コンピュータビジョン　80－6
(1992. 11. 19)

相対カルフーネン·ロエーブ作用素

山下 幸彦　　小川 英光

東京工業大学工学部情報工学科

東京都目黒区大岡山 2–12–1

あらまし　カルフーネン·ロエーブ (K–L) 部分空間は, ある決められた次元の部分空間の中で, 確率的に分布する信号を最も精度良く近似する部分空間である. K–L 部分空間は, 通信における情報圧縮, および, パターン認識における部分空間法などに用いられている. しかし, K–L 部分空間による近似では, 通信の分野における雑音, パターン認識の分野における雑音や他カテゴリのパターンの影響などは考慮されていない. このために, 通信における雑音の抑制が不十分であったり, パターン認識において, 近接パターンに誤認識が生じやすいという問題が生じる. この問題を解決するために, 相対 K–L 作用素の概念を提案する. 相対 K–L 作用素は, 値域の次元が決められた作用素の中で, 近似した信号と元の信号の平均 2 乗誤差と雑音による平均 2 乗誤差の和を最小にするものである. さらに, 本論文では, 相対 K–L 作用素が存在するための必要十分条件とその一般形を与える.

和文キーワード　カルフーネン·ロエーブ部分空間, カルフーネン·ロエーブ展開, パターン認識, 情報圧縮

Relative Karhunen-Loève operator

Y. Yamashita　and　H. Ogawa

Department of Computer Science, Faculty of Engineering

Tokyo Institute of Technology

2–12–1 Ookayama, Meguro-ku, Tokyo

Abstract　The Karhunen-Loève (K–L) subspace is a subspace which provides the best approximation for a stochastic signal under the condition that its dimension is fixed. The K–L subspace has been successfully used in the data compression in communication and the subspace method in pattern recognition. The K–L subspace, however, does not consider a noise in communication and a noise and other patterns in pattern recognition. Therefore, its noise suppression is not sufficient in communication and it gives a wrong recognition result for patterns which are similar to each other.

In order to solve this problem, we propose a concept of the relative K–L operator. The relative K–L operator minimizes the sum of the mean square error between the original signal and the approximated signal and the mean square error caused by noise under the condition that the dimension of its range is fixed. We provide the conditions under which the relative K–L operator exists. We also provide its general form.

英文 key words　Karhunen-Loève subspace, Karhunen-Loève expansion, pattern recognition, data compression

1 Introduction

The Karhunen-Loève (K–L) subspace is a subspace which provides the best approximation for a stochastic signal under the condition that its dimension is fixed [1]–[7]. The K–L subspace has been successfully used for many purposes. In the field of communication, it is used for data compression [2]. In the field of pattern recognition, it is used in the subspace method [4].

Let $\mathcal{H}$ be an N-dimensional Hilbert space. Let $(\cdot,\cdot)$ and $\|\cdot\|$ be the inner product and the norm in $\mathcal{H}$, respectively. Let f be a stochastic variable in $\mathcal{H}$. Let E_f be the ensemble average for f. Let $P_{\mathcal{S}}$ be the orthogonal projection operator onto a subspace $\mathcal{S}$. Finally, let M be an integer such that $M \leq N$.

A subspace $\mathcal{S}$ is said to be the M-dimensional K–L subspace if and only if it maximizes

$$E_f\|P_{\mathcal{S}}f\|^2 \tag{1}$$

under the condition that the dimension of $\mathcal{S}$ is equal to M [6], [7].

For a pair of elements f and g in $\mathcal{H}$, the Schatten product $f \otimes \overline{g}$, which is an operator in $\mathcal{H}$, is defined as

$$(f \otimes \overline{g})h = (h, g)f \tag{2}$$

with any element $h \in \mathcal{H}$ [11]. The correlation operator R with respect to f is defined as

$$R = E_f(f \otimes \overline{f}). \tag{3}$$

Let λ_i be eigenvalues of R such that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N$. Let u_i be the eigenelement of R corresponding to λ_i. We can choose $\{u_i\}_{i=1}^N$ so that it is an orthonormal base in $\mathcal{H}$. If an eigenvalue is degenerate, a corresponding eigenelement is not unique. Let Λ be the set of sets $\{u_i\}_{i=1}^N$ such that u_i is a eigenelement corresponding to λ_i and $\{u_i\}_{i=1}^N$ is an orthonormal base in $\mathcal{H}$. A subspace $\mathcal{S}$ becomes the M-dimensional K–L subspace if and only if $\mathcal{S}$ is spanned by the first M elements u_i $(i = 1, 2, \cdots, M)$ with some set of eigenelements in Λ [6], [7].

The K–L subspace, however, does not consider a noise in communication and a noise and other patterns in pattern recognition. Therefore, its noise suppression is not sufficient in communication and it gives a wrong recognition result for patterns which are similar to each other. In this paper, we shall consider an extra stochastic variable n. This n may be a noise or an extra signal. In this paper, n is called a noise.

Historically, such a problem has been first discussed in [8]. It based on an extension of the evaluation function(1). Let $P_{\mathcal{S},\mathcal{L}}$ be an oblique projection operator onto a subspace $\mathcal{S}$ along a subspace $\mathcal{L}$. We wish to maximize $E_f\|P_{\mathcal{S},\mathcal{L}}f\|^2$, while we wish to minimize $E_n\|P_{\mathcal{S},\mathcal{L}}n\|^2$. One way to satisfy these two demands at the same time is maximaizing the following ratio:

$$\frac{E_f\|P_{\mathcal{S},\mathcal{L}}f\|^2}{E_n\|P_{\mathcal{S},\mathcal{L}}n\|^2}. \tag{4}$$

Indeed, the article[8] solved this maximization problem. However, this definition has the following problems. Since $\mathcal{S}$ is allowed to be any complementary subspace of $\mathcal{L}$ which is determined by this evaluation, eq.(4) is not a good evaluation in the sense that $P_{\mathcal{S},\mathcal{L}}f$ must provide an approximation of f. If no eigenvalue is degenerate, then it follows that $\dim \mathcal{S} = 1$. Furthermore, if no eigenvalue is degenerate, then the problem maximizing eq.(4) under the condition that $\dim \mathcal{S} = M$ $(M \geq 2)$ has no solution. This does not agree with our intuition in that sense.

The K–L subspace can be characterized in a different way from eq.(1) as follows. Let B be a linear operator in $\mathcal{H}$. Consider the following problem that B minimizes

$$E_f\|f - Bf\|^2, \tag{5}$$

under the condition that the dimension of the range of B is equal to M. If the range of R is equal to $\mathcal{H}$, then it follows that

$$B = P_{\mathcal{S}}, \tag{6}$$

with an M-dimensional K–L subspace $\mathcal{S}$. If the range of R is not equal to $\mathcal{H}$, eq.(6) does not necessarily holds. But it always follows that

$$E_f\|Bf\| = E_f\|P_{\mathcal{S}}f\|^2. \tag{7}$$

This characterization leads to a new approach to our problem. In the case of eq.(4), we wish to maximize one part in it and to minimize the other, so that the ratio is used. On the other hand, when we use eq.(5), we wish to minimize both eq.(5) and $E_n\|Bn\|^2$, so that we consider that the sum of them is minimized. Now, we propose a concept of the *relative Karhunen-Loève (K-L) operator*. The relative K–L operator B minimizes the sum of the mean square error between the original signal and the approximated signal and the mean square error caused by the noise:

$$J[B] = E_f\|f - Bf\|^2 + E_n\|Bn\|^2 \tag{8}$$

under the condition that the dimension of the range of B is fixed.

We provide the conditions under which the relative K–L operator exists. We also provide its general form.

In Section 2, we provide a mathematical preliminaries. In Section 3, we provide the relative K-L operator. In Section 4, we prove the result of Section 3.

2 Mathematical preliminaries

The following notations and terminologies are used in this paper.

Let A^* be the adjoint operator of an operator A. Let $\mathcal{R}(A)$ and $\mathcal{N}(A)$ be the range and the null space of an operator A, respectively. For any operator A there exists a unique operator $A^\dagger$ [12], [13] such that

$$AA^\dagger A = A, \tag{9}$$
$$A^\dagger AA^\dagger = A^\dagger, \tag{10}$$
$$(AA^\dagger)^* = AA^\dagger, \tag{11}$$
$$(A^\dagger A)^* = A^\dagger A. \tag{12}$$

The operator $A^\dagger$ is called the Moore-Penrose generalized inverse of A.

A self-adjoint operator A in $\mathcal{H}$ is said to be a nonnegative definite operator if and only if $(Ax, x) \geq 0$ for every $x \in \mathcal{H}$. It is denoted by $A \geq 0$. For any operator $A \geq 0$ there exists a unique operator $A^{1/2}$ such that $A^{1/2} \geq 0$ and

$$A^{1/2}A^{1/2} = A. \tag{13}$$

It follows that

$$(A^{\dagger})^{1/2} = (A^{1/2})^{\dagger}. \tag{14}$$

Let Q be a correlation operator with respect to a noise ensemble which is defined as

$$Q = E_n(n \otimes \overline{n}). \tag{15}$$

Let tr$[A]$ be the trace of an operator A. Let $\|A\|_2$ be the Schmidt norm of an operator A [11]. It follows that

$$\|A\|_2^2 = \mathrm{tr}[AA^*]. \tag{16}$$

For an orthonormal base $\{\phi_i\}_{i=1}^N$, it follows that

$$\|A\|_2^2 = \sum_{i=1}^{N} \|A\phi_i\|^2. \tag{17}$$

Let $\dim \mathcal{S}$ be the dimension of a subspace $\mathcal{S}$. Let $\min(M, N)$ be the smaller value of integers M or N.

3 Relative K–L operator

Definition 1. Let M be an integer such that $M \leq N$. An operator B is said to be a relative Karhunen-Loève (K–L) operator of degree M if and only if B minimizes J in eq.(8) subject to $\dim \mathcal{R}(B) = M$.

In order to obtain the relative K–L operator, we shall introduce some operators and eigenelements.

We define U as

$$U = R + Q. \tag{18}$$

Lemma 5 in [9] yields that

$$\begin{aligned} \mathcal{N}(U) &= \mathcal{N}(R) \cap \mathcal{N}(Q), & (19) \\ \mathcal{R}(U) &= \mathcal{R}(R) + \mathcal{R}(Q). & (20) \end{aligned}$$

We define operators B_1 and C_1 as

$$\begin{aligned} B_1 &= RU^{\dagger}, & (21) \\ C_1 &= R(U^{1/2})^{\dagger}, & (22) \end{aligned}$$

respectively.

Let K be $\dim \mathcal{R}(R)$. Let L be $\dim \mathcal{R}(U)$. Eq.(20) yields that $K \leq L$. Since $U = U^*$ and $\mathcal{R}((U^{1/2})^{\dagger}) = \mathcal{R}(U)$, eq.(19) yields that $\mathcal{R}(C_1) = \mathcal{R}(R)$, so that it follows that

$$\dim \mathcal{R}(C_1) = K. \tag{23}$$

A singular value decomposition of C_1 can be given as

$$C_1 = \sum_{i=1}^{N} \lambda_i (v_i \otimes \overline{u_i}), \tag{24}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N \geq 0$, and $\{u_i\}_{i=1}^N$ and $\{v_i\}_{i=1}^N$ are orthonormal bases in $\mathcal{H}$. It follows that

$$\begin{aligned} C_1 u_i &= \lambda_i v_i, & (25) \\ C_1^* v_i &= \lambda_i u_i. & (26) \end{aligned}$$

Eq.(23) yields that $\lambda_i > 0$ for $i \leq K$ and $\lambda_i = 0$ for $i > K$. When some of λ_i are equal to each other, the singular value decomposition is not unique. Let Λ be the set of sets $\{u_i, v_i\}_{i=1}^N$ such that they satisfy eq.(24) and $\{u_i\}_{i=1}^N$ and $\{v_i\}_{i=1}^N$ are orthonormal bases in $\mathcal{H}$. Now, we shall provide the relative K–L operator of degree M.

Theorem 1. The relative K–L operator of degree M exists if and only if $N - L \geq M - K$. An operator B is a relative K–L operator of degree M if and only if

$$B = B_0 + W(I - UU^{\dagger}), \tag{27}$$

where

$$B_0 = \sum_{i=1}^{M} \lambda_i (v_i \otimes \overline{(U^{1/2})^{\dagger} u_i}) \tag{28}$$

with a set $\{u_i, v_i\}_{i=1}^N$ in Λ and an arbitrary operator W such that $\dim R(B) = M$.

Note that the sum eq.(28) is truncated after M terms.

Since almost all signals f are contained in $\mathcal{R}(R)$ [10], it is sufficient to approximate signals only in $\mathcal{R}(R)$. Hence, it is usually assumed that $\mathcal{R}(B) \subset \mathcal{R}(R)$, so that it follows that $M \leq K$. In this case, it follows that $\dim \mathcal{R}(B_0) = M$, so that the following corollary holds.

Corollary 1. If $M \leq K$, the relative K–L operator of degree M exists. An operator B is a relative K–L operator of degree M if and only if B is given by eq.(27) with a set $\{u_i, v_i\}_{i=1}^N$ in Λ and an arbitrary operator W such that $\mathcal{R}(W) \subset \mathcal{R}(B_0)$.

Corollary 1 implies that if $M \leq K$, we can use B_0 itself as the relative K–L operator of degree M.

The proof of Theorem 1 yields Corollary 2.

Corollary 2. There exists an operator B which minimizes $J[B]$ in eq.(8) subject to $\dim \mathcal{R}(B) \leq M$. An operator B is such operator if and only if B is given by eq.(27) with a set $\{u_i, v_i\}_{i=1}^N$ in Λ and an arbitrary operator W such that $\dim \mathcal{R}(B) \leq M$.

The essential term of the relative K–L operator is B_0. Because $J[B]$ in eq.(8) is independent of W. W is used to make $\dim \mathcal{R}(B)$ be M.

In the case of the relative K–L operator, even if $\mathcal{R}(B)$ and $\mathcal{N}(B)$ are given, we can not construct B. In this case, not subspaces but an operator itself is important.

Assume that $K = N$. Corollaries 1 and 2 yield that when we increase M, then $J[B]$ in eq.(8) of corresponding relative K–L operators decreases. However, in the case of the relative K–L subspace defined by eq.(4), if no eigenvalue is degenerate, a subspace $\mathcal{S}$ whose dimension is equal to one gives the maximum value of eq.(4). Therefore, the definition of the relative K–L operator is more natural in order to get an approximation of f.

Let f be the original signal. Let n be the additive noise. We consider the case that only $g = f + n$ can be observed and f is approximated by Bg where B is an operator such that $\dim \mathcal{R}(B) = M$. The mean square error between f and Bg is given as

$$E_f E_n \|f - Bg\|^2. \tag{29}$$

Assume that f is independent of n and the average of noise $E_n n$ is equal to 0. In this case eq.(29) is equal to (8), so that all discussions above hold for this problem.

4 Proof of Theorem 1

In order to prove Theorem 1, we shall prepare the following Lemma 1.

Lemma 1. Let M be an integer such that $M \leq N$. An operator B minimizes

$$J_1[B] = \|B - A\|_2 \tag{30}$$

subject to $\dim \mathcal{R}(B) \leq M$ if and only if

$$B = \sum_{i=1}^{M} \lambda_i (v_i \otimes \overline{u_i}) \tag{31}$$

with some singular value decomposition of A defined as

$$A = \sum_{i=1}^{N} \lambda_i (v_i \otimes \overline{u_i}), \tag{32}$$

where $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_N \geq 0$.

Proof. Let K be $\dim \mathcal{R}(A)$. K is the number of positive singular values λ_i.

When $M \geq K$, eq.(31) yields that $B = A$, so that Lemma 1 holds. Therefore, we assume that $M < K$ hereafter.

It follows that

$$\|Cx - y\| \geq \|P_{\mathcal{R}(C)} y - y\| \tag{33}$$

for any operator C and any x, y in $\mathcal{H}$. Since $\{u_i\}_{i=1}^{N}$ is an orthonormal base in $\mathcal{H}$, eqs.(17) and (33) yield that for any operator C

$$\begin{aligned} \|C - A\|_2^2 &= \sum_{i=1}^{N} \|(C - A)u_i\|^2 \\ &= \sum_{i=1}^{N} \|Cu_i - \lambda_i v_i\|^2 \\ &\geq \sum_{i=1}^{N} \|P_{\mathcal{R}(C)}(\lambda_i v_i) - \lambda_i v_i\|^2 \\ &= \sum_{i=1}^{K} \lambda_i^2 \|P_{\mathcal{R}(C)} v_i - v_i\|^2. \end{aligned} \tag{34}$$

Since λ_i $(i \leq K)$ are positive, the uniqueness of the orthogonal projection yields the following relation. If an operator B minimizes J_1 subject to $\dim \mathcal{R}(B) = M$, then eq.(34) yields that

$$Bu_i = \lambda_i P_{\mathcal{R}(B)} v_i \tag{35}$$

for all $i \leq K$ since $\{u_i\}_{i=1}^{N}$ is an orthonormal base in $\mathcal{H}$. Therefore, we consider the minimizing problem of

$$J_2[B] = \sum_{i=1}^{K} \lambda_i^2 \|P_{\mathcal{R}(B)} v_i - v_i\|^2 \tag{36}$$

subject to $\dim \mathcal{R}(B) \leq M$. Since $P_{\mathcal{R}(B)}$ is an orthogonal projection operator, it follows that

$$J_2[B] = \sum_{i=1}^{K} \lambda_i^2 - \sum_{i=1}^{K} \lambda_i^2 \|P_{\mathcal{R}(B)} v_i\|^2. \tag{37}$$

Hence, J_2 becomes minimum if and only if

$$J_3[B] = \sum_{i=1}^{K} \lambda_i^2 \|P_{\mathcal{R}(B)} v_i\|^2 \tag{38}$$

is maximum.

Since

$$0 \leq \|P_{\mathcal{R}(B)} v_i\|^2 \leq 1, \tag{39}$$

and

$$\sum_{i=1}^{K} \|P_{\mathcal{R}(B)} v_i\|^2 \leq M, \tag{40}$$

if B maximizes J_3, then it follows that

$$\|P_{\mathcal{R}(B)} v_i\| = \begin{cases} 1 & \text{if } 1 \leq i \leq M \\ 0 & \text{otherwise} \end{cases} \tag{41}$$

with $\{v_i\}_{i=1}^{N}$ a orthonormal base which satisfies a singular value decomposition(32) of A.

Eq.(41) yields that

$$P_{\mathcal{R}(B)} v_i = \begin{cases} v_i & \text{if } 1 \leq i \leq M \\ 0 & \text{otherwise,} \end{cases} \tag{42}$$

so that eq.(35) yields that

$$Bu_i = \begin{cases} \lambda_i v_i & \text{if } 1 \leq i \leq M \\ 0 & \text{otherwise.} \end{cases} \tag{43}$$

Therefore, eq.(31) holds.

We shall prove the converse. Assume that eq.(31) holds. In this case, eq.(41) holds. For any operator C such that $\dim \mathcal{R}(C) = M$ eqs.(34), (37), (39), (40), and (41) yield that

$$\begin{aligned} \|C - A\|_2^2 &\geq \sum_{i=1}^{K} \lambda_i^2 \|P_{\mathcal{R}(C)} v_i - v_i\|^2 \\ &= \sum_{i=1}^{K} \lambda_i^2 - \sum_{i=1}^{K} \lambda_i^2 \|P_{\mathcal{R}(C)} v_i\|^2. \\ &\geq \sum_{i=M}^{K} \lambda_i^2 \\ &= \sum_{i=1}^{K} \lambda_i^2 - \sum_{i=1}^{K} \lambda_i^2 \|P_{\mathcal{R}(B)} v_i\|^2. \\ &\geq \sum_{i=1}^{K} \lambda_i^2 \|P_{\mathcal{R}(B)} v_i - v_i\|^2 \\ &= \|B - A\|_2^2, \end{aligned} \tag{44}$$

so that Lemma 1 holds. □

Now we shall prove Theorem 1.

Proof of Theorem 1.

Eqs.(8), (3), and (15) yield that

$$\begin{aligned} J[B] &= E_f \|f - Bf\|^2 + E_n \|Bn\|^2 \\ &= E_f \mathrm{tr}[(f - Bf) \otimes \overline{(f - Bf)}] \\ &\quad + E_n \mathrm{tr}[Bn \otimes \overline{Bn}] \\ &= \mathrm{tr}[(I - B) E_f (f \otimes \overline{f})(I - B)^*] \\ &\quad + \mathrm{tr}[B E_n (n \otimes \overline{n}) B^*] \\ &= \mathrm{tr}[(I - B) R (I - B)^*] \\ &\quad + \mathrm{tr}[B Q B^*]. \end{aligned} \tag{45}$$

Eqs.(21) and (20) and $R = R^*$ yield that

$$B_1 U = BU^\dagger U = R. \tag{46}$$

Since $R^* = R$ and $U^* = U$, eqs.(45) and (46) yield that

$$\begin{aligned} J[B] &= \mathrm{tr}[BRB^* - BR + RB^* + R + BQB^*] \\ &= \mathrm{tr}[BUB^* - BUB_1^* + B_1UB^* + R]. \end{aligned} \tag{47}$$

We define $J_4[B]$ as

$$J_4[B] = J[B] - J[B_1]. \tag{48}$$

Eq.(47) yields that

$$\begin{aligned} J_4[B] &= \mathrm{tr}[BUB^* - BUB_1^* + B_1UB^* + R] \\ &\quad -\mathrm{tr}[B_1UB_1^* + R] \\ &= \mathrm{tr}[(B - B_1)U(B - B_1)^*]. \end{aligned} \tag{49}$$

Since $U^* = U$, a general formula $(A^*A)^\dagger A^* = A^\dagger$ yields that

$$U^\dagger U^{1/2} = (U^{1/2})^\dagger. \tag{50}$$

Hence, eqs.(21) and (22) yield that

$$B_1 U^{1/2} = RU^\dagger U^{1/2} = R(U^{1/2})^\dagger = C_1,$$

so that

$$B_1 U^{1/2} = C_1 \tag{51}$$

Eqs.(49), (51), and (16) yield that

$$\begin{aligned} J_4[B] &= \mathrm{tr}[(B - B_1)U^{1/2}\{(B - B_1)U^{1/2}\}^*] \\ &= \|[(B - B_1)U^{1/2}\|_2^2 \\ &= \|BU^{1/2} - C_1\|_2^2. \end{aligned} \tag{52}$$

Since $J[B_1]$ is a constant value, J becomes minimum if and only if J_4 is minimum. Since $\dim \mathcal{R}(B) = M$, it follows that

$$\dim \mathcal{R}(BU^{1/2}) \leq M. \tag{53}$$

Since $U^{1/2} = (U^{1/2})^*$, eqs.(22) and (24) yield that

$$\begin{aligned} \mathcal{N}(U^{1/2}) &= \mathcal{N}((U^{1/2})^*) \\ &= \mathcal{N}((U^{1/2})^\dagger) \\ &\subset \mathcal{N}(C_1) \\ &\subset \sum_{i=1}^{M} \lambda_i (v_i \otimes \overline{u_i}). \end{aligned} \tag{54}$$

First, we assume that there exists an operator B which satisfies $\dim \mathcal{R}(B) = M$ and

$$BU^{1/2} = \sum_{i=1}^{M} \lambda_i (v_i \otimes \overline{u_i}) \tag{55}$$

with a set $\{u_i, v_i\}_{i=1}^N$ in Λ. In this case, eqs.(52) and (53) and Lemma 1 yield that the relative K–L operator exists and an operator B is a relative K–L operator if and only if B satisfies eq.(55) with a set $\{u_i, v_i\}_{i=1}^N$ in Λ. Eq.(54) yields that eq.(55) always has a solution and the general solution of eq.(55) is given by eq.(27) with an arbitrary operator W. Therefore, in this case the relative K–L operator exists and an operator B is a relative K–L operator of degree M if and only if B is given by eq.(27) with a set $\{u_i, v_i\}_{i=1}^N$ in Λ and an arbitrary operator W such that $\dim R(B) = M$.

Now, we clarify that the case above holds if and only if $N - L \geq M - K$. Eq.(28) yields that

$$\dim \mathcal{R}(B_0) = \min(K, M). \tag{56}$$

Since $I - UU^\dagger$ is an orthogonal projection operator onto $R(U)^\perp$, it follows that

$$\dim \mathcal{R}(I - UU^\dagger) = N - L. \tag{57}$$

Since $\mathcal{N}(B_0) \supset \mathcal{N}((U^{1/2})^\dagger) = \mathcal{N}(U)$ and $\mathcal{N}(I - UU^\dagger) = \mathcal{N}(U)^\perp$, it follows that

$$\mathcal{N}(B_0)^\perp \perp \mathcal{N}(I - UU^\dagger)^\perp. \tag{58}$$

If eq.(55) holds, eqs.(27), (56), (57), and (58) yield that $\dim \mathcal{R}(B) \leq min(K, M) + N - L$. Therefore, if eq.(55) and $\dim \mathcal{R}(B) = M$ hold, it follows that $N - L \geq M - K$. Conversely, if $N - L \geq M - K$ holds, eqs.(27), (56), (57), and (58) yield that there exists a W such that eq.(55) and $\dim \mathcal{R}(B) = M$ hold.

We assume that $N - L < M - K$. We shall prove that there exists no relative K–L operator on degree M in this case. Suppose that B is a relative K–L operator of degree M. Since

$$B = P_{\mathcal{R}(C_1)}B + (I - P_{\mathcal{R}(C_1)})B, \tag{59}$$

eq.(52) yields that

$$\begin{aligned} J_4[B] &= \|P_{\mathcal{R}(C_1)}BU^{1/2} - C_1\|_2^2 \\ &\quad + \|(I - P_{\mathcal{R}(C_1)})BU^{1/2}\|_2^2. \end{aligned} \tag{60}$$

Since $\dim \mathcal{N}(U^{1/2}) = \dim \mathcal{N}(U) = N - L$, it follows that

$$\dim \mathcal{R}(BU^{1/2}) \geq M - (N - L) > K. \tag{61}$$

Eq.(61) yields that

$$\dim \mathcal{R}((I - P_{\mathcal{R}(C_1)})BU^{1/2}) > 0, \tag{62}$$

so that it follows that

$$\|(I - P_{\mathcal{R}(C_1)})BU^{1/2}\|_2 > 0. \tag{63}$$

Let us define an operator B_2 as

$$B_2 = P_{\mathcal{R}(C_1)}B + \frac{1}{2}(I - P_{\mathcal{R}(C_1)})B. \tag{64}$$

Eqs.(60) and (63) yield that

$$\begin{aligned} J_4[B_2] &= \|P_{\mathcal{R}(C_1)}BU^{1/2} - C_1\|_2^2 \\ &\quad + \frac{1}{2}\|(I - P_{\mathcal{R}(C_1)})BU^{1/2}\|_2^2 \\ &< \|P_{\mathcal{R}(C_1)}BU^{1/2} - C_1\|_2^2 \\ &\quad + \|(I - P_{\mathcal{R}(C_1)})BU^{1/2}\|_2^2 \\ &= J_1[B]. \end{aligned} \tag{65}$$

Eqs.(64) and $(2I - P_{\mathcal{R}(C_1)})B_2 = B$ yield that $\dim \mathcal{R}(B_2) = M$. This contradicts that B is the relative K–L operator. Therefore, if $N - L < M - K$, there exists no relative K–L operator.

This completes the proof. □

5 Conclusion

We proposed the concept of the relative Karhunen-Loève (K–L) operator. That is an operator which minimizes the sum of the mean square error between the original signal and the approximated signal and the mean square error caused by noise under the condition that the dimension of its range is fixed.

We provided a necessary and sufficient condition that the relative K–L operator exists. We also provided its general form.

References

[1] K. Fukunaga: "Introduction to Statistical Pattern Recognition", Academic Press, London(1972).

[2] H. C. Andrews : "Two-dimensional Transforms", in Picture Processing and Digital Filtering, T. S. Huang ed., pp.21–68, Springer-Verlag, Berlin(1979).

[3] P. A. Devijver and J. Kittler : "Pattern Recognition : A Statistical Approach", Prentice Hall, Englewood Cliffs(1982).

[4] E. Oja : "Subspace Methods of Pattern Recognition ", Research Studies Press, Letchworth, Hertfordshire, England(1983).

[5] H. Ogawa and E. Oja : "Projection filter, Wiener filter and Karhunen-Loève subspaces in digital image restoration": J. Math. Anal. Appl., vol. 114, no. 1, pp.37–51(Feb. 1986).

[6] H. Ogawa : "On the subspace which provides the best approximation to a set of patterns", Technical report of IEICE, no. PRU90-27, pp.67–72(Oct. 1990).

[7] H. Ogawa : "Karhunen-Loève subspace", Proc. of the 11th IAPR Inter. Conf. on Pattern Recognition, vol. 2, The Hague, The Netherlands, pp.75–78(Aug. 30–Sept. 3 1992).

[8] Y. Yamashita, Y. Taniguchi, and H. Ogawa : "Relative Karhunen-Loève subspace", Proc. of the 1992 IEICE Spring Conf., Noda, vol. 7, p.263(March 24–27 1992).

[9] Y. Yamashita and H. Ogawa : "Image restoration by averaged projection filter" Trans. IEICE, Part D-II, vol. J74-D-II, no. 2, pp150–157(Feb. 1991).

[10] Y. Yamashita and H. Ogawa : "Properties of averaged projection filter for image restoration", Trans. IEICE, Part D-II, vol. J74-D-II, no. 2, pp.142–149(Feb. 1991).

[11] R. Schatten : "Norm Ideals of Completely Continuous Operators", Springer-Verlag, Berlin(1970).

[12] A. Albert : "Regression and the Moore-Penrose Pseudo-inverse", Academic Press, London(1972).

[13] A. Ben-Israel and T. N. E. Greville : "Generalized Inverses: Theory and Applications", John Wiley & Sons, New York(1974).