{n} = [n_1, \ldots, n_\ell]'$ with the superscript $'$ denoting the transposition operator for a matrix (or a vector), then applying the Schatten product to Eq.(6) yields

$$\boldsymbol{y} = \left(\sum_{k=1}^{\ell} [\boldsymbol{e}_k^{(\ell)} \otimes K(\cdot, \boldsymbol{x}_k)] \right) f(\cdot) + \boldsymbol{n}, \tag{7}$$

where $\boldsymbol{e}_k^{(\ell)}$ denotes the k-th vector of the canonical basis of $\mathbf{R}^\ell$. For convenience of description, we write

$$A_{K,X} = \left(\sum_{k=1}^{\ell} [\boldsymbol{e}_k^{(\ell)} \otimes K(\cdot, \boldsymbol{x}_k)] \right), \tag{8}$$

where $X = \{\boldsymbol{x}_k \in \mathbf{R}^\ell \mid k \in \{1, \ldots, \ell\}\}$ be input training data set. $A_{K,X}$ is a linear operator that maps an element in $\mathcal{H}_K$ onto $\mathbf{R}^\ell$ and Eq.(7) can be written by

$$\boldsymbol{y} = A_{K,X} f(\cdot) + \boldsymbol{n}, \tag{9}$$

which represents the relation between the unknown true function $f(\cdot)$ and an output vector $\boldsymbol{y}$. The information of input vectors is integrated in the operator $A_{K,X}$. Therefore, a machine learning problem can be interpreted as an inversion problem of the linear equation Eq.(9) [12].

4. GENERALIZATION ERROR IN A MODEL SPACE

As is well known, the minimum-norm least-squares solution of Eq.(9) is given as

$$\hat{f}(\cdot) = A_{K,X}^{+} \boldsymbol{y} = A_{K,X}^{+} A_{K,X} f(\cdot) + A_{K,X}^{+} \boldsymbol{n} \tag{10}$$

where $A_{K,X}^{+}$ denotes the Moore-Penrose generalized inverse operator of $A_{K,X}$ [13]. Note that $P_{K,X} = A_{K,X}^{+} A_{K,X}$ is the orthogonal projector onto the model space $\mathcal{R}(A_{K,X}^{*})$ (the range space of $A_{K,X}^{*}$) which is the linear subspace in $\mathcal{H}_K$ spanned by $\{K(\cdot, \boldsymbol{x}_k) \mid k \in \{1, \ldots, \ell\}\}$.

In general, the generalization error of a learning result is defined by $||f(\cdot) - \hat{f}(\cdot)||_{\mathcal{H}_K}^2$, where $|| \cdot ||_{\mathcal{H}_K}$ denotes the induced norm of $\mathcal{H}_K$. According to the Pythagorean theorem, we have

$$\begin{aligned} &||f(\cdot) - \hat{f}(\cdot)||_{\mathcal{H}_K}^2 \\ &\quad = ||f(\cdot) - P_K f(\cdot)||_{\mathcal{H}_K}^2 + ||P_K f(\cdot) - \hat{f}(\cdot)||_{\mathcal{H}_K}^2 \\ &\quad = ||f(\cdot) - P_K f(\cdot)||_{\mathcal{H}_K}^2 + ||A_{K,X}^{+} \boldsymbol{n}||_{\mathcal{H}_K}^2. \end{aligned} \tag{11}$$

The first term in Eq.(11) can be regarded as the generalization error of the model space $\mathcal{R}(A_{K,X}^{*})$, which is the squared distance between the unknown true function $f(\cdot)$ and the model space $\mathcal{R}(A_{K,X}^{*})$ whose behavior with respect to a kernel was discussed in [6].

On the other hand, the second term in Eq.(11) can be regarded as the generalization error in a model space $\mathcal{R}(A_{K,X}^{*})$, which is the target of our analysis in this paper. Applying Eq.(4) to the equation $A_{K,X}^{+} = A_{K,X}^{*} (A_{K,X} A_{K,X}^{*})^{+}$ yields

$$A_{K,X}^{+} = \left(\sum_{k=1}^{\ell} [K(\cdot, \boldsymbol{x}_k) \otimes \boldsymbol{e}_k^{(\ell)}] \right) G_{K,X}^{+}, \tag{12}$$

where $G_{K,X}$ denotes the Gramian matrix of K with X defined as $G_{K,X} = (K(\boldsymbol{x}_i, \boldsymbol{x}_j))$. Therefore, we have

$$\begin{aligned} A^+_{K,X}\boldsymbol{n} &= \left(\sum_{k=1}^{\ell}[K(\cdot,\boldsymbol{x}_k)\otimes \boldsymbol{e}_k^{(\ell)}]\right) G^+_{K,X}\boldsymbol{n} \\ &= \sum_{j=1}^{\ell}\sum_{k=1}^{\ell} n_j (G^+_{K,X})_{jk} K(\cdot,\boldsymbol{x}_k) \end{aligned}$$

and

$$\begin{aligned} &J(K,X;\boldsymbol{n}) \\ &= ||A^+_{K,X}\boldsymbol{n}||^2_{\mathcal{H}_K} = \langle A^+_{K,X}\boldsymbol{n}, A^+_{K,X}\boldsymbol{n}\rangle_{\mathcal{H}_K} \\ &= \sum_{j,k,s,t=1}^{\ell} n_j (G^+_{K,X})_{jk} n_s (G^+_{K,X})_{st} \\ &\quad \times \langle K(\cdot,\boldsymbol{x}_k), K(\cdot,\boldsymbol{x}_t)\rangle_{\mathcal{H}_K} \\ &= \sum_{j,k,s,t=1}^{\ell} n_j (G^+_{K,X})_{jk} n_s (G^+_{K,X})_{st} (G_{K,X})_{kt} \\ &= \boldsymbol{n}' G^+_{K,X} G_{K,X} G^+_{K,X}\boldsymbol{n} = \boldsymbol{n}' G^+_{K,X}\boldsymbol{n}, \end{aligned}$$

since

$$\langle K(\cdot,\boldsymbol{x}_k), K(\cdot,\boldsymbol{x}_t)\rangle_{\mathcal{H}_K} = K(\boldsymbol{x}_k,\boldsymbol{x}_t)$$

holds and a Gramian matrix is symmetric.

5. ANALYSES OF GENERALIZATION ERROR WITH NESTED RKHS'S

In this section, we consider a class of kernels forming a nested class of RKHS's with an invariant metric and analyze the generalization error in a model space defined in the previous section with those kernels.

Let K_1 and K_2 be kernels corresponding to RKHS's $\mathcal{H}_{K_1}$ and $\mathcal{H}_{K_2}$ satisfying

$$\mathcal{H}_{K_1} \subset \mathcal{H}_{K_2} \tag{13}$$

and

$$||f(\cdot)||^2_{\mathcal{H}_{K_1}} = ||f(\cdot)||^2_{\mathcal{H}_{K_2}} \tag{14}$$

for any $f(\cdot) \in \mathcal{H}_{K_1}$.

The following theorems hold for kernels corresponding to nested class of RKHS's.

Theorem 1 *[9] If K is the reproducing kernel of the class F with the norm $||\cdot||$, and if the linear class $F_1 \subset F$ forms a Hilbert space with the norm $||\cdot||_1$, such that $||f||_1 \geq ||f||$ for any $f \in F_1$, then the class F_1 possesses a reproducing kernel K_1 such that $K^c = K - K_1$ is also a reproducing kernel.*

Theorem 2 *[9] If K and K_1 are the reproducing kernels of the classes of F and F_1 with the norms $||\cdot||$, $||\cdot||_1$, and if $K - K_1$ is a reproducing kernel, then $F_1 \subset F$ and $||f_1||_1 \geq ||f_1||$ for every $f_1 \in F_1$.*

From Theorem 1 and Eq.(14), it is trivial that

$$K_2 = K_1 + K^c \tag{15}$$

holds with a certain kernel K^c. Therefore, we have

$$G_{K_2,X} = G_{K_1,X} + G_{K^c,X}. \tag{16}$$

Here, we introduce the following lemma.

Lemma 1 *[6] Let $H_1,\ H_2$ be Hermitian matrices and let $\boldsymbol{y} \in \mathcal{R}(H_1)$, then*

$$\boldsymbol{y}^*(H_1^+ - (H_1+H_2)^+)\boldsymbol{y} \geq 0 \tag{17}$$

holds.

The following theorem is the main result in this paper.

Theorem 3 *Let K_1 and K_2 be kernels satisfying Eqs.(13) and (14); and let $X = \{\boldsymbol{x}_k \in \mathbf{R}^\ell \mid k \in \{1,\ldots,\ell\}\}$ and $\boldsymbol{n} \in \mathbf{R}^\ell$ be an arbitrary set of input training vectors and an arbitrary noise vector, respectively. If $G_{K_1,X}$ is non-singular,*

$$||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} \geq ||A^+_{K_2,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} \tag{18}$$

holds.

Proof From the assumption Eq.(14) and the trivial fact that $\mathcal{R}(A^+_{K_1,X}) = \mathcal{R}(A^*_{K_1,X}) \subset \mathcal{H}_{K_1}$ holds, we have

$$||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} = ||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_1}}.$$

Therefore, for any $\boldsymbol{n} \in \mathbf{R}^\ell = \mathcal{R}(G_{K_1,X})$,

$$\begin{aligned} &||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} - ||A^+_{K_2,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} \\ &= ||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_1}} - ||A^+_{K_2,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} \\ &= J(K_1,X;\boldsymbol{n}) - J(K_2,X;\boldsymbol{n}) \\ &= \boldsymbol{n}' G^{-1}_{K_1,X}\boldsymbol{n} - \boldsymbol{n}' G^{-1}_{K_2,X}\boldsymbol{n} \\ &= \boldsymbol{n}'(G^{-1}_{K_1,X} - (G_{K_1,X} + G_{K^c,X})^{-1})\boldsymbol{n} \geq 0 \end{aligned}$$

is obtained by Lemma 1, which concludes the proof. □

From Theorem 3, it is concluded that when we adopt a kernel corresponding to a larger RKHS, the generalization error in the model space is reduced, while the contrary conclusion was obtained for the generalization error of a model space as shown in [6].

This result can be trivially extended to the variance of the generalization error in a model space as follows.

Corollary 1 *Under the conditions in Theorem 3,*

$$E_{\boldsymbol{n}}||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} \geq E_{\boldsymbol{n}}||A^+_{K_2,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} \tag{19}$$

holds, where $E_{\boldsymbol{n}}$ denotes the expectation operator over the random vector $\boldsymbol{n}$.

Note that it is not guaranteed that Eq.(18) holds when the metrics of $\mathcal{H}_{K_1}$ and $\mathcal{H}_{K_2}$ differ. In fact, Theorem 2 implies

$$||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_1}} \geq ||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}}, \tag{20}$$

which may break the proof of Theorem 3. When a nested class of RKHS's does not have an invariant metric, the generalization error in a model space $\mathcal{R}(A^*_{K_1,X})$ evaluated in $\mathcal{H}_{K_2}$ is reduced to

$$\begin{aligned} &||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} \\ &= \sum_{j,k,s,t=1}^{\ell} n_j (G^{-1}_{K_1,X})_{jk} n_s (G^{-1}_{K_1,X})_{st} \\ &\quad \times \langle K_1(\cdot,\boldsymbol{x}_k), K_1(\cdot,\boldsymbol{x}_t)\rangle_{\mathcal{H}_{K_2}} \\ &= \boldsymbol{n}' G^{-1}_{K_1,X} \tilde{G}_{K_1,X} G^{-1}_{K_1,X} \boldsymbol{n}, \end{aligned}$$

where $\tilde{G}_{K_1,X}$ denotes the Gramian matrix of K_1 with X evaluated in $\mathcal{H}_{K_2}$ defined as

$$\tilde{G}_{K_1,X} = (\langle K_1(\cdot,\boldsymbol{x}_i), K_1(\cdot,\boldsymbol{x}_j)\rangle_{\mathcal{H}_{K_2}}). \tag{21}$$

Therefore, we have

$$\begin{aligned} &||A^+_{K_1,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} - ||A^+_{K_2,X}\boldsymbol{n}||^2_{\mathcal{H}_{K_2}} \\ &= \boldsymbol{n}' G^{-1}_{K_1,X} \boldsymbol{n} - \boldsymbol{n}' G^{-1}_{K_2,X} \boldsymbol{n} \\ &= \boldsymbol{n}' (G^{-1}_{K_1,X} \tilde{G}_{K_1,X} G^{-1}_{K_1,X} - G^{-1}_{K_2,X}) \boldsymbol{n}. \end{aligned}$$

Let

$$M = G^{-1}_{K_1,X} \tilde{G}_{K_1,X} G^{-1}_{K_1,X} - G^{-1}_{K_2,X}. \tag{22}$$

If all eigenvalues of M are non-negative, it is concluded that the same result with Theorem 3 is obtained for variant metric cases. On the other hand, if there exist negative eigenvalues in M, it implies that Theorem 3 does not hold for variant metric cases. We conducted some numerical experiments for eigenstructure of M with Gaussian kernels which form a nested class of RKHS's with variant metrics as shown in [7]. However, we could not find a case where M has negative eigenvalues. Theoretical analyses for this issue is one of our future works.

6. CONCLUSION

In this paper, we discussed the generalization error in a model space of kernel regressors and showed that the generalization error in a model space is reduced when we adopt a kernel corresponding to a larger reproducing kernel Hilbert space among a class of kernels forming a class of nested reproducing kernel Hilbert spaces with an invariant metric, which is a contrary conclusion for the generalization error of a model space. Theoretical analyses for a class of RKHS's with variant metrics and extension of our result for useful kernel machines such as support vector machines and kernel ridge regressors is one of our future works to be undertaken.

7. REFERENCES

[1] K. Muller, S. Mika, G. Ratsch, K. Tsuda, and B. Scholkopf, "An Introduction to Kernel-based Learning Algorithms," *IEEE Transactions on Neural Networks*, vol. 12, pp. 181–201, 2001.

[2] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, 1999.

[3] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Recognition*, Cambridge University Press, Cambridge, 2004.

[4] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, Cambridge, 2000.

[5] M. Sugiyama, M. Kawanabe, and K. Muller, "Trading Variance Reduction with Unbiasedness: The Regularized Subspace Information Criterion for Robust Model Selection in Kernel Regression," *Neural Computation*, vol. 16, no. 5, pp. 1077–1104, 2004.

[6] A. Tanaka, H. Imai, M. Kudo, and M. Miyakoshi, "Optimal Kernel in a Class of Kernels with an Invariant Metric," in *Joint IAPR Internatioanl Workshops SSPR 2008 and SPR 2008*. 2008, pp. 530–539, Springer.

[7] A. Tanaka, , H. Imai, M. Kudo, and M. Miyakoshi, "Theoretical Analyses on a Class of Nested RKHS's," in *2011 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2011)*, 2011, pp. 2072–2075.

[8] A. Tanaka, , H. Imai, M. Kudo, and M. Miyakoshi, "A Relationship Between Generalization Error and Training Samples in Kernel Regressors," in *20th International Conference on Patter Recognition (ICPR2010)*, 2010, pp. 1421–1424.

[9] N. Aronszajn, "Theory of Reproducing Kernels," *Transactions of the American Mathematical Society*, vol. 68, no. 3, pp. 337–404, 1950.

[10] J. Mercer, "Functions of Positive and Negative Type and Their Connection with The Theory of Integral Equations," *Transactions of the London Philosophical Society*, vol. A, no. 209, pp. 415–446, 1909.

[11] R. Schatten, *Norm Ideals of Completely Continuous Operators*, Springer-Verlag, Berlin, 1960.

[12] H. Ogawa, "Neural Networks and Generalization Ability," *IEICE Technical Report*, vol. NC95-8, pp. 57–64, 1995 (in Japanese).

[13] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and Its Applications*, John Wiley & Sons, 1971.