Towards theory of generic Principal Component Analysis

Anatoli Torokhti and Shmuel Friedland

Abstract

In this paper, we consider a technique called the generic Principal Component Analysis (PCA) which is based on an extension and rigorous justification of the standard PCA. The generic PCA is treated as the best weighted linear estimator of a given rank under the condition that the associated covariance matrix is singular. As a result, the generic PCA is constructed in terms of the pseudo-inverse matrices that imply a development of the special technique. In particular, we give a solution of the new low-rank matrix approximation problem that provides a basis for the generic PCA. Theoretical aspects of the generic PCA are carefully studied.

1 Introduction

In this paper, we consider an extension and rigorous justification of Principal Component Analysis (PCA) for the case of singular data and weighting matrices used in the PCA structure. Such a technique is here called the generic PCA. Differences from the known results and the innovation of the proposed methodology are specified in Section 3.

The PCA is a procedure of finding the so called principal components of observed data presented by a large random vector, i.e. of components of a smaller vector which preserves principal features of observed data. In particular, this means that the original vector can be reconstructed from the smaller one with the least possible error.

The standard PCA [12] works under a strong assumption on non-singularity of the associated covariance matrix. At the same time, intrinsic data features imply singularity of the covariance matrix that leads to the necessity of exploiting the pseudo-inverse operator in constructing the PCA. Although an approach to a derivation of the PCA in the case of the pseudo-inverse operator has been outlined in [11], the technique associated with the pseudo-inverse operator is not straightforward and requires a more detailed and rigorous analysis. It is shown below (see Theorem 4 and Remark 4 in Section 5.1), that such a technique requires an extension of the known low-rank matrix approximation [4] to the more general cases presented by [5] and Theorem 1 in Section 4.

Observed data is normally corrupted with random noise. Therefore, a procedure of finding the principal components (such a procedure is often called data compression) should be accompanied by filtering. We note that filtering and data compression could be separated. Nevertheless, simultaneous filtering and compression is more effective in the sense of minimizing the associated error (see [30, 33], for example). The generic PCA considered below performs these operations simultaneously. See Section 5.2 in this regard.

Next, many applied problems require a development of weighted estimators. Examples of such problems are determining the electroencephalography (EEG) envelope in neuropsychobiology [7], a statistical analysis of shells in biology [2], data dimensionality reduction in neural systems [3, 23], modelling automobile assembly operation [34], parameter estimation in linear regression [32], filter design [17, 18, 19, 26], state-space modelling [9, 25], array signal processing [31] and channel estimation [20]. For other relevant applications, see, e.g., [12, 24].[1] The proposed generic PCA is the new effective weighted estimator. In this regard, see (4)–(5) in Section 3, and Theorems 2 and 4 in Section 5.1.

The main question addressed in the paper is as follows: What is a constructive representation of the generic PCA and its rigorous theoretical justification? In turn, this implies the following questions. What kind of an extension of the low-rank matrix approximation problem should be used in a derivation of the generic PCA? Is a solution of such a problem unique? If not, in what analytical form can we represent its non-uniqueness? What is a condition that leads to the uniqueness? The answers are given below.

2 Standard Principal Component Analysis (PCA)

By Jolliffe [12], 'Principal component analysis is probably the oldest and best known of the techniques of multivariate analysis.' The PCA was discovered by Pearson [22] in 1901 and then independently developed by Hotelling [10] in 1933, by Karhunen [13] in 1947 and by Loève [16] in 1948. Owing to its versatility in applications, PCA has been extended in many directions (see, in particular, [11, 21, 24, 33] and the corresponding bibliographies). In engineering literature, PCA is normally called the Karhunen-Loève transform.

Note that PCA can be reformulated as a technique which provides the best linear estimator of a given rank for a random vector (see [11, 24]). The error associated with the estimators [11, 12, 21] based on the PCA idea is the smallest in the corresponding class of linear estimators with the same rank.

To represent the PCA, we begin with some notation which will be used here and in the following Sections. Let (Ω, Σ, μ) signify a probability space, where $\Omega = \{\omega\}$ is the set of outcomes, Σ a σ–field of measurable subsets in Ω and $\mu : \Sigma \to [0, 1]$ an associated probability measure on Σ with $\mu(\Omega) = 1$.

Suppose that $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$ and $\mathbf{y} \in L^2(\Omega, \mathbb{R}^n)$ are random vectors such that $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^T$ and $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)^T$ with $\mathbf{x}_i, \mathbf{y}_k \in L^2(\Omega, \mathbb{R})$ for $i = 1, \ldots, m$ and $k = 1, \ldots, n$, respectively. Here, $\mathbf{y}$ is noisy observable data and $\mathbf{x}$ represents unknown random data to be estimated from $\mathbf{y}$. No special relationship between $\mathbf{y}$ and $\mathbf{x}$ is assumed except a covariance matrix formed from $\mathbf{y}$ and $\mathbf{x}$. In particular, $\mathbf{y}$ can be a sum of $\mathbf{x}$ and some additive noise. If hypothetically $\mathbf{y}$ contains no noise then $\mathbf{y} = \mathbf{x}$.

We write

$$\langle \mathbf{x}_i \mathbf{y}_j \rangle = \int_\Omega \mathbf{x}_i(\omega) \mathbf{y}_j(\omega) d\mu(\omega), \quad E_{xy} = \{\langle \mathbf{x}_i \mathbf{y}_j \rangle\}_{i,j=1}^{m,n} \quad \text{and} \quad \|\mathbf{x}\|_E^2 = \int_\Omega \|\mathbf{x}(\omega)\|^2 d\mu(\omega)$$

where $\langle \mathbf{x}_i \mathbf{y}_j \rangle < \infty$ and $\|\mathbf{x}(\omega)\|$ is the Euclidean norm of $\mathbf{x}(\omega)$.

Let the eigendecomposition of E_{xx} be given by $E_{xx} = \sum_{j=1}^{m} \lambda_j u_j u_j^T$, where u_j and λ_j are eigenvectors and corresponding eigenvalues of E_{xx}. Let $\mathcal{A} : L^2(\Omega, \mathbb{R}^n) \to L^2(\Omega, \mathbb{R}^m)$ be a linear operator[2] defined by the matrix $A \in \mathbb{R}^{m \times n}$ so that

$$[\mathcal{A}(\mathbf{y})](\omega) = A[\mathbf{y}(\omega)]. \tag{1}$$

The PCA can be represented in the following way. Given $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$ and E_{xx}, the PCA produces a linear operator $\mathcal{P}_0 : L^2(\Omega, \mathbb{R}^m) \to L^2(\Omega, \mathbb{R}^m)$ of maximum possible rank $r (\leq m)$ that minimizes

$$\mathcal{J}(P) = \|\mathbf{x} - \mathcal{P}(\mathbf{x})\|_E^2$$

[1] Concrete application examples can be found, for example, in [2, 3, 15, 23, 34].

[2] Hereinafter, an operator defined similarly to that by (1) will be denoted with a calligraphic letter.

over all linear operators $\mathcal{P} : L^2(\Omega, \mathbb{R}^m) \to L^2(\Omega, \mathbb{R}^m)$ of the same rank r. Here, $\text{rank}(\mathcal{P}) = \dim \mathcal{P}(L^2(\Omega, \mathbb{R}^m))$.

The matrix P_0, associated with operator $\mathcal{P}_0$, is given by

$$P_0 = U_r U_r^T,$$

where $U_r = [u_1, u_2, \dots, u_r]$. Thus, U_r^T performs a determination of the principal components in the form of a shorter vector in $\mathcal{U}_r^T(\mathbf{x}) \in L^2(\Omega, \mathbb{R}^r)$ and U_r performs a reconstruction of the vector of principal components to $\hat{\mathbf{x}}$ so that $\hat{\mathbf{x}} = \mathcal{P}_0(\mathbf{x})$. The ratio

$$c = \frac{r}{m}, \tag{2}$$

where r is the number of principal components of vector $\mathbf{x}$ and c is often called the compression ratio.

3 Differences from known results. Statement of the problem

While, as we have mentioned above, the standard PCA has been extended in many directions, we consider here the works that are directly concerned with the result derived in this paper.

Scharf [24] presented an extension of the standard PCA as a solution of the problem

$$\min_{\text{rank } P \le r \le m} \|\mathbf{x} - \mathcal{P}(\mathbf{y})\|_E^2 \tag{3}$$

where $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$ and $\mathbf{y} \in L^2(\Omega, \mathbb{R}^m)$. A difference of the PCA extension in [24] from the standard PCA is that $\mathcal{P}$ transforms an arbitrary $\mathbf{y}$, not $\mathbf{x}$. The crucial assumption in [24] is that the covariance matrix $E[\mathbf{y}\mathbf{y}^T]$ is nonsingular.

Yamashita and Ogawa [33] proposed a version of the PCA for the case where $E[\mathbf{y}\mathbf{y}^T]$ is singular and $\mathbf{y} = \mathbf{x} + \mathbf{w}$ with $\mathbf{w}$ an additive noise.

Hua and Liu [11] outlined the *generalized* PCA with a replacement of the inverse of matrix $E[\mathbf{y}\mathbf{y}^T]$ by its pseudo-inverse.

An attractive feature of the methods [11, 33] is that they are constructed in terms of pseudo-inverse matrices (i.e. invertibility of the covariance matrix $E[\mathbf{y}\mathbf{y}^T]$ is not assumed) and therefore, they always exist. Some other known extensions of the PCA work under the condition that $E[\mathbf{y}\mathbf{y}^T]$ is nonsingular, and this restriction can impose certain limitations on the applicability of the method. In many practical situations, the matrix $E[\mathbf{y}\mathbf{y}^T]$ is singular. See, for example, [27, 28] and [29] in this regard.

At the same time, the usage of pseudo-inverse matrices, as it proposed in [11, 33], requires special techniques for its justification and implementation. In Sections 4 and 5 we present a rigorous generalization of the methods [11, 24, 33] in terms of the pseudo-inverse matrices.

The problem we consider is as follows. Let $W_x \in \mathbb{R}^{p\times m}$, $W_F \in \mathbb{R}^{p\times s}$ and $W_y \in \mathbb{R}^{q\times n}$ be weighting matrices.[3] For $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$, $\mathbf{y} \in L^2(\Omega, \mathbb{R}^n)$ and $F \in \mathbb{R}^{s\times q}$, let

$$J(F) = \|\mathcal{W}_x(\mathbf{x}) - \mathcal{W}_F(\mathcal{F}[\mathcal{W}_y(\mathbf{y})])\|_E^2. \tag{4}$$

We wish to find a linear operator $\mathcal{F}^0 : L^2(\Omega, \mathbb{R}^q) \to L^2(\Omega, \mathbb{R}^s)$ such that

$$J(F^0) = \min_{\text{rank } F \le k \le \min\{m,n\}} J(F). \tag{5}$$

We say that $\mathcal{F}^0$ provides the generic PCA.

[3] They are assumed to be known from particular problems such as those mentioned in references in Section 1. For the sake of generality, here, the weighting matrices are assumed to be arbitrary.

Differences from [11, 24, 33] and an innovation are as follows.

First, we give a solution of the new low-rank matrix approximation problem (Theorem 1 in Section 4) that provides a basis for the generic PCA. See Remark 4 in Section 5.1 for more detail.

Second, the problem (5) is formulated in terms of weighting matrices. The motivation for using the weighting matrices W_x, W_y and W_F follows from a number of applied problems mentioned in Section 1. We note that weighted estimators are normally studied when $s = m$, $q = n$, $W_x = W_F$ and $W_y = I$ (see [18], for example). Such a simplified case follows directly from (4)–(5).

The solution to the problem (5) is given in a completed form by Theorem 2. The generic PCA follows from Theorem 2 and is described in Section 5.2.

4 Generic low-rank matrix approximation problem

A solution to the problem (4)–(5) is based on the results presented in this section.

Let $\mathbb{C}^{m\times n}$ be a set of $m \times n$ complex valued matrices, and denote by $\mathcal{R}(m, n, k) \subseteq \mathbb{C}^{m\times n}$ the variety of all $m \times n$ matrices of rank k at most. Fix $A = [a_{ij}]_{i,j=1}^{m,n} \in \mathbb{C}^{m\times n}$. Then $A^* \in \mathbb{C}^{n\times m}$ is the conjugate transpose of A. Let the SVD of A be given by

$$A = U_A \Sigma_A V_A^*, \tag{6}$$

where $U_A \in \mathbb{C}^{m\times m}$ and $V_A \in \mathbb{C}^{n\times n}$ are unitary matrices, $\Sigma_A := \text{diag}(\sigma_1(A), \ldots, \sigma_{\min(m,n)}(A)) \in \mathbb{C}^{m\times n}$ is a generalized diagonal matrix, with the singular values $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq 0$ on the main diagonal.

Let $U_A = [u_1\ u_2\ \ldots u_m]$ and $V_A = [v_1\ v_2\ \ldots v_n]$ be the representations of U and V in terms of their m and n columns, respectively. Let

$$P_{A,L} := \sum_{i=1}^{\text{rank } A} u_i u_i^* \in \mathbb{C}^{m\times m} \quad \text{and} \quad P_{A,R} := \sum_{i=1}^{\text{rank } A} v_i v_i^* \in \mathbb{C}^{n\times n} \tag{7}$$

be the orthogonal projections on the range of A and A^*, correspondingly. Define

$$A_k := (A)_k := \sum_{i=1}^{k} \sigma_i(A) u_i v_i^* = U_{Ak} \Sigma_{Ak} V_{Ak}^* \in \mathbb{C}^{m\times n} \tag{8}$$

for $k = 1, \ldots, \text{rank } A$, where

$$U_{Ak} = [u_1\ u_2\ \ldots u_k], \quad \Sigma_{Ak} = \text{diag}(\sigma_1(A), \ldots, \sigma_k(A)) \quad \text{and} \quad V_{Ak} = [v_1\ v_2\ \ldots v_k]. \tag{9}$$

For $k > \text{rank } A$, we write $A_k := A\ (= A_{\text{rank } A})$. For $1 \leq k < \text{rank } A$, the matrix A_k is uniquely defined if and only if $\sigma_k(A) > \sigma_{k+1}(A)$.

Recall that

$$A^\dagger = (V_A)_{\text{rank } A} (\Sigma_A)^{-1}_{\text{rank } A} (U_A)^*_{\text{rank } A} \tag{10}$$

is the Moore-Penrose generalized inverse.

Henceforth $\|\cdot\|$ designates the Frobenius norm.

In this section, we consider a problem of finding a matrix X_0 such that

$$||A - BX_0C|| = \min_{X\in\mathcal{R}(p,q,k)} ||A - BXC||. \tag{11}$$

Theorem 1 below provides a solution to the problem (11) and is based on the fundamental result in [5] (Theorem 2.1) which is a generalization of the well known Eckart-Young theorem

[4, 8]. The Eckart-Young theorem states that for the case when $m = p$, $q = n$ and $B = C = I$, the solution is given by $X_0 = A_k$, i.e.

$$||A - A_k|| = \min_{X \in \mathcal{R}(m,n,k)} ||A - X||, \quad k = 1, \dots, \min\{m, n\}. \tag{12}$$

Some related references are [6], [17]–[21], [26]–[33].

Theorem 1 *Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{m\times p}$ and $C \in \mathbb{C}^{q\times n}$ be given matrices. Let*

$$K := (I_p - P_{B,R})S \quad \text{and} \quad L := T(I_q - P_{C,L}) \tag{13}$$

where $S \in \mathbb{C}^{p\times p}$ and $T \in \mathbb{C}^{q\times q}$ any matrices, and I_p is the $p \times p$ identity matrix. Then the matrix

$$X_0 := (I_p + K)B^\dagger (P_{B,L} A P_{C,R})_k C^\dagger (I_q + L) \tag{14}$$

is a minimizing matrix for the minimal problem (11). *Any minimizing X_0 has the above form if and only if either*

$$k \geq \text{rank } P_{B,L} A P_{C,R} \tag{15}$$

or

$$1 \leq k < \text{rank } P_{B,L} A P_{C,R} \quad \text{and} \quad \sigma_k(P_{B,L} A P_{C,R}) > \sigma_{k+1}(P_{B,L} A P_{C,R}). \tag{16}$$

Proof. The proof follows a line of reasoning in [5]. We have $||A - BXC|| = ||\widetilde{A} - \Sigma_B \widetilde{X} \Sigma_C||$, where $\widetilde{A} := U_B^* A V_C$ and $\widetilde{X} := V_B^* X U_C$. Matrices X and $\widetilde{X}$ have the same rank and the same Frobenius norm. Thus, it is enough to consider the minimal problem

$$\min_{\widetilde{X} \in \mathcal{R}(p,q,k)} ||\widetilde{A} - \Sigma_B \widetilde{X} \Sigma_C|| \tag{17}$$

in the following sense: we can first find $\widetilde{X}_0$ that minimizes $||\widetilde{A} - \Sigma_B \widetilde{X} \Sigma_C||$ and then find $X_0 = V_B \widetilde{X}_0 U_C^*$ that minimizes $||A - BXC||$. Therefore, matrices B and C can be identified with matrices Σ_B and Σ_C, respectively.

Let $s = \text{rank } B$ and $t = \text{rank } C$. Clearly if B or C is a zero matrix, then $X = \mathbb{O}$ is the solution to the minimal problem (11). Here, $\mathbb{O}$ is the zero matrix. Let us consider the case $1 \leq s, 1 \leq t$. Define $B_1 := \text{diag}(\sigma_1(B), \dots, \sigma_s(B)) \in \mathbb{C}^{s\times s}, C_1 := \text{diag}(\sigma_1(C), \dots, \sigma_t(C)) \in \mathbb{C}^{t\times t}$. Partition $\widetilde{A}$ and $\widetilde{X}$ into four block matrices A_{ij} and X_{ij} with $i, j = 1, 2$ so that $\widetilde{A} = [A_{ij}]_{i,j=1}^2$ and $\widetilde{X} = [X_{ij}]_{i,j=1}^2$, where $A_{11}, X_{11} \in \mathbb{C}^{s\times t}$. Next, observe that $Z := \Sigma_B \tilde{X} \Sigma_C = \begin{bmatrix} Z_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$, where $Z_{11} = B_1 X_{11} C_1$. Since B_1 and C_1 are invertible we deduce $\text{rank } Z = \text{rank } Z_{11} = \text{rank } X_{11} \leq \text{rank } \widetilde{X} \leq k$.

The approximation property of $(A_{11})_k$ yields the inequality $||A_{11} - Z_{11}|| \geq ||A_{11} - (A_{11})_k||$ for any Z_{11} of rank k at most. Thus

$$\widetilde{X}_0 = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \tag{18}$$

where $X_{11} = B_1^{-1}(A_{11})_k C_1^{-1}$, and X_{12}, X_{21}, X_{22} are arbitrary such that $\widetilde{X}_0 \in \mathcal{R}(p, q, k)$, is a solution to the problem (17). Any solution to this problem has such a presentation if and only if the solution $Z_{11} = (A_{11})_k$ is the unique solution to the problem $\min\limits_{Z_{11} \in \mathcal{R}(s,t,k)} ||A_{11} - Z_{11}||$. This happens if either $k \geq \text{rank } A_{11}$ or $1 \leq k < \text{rank } A_{11}$ and $\sigma_k(A_{11}) > \sigma_{k+1}(A_{11})$.

Next, to preserve $\text{rank } \widetilde{X}_0 = k$ we must have

$$X_{12} = X_{11} G_{12}, \quad X_{21} = H_{21} X_{11} \quad \text{and} \quad X_{22} = H_{21} X_{11} G_{12}, \tag{19}$$

where G_{12} and H_{21} are arbitrary matrices. Let us show that $\widetilde{X}_0$ by (18)-(19) can equivalently be represented in the form

$$\widetilde{X}_0 = (I_p + \tilde{K})\Sigma_B^{\dagger}(P_{\Sigma_B,L}\widetilde{A}P_{\Sigma_C,R})_k\Sigma_C^{\dagger}(I_q + \tilde{L}), \tag{20}$$

where $\tilde{K} = (I_p - P_{\Sigma_B,R})\tilde{S}$ and $\tilde{L} = \tilde{T}(I_q - P_{\Sigma_C,L})$ with $\tilde{S}$ and $\tilde{T}$ arbitrary.

First, we observe that $P_{\Sigma_B,R} = \begin{bmatrix} I_s & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$ and $P_{\Sigma_C,L} = \begin{bmatrix} I_t & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$. Partition $\tilde{S}$ and $\tilde{T}$ into four block matrices S_{ij} and T_{ij} with $i,j = 1,2$, respectively, so that $\tilde{S} = [S_{ij}]_{i,j=1}^2$ and $\tilde{T} = [T_{ij}]_{i,j=1}^2$, where $S_{11} \in \mathbb{C}^{s\times s}$ and $T_{11} \in \mathbb{C}^{t\times t}$. Then we have $\tilde{K} = \begin{bmatrix} \mathbb{O} & \mathbb{O} \\ S_{21} & S_{22} \end{bmatrix}$ and $\tilde{L} = \begin{bmatrix} \mathbb{O} & T_{12} \\ \mathbb{O} & T_{22} \end{bmatrix}$. We also have

$$\Sigma_B^{\dagger}(P_{\Sigma_B,L}\widetilde{A}P_{\Sigma_C,R})_k\Sigma_C^{\dagger} = \begin{bmatrix} B_1^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}\begin{bmatrix} A_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}_k\begin{bmatrix} C_1^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} = \begin{bmatrix} X_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \tag{21}$$

because $\begin{bmatrix} (A_{11})_k & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} = \begin{bmatrix} A_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}_k$. Then (20) implies

$$\widetilde{X}_0 = \begin{bmatrix} X_{11} & X_{11}T_{12} \\ S_{21}X_{11} & S_{21}X_{11}T_{12} \end{bmatrix}. \tag{22}$$

In (19), G_{12} and H_{21} are arbitrary, therefore, (20)-(22) implies (18)-(19) with $G_{12} = T_{12}$ and $H_{21} = S_{21}$.

Conversely, on the basis of (21), it is shown that (18)-(19) implies (20) as follows:

$$\begin{aligned}\widetilde{X}_0 = \begin{bmatrix} X_{11} & X_{11}G_{12} \\ H_{21}X_{11} & H_{21}X_{11}G_{12} \end{bmatrix} &= \begin{bmatrix} I_s & \mathbb{O} \\ H_{21} & I_{p-s} + H_{22} \end{bmatrix}\begin{bmatrix} X_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \\ &\quad \times \begin{bmatrix} I_t & G_{12} \\ \mathbb{O} & I_{q-t} + G_{22} \end{bmatrix} \\ &= \begin{bmatrix} I_s & \mathbb{O} \\ H_{21} & I_{p-s} + H_{22} \end{bmatrix}\Sigma_B^{\dagger}(P_{\Sigma_B,L}\widetilde{A}P_{\Sigma_C,R})_k\Sigma_C^{\dagger}\begin{bmatrix} I_t & G_{12} \\ \mathbb{O} & I_{q-t} + G_{22} \end{bmatrix} \end{aligned} \tag{23}$$

with G_{22} and H_{22} arbitrary. Putting $H_{21} = S_{21}$, $H_{22} = S_{22}$, $G_{12} = T_{12}$ and $G_{22} = T_{22}$, we have $\begin{bmatrix} I_s & \mathbb{O} \\ S_{21} & I_{p-s} + S_{22} \end{bmatrix} = I_p + \tilde{K}$ and $\begin{bmatrix} I_t & T_{12} \\ \mathbb{O} & I_{q-t} + T_{22} \end{bmatrix} = I_q + \tilde{L}$, and then (20) follows.

Thus, if we denote $Z := \Sigma_B^{\dagger}(P_{\Sigma_B,L}\widetilde{A}P_{\Sigma_C,R})_k\Sigma_C^{\dagger}$ then

$$X_0 = V_B\widetilde{X}_0U_C^* = V_BZU_C^* + V_BZ\tilde{L}U_C^* + V_B\tilde{K}ZU_C^* + V_B\tilde{K}Z\tilde{L}U_C^*. \tag{24}$$

Here, $V_BZU_C^* = \bar{X}$, where $\bar{X} = B^{\dagger}(P_{B,L}AP_{C,R})_kC^{\dagger}$, and

$$V_BZ\tilde{L}U_C^* = V_BZU_C^*U_C\tilde{L}U_C^* = \bar{X}L,$$

where

$$U_C\tilde{L}U_C^* = U_C\tilde{T}U_C^*U_C(I_q - P_{\Sigma_C,L})U_C^* = L.$$

Similarly,

$$V_B\tilde{K}ZU_C^* = K\bar{X} \quad \text{and} \quad V_B\tilde{K}Z\tilde{L}U_C^* = K\bar{X}L.$$

Therefore, (24) implies (14).

The representation (14) is unique if either (15) or (16) is true. ■

Corollary 1 *If $p = m$, $q = n$ and B, C are non-singular then the solution to* (11) *is unique and given by $X_0 = B^{-1}A_kC^{-1}$.*

Proof. Under the conditions of the Corollary, rank (BXC) = rank X. In this case, also $P_{B,L} = I_m$, $P_{C,R} = I_n$ and $K = L = \mathbb{O}$. Then $X_0 = B^{-1}A_kC^{-1}$ follows from (14). ■

5 Generic PCA

5.1 Solution to the problem (4)–(5)

Now, we are in the position to give a solution to the problem (4)–(5). To formulate and prove our main result in Theorem 2 below, we need the following notation:

$$\tilde{A} = W_x E_{xy} (E_{yy}^{1/2})^\dagger, \quad \tilde{B} = W_F \quad \text{and} \quad \tilde{C} = W_y E_{yy}^{1/2}. \tag{25}$$

Theorem 2 *Let*

$$\tilde{K} := (I_s - P_{\tilde{B},R})\tilde{S} \quad \text{and} \quad \tilde{L} := \tilde{T}(I_n - P_{\tilde{C},L}) \tag{26}$$

where $\tilde{S} \in \mathbb{R}^{s\times s}$ and $\tilde{T} \in \mathbb{R}^{n\times n}$ any matrices. Then the matrix

$$F^0 := (I_s + \tilde{K})\tilde{B}^\dagger (P_{\tilde{B},L}\tilde{A}P_{\tilde{C},R})_k \tilde{C}^\dagger (I_n + \tilde{L}) \tag{27}$$

provides the generic PCA, i.e. is a minimizing matrix for the problem (4)–(5). Any generic PCA F^0 has the above form if and only if either

$$k \geq \text{rank}\,(P_{\tilde{B},L}\tilde{A}P_{\tilde{C},R}) \tag{28}$$

or

$$1 \leq k < \text{rank}\,(P_{\tilde{B},L}\tilde{A}P_{\tilde{C},R}) \quad \text{and} \quad \sigma_k(P_{\tilde{B},L}\tilde{A}P_{\tilde{C},R}) > \sigma_{k+1}(P_{\tilde{B},L}\tilde{A}P_{\tilde{C},R}). \tag{29}$$

Proof. Let us denote $\Delta_{xy} = \mathcal{W}_x(\mathbf{x}) - \mathcal{W}_F\{\mathcal{F}[\mathcal{W}_y(\mathbf{y})]\}$. We have

$$\begin{aligned}
\|\Delta_{xy}\|_E^2 &= \text{tr}\{E[\Delta_{xy}\Delta_{xy}^T]\}\\
&= \text{tr}\{W_x E_{xx} W_x^T - W_x E_{xy} W_y^T F^T W_F^T - W_F F W_y E_{yx} W_x^T\\
&\qquad\qquad + W_F F W_y E_{yy} W_y^T F^T W_F^T\}\\
&= \|W_x E_{xx}^{1/2}\|^2 - \|W_x E_{xy}(E_{yy}^{1/2})^\dagger\|^2 + \|W_x E_{xy}(E_{yy}^{1/2})^\dagger - W_F F W_y E_{yy}^{1/2}\|^2 \qquad (30)
\end{aligned}$$

because $E_{yy}^\dagger E_{yy}^{1/2} = (E_{yy}^{1/2})^\dagger$ and $E_{xy}E_{yy}^\dagger E_{yy} = E_{xy}$. The latter expression is true by Lemma 2 in [27].

In (30), the only term that depends on F is $\|W_x E_{xy}(E_{yy}^{1/2})^\dagger - W_F F W_y E_{yy}^{1/2}\|^2$. In notation (25), this term is represented as

$$\|W_x E_{xy}(E_{yy}^{1/2})^\dagger - W_F F W_y E_{yy}^{1/2}\|^2 = \|\tilde{A} - \tilde{B}F\tilde{C}\|^2. \tag{31}$$

Let $R(m,n,k) \subseteq \mathbb{R}^{m\times n}$ be the variety of all $m \times n$ matrices of rank k at most. It follows from Theorem 1 that a solution to the minimal problem

$$\min_{F\in R(m,n,k)} \|\tilde{A} - \tilde{B}F\tilde{C}\|^2$$

is given by (27)–(29). ■

Remark 1 *It follows from (27)–(29) that, for arbitrary W_x, W_F and W_y, the solution to the problem (4)–(5) is still not unique even if E_{yy} is non-singular. In such a case, $\tilde{A} = W_x E_{xy}(E_{yy}^{1/2})^{-1}$. If $p = m$, $q = n$ and $\tilde{B}$, $\tilde{C}$ are non-singular then the solution to the problem (4)–(5) is unique and is given by $F^0 = \tilde{B}^{-1}\tilde{A}_k\tilde{C}^{-1}$. This is true by Corollary 1.*

5.2 Structure of generic PCA

By Theorem 2, $\mathcal{F}^0$ determined by (26)–(29) provides the generic Principal Component Analysis.

For $G = P_{\tilde{B},L}\tilde{A}P_{\tilde{C},R}$, we write $G = U_G\Sigma_G V_G^T$ and $G_k = U_{Gk}\Sigma_{Gk}V_{Gk}^T$ where $U_G\Sigma_G V_G^T$ and $U_{Gk}\Sigma_{Gk}V_{Gk}^T$ are the SVD and the truncated SVD defined similarly to (6) and (8), respectively.

Compression of vector $\mathbf{y}$ (in fact, the simultaneous filtering and compression of data $\mathbf{y}$) by the generic PCA is provided by the matrix $V_{Gk}^T\tilde{C}^\dagger(I_n+\tilde{L})$ or by the matrix $\Sigma_{Gk}V_{Gk}^T\tilde{C}^\dagger(I_n+\tilde{L})$. Reconstruction of the compressed vector is performed by the matrix $(I_s+\tilde{K})B^\dagger U_{Gk}\Sigma_{Gk}$ or by the matrix $(I_s+\tilde{K})B^\dagger U_{Gk}$, respectively. *The compression ratio* is given by

$$c = \frac{k}{m}$$

where k is the number of principal components, i.e. the number of components in the compressed vector $x^{(1)} = V_{Gk}^T\tilde{C}^\dagger y$ or $x^{(2)} = \Sigma_{Gk}V_{Gk}^T\tilde{C}^\dagger y$ with $x^{(1)}, x^{(2)} \in \mathbb{R}^k$.

As we have pointed out above, the generic PCA always exists since $\mathcal{F}^0$ is constructed from the pseudo-inverse matrices.

5.3 Description of algorithm

It follows from the above that the numerical realization of the generic PCA consists of computation of its components presented in Theorem 2 and Section 5.2.

The device of numerical realization for the generic PCA is summarized as follows.

Initial parameters: $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$, $\mathbf{y} \in L^2(\Omega, \mathbb{R}^n)$, $p \in \mathbb{N}$.
Final parameters: $\check{x}^{(1)}$, $\check{x}^{(2)}$, $\hat{x}^{(1)}$, $\hat{x}^{(2)}$.

Algorithm:
Compute:

$$\tilde{A} := W_x E_{xy}(E_{yy}^{1/2})^\dagger; \quad \tilde{B} := W_F; \quad \tilde{C} := W_y E_{yy}^{1/2};$$

$$\tilde{K} := (I_s - P_{\tilde{B},R})\tilde{S}; \quad \tilde{L} := \tilde{T}(I_n - P_{\tilde{C},L});$$

% The truncated SVD for $G = P_{\tilde{B},L}\tilde{A}P_{\tilde{C},R}$ defined similarly to (6) and (8):

$$G_k := U_{Gk}\Sigma_{Gk}V_{Gk}^T;$$

% Compression of data $\mathbf{y}$:

$$\check{x}^{(1)} := V_{Gk}^T\tilde{C}^\dagger(I_n+\tilde{L})y; \quad \text{or} \quad \check{x}^{(2)} := \Sigma_{Gk}V_{Gk}^T\tilde{C}^\dagger(I_n+\tilde{L})y;$$

% Reconstruction of the compressed data:

$$\hat{x}^{(1)} = (I_s+\tilde{K})\tilde{B}^\dagger U_{Gk}\Sigma_{Gk}\check{x}^{(1)}; \quad \text{or} \quad \hat{x}^{(2)} = (I_s+\tilde{K})\tilde{B}^\dagger U_{Gk}\check{x}^{(2)}\ ;$$

end.

Remark 2 *Here, as has been mentioned in Section* 5.2, *the compressed data can be represented either by* $\check{x}^{(1)}$ *or by* $\check{x}^{(2)}$. *Consequently, the reconstructed data can be represented either by* $\hat{x}^{(1)}$ *or by* $\hat{x}^{(2)}$, *respectively.*

5.4 The minimum norm generic PCA

The generic PCA presented by (26)–(29) depends on arbitrary matrices $\tilde{S}$ and $\tilde{T}$ and therefore, it is not unique. This implies a natural question: What kind of condition should be imposed on the statement of the problem (4)–(5) and the solution (26)–(29) to make it unique?

The answer follows from imposing an additional constraint of the minimum norm for F^0 and is based on the main result in [5].

Theorem 3 [5] *The minimal norm solution to the problem* (11) *follows when* $X_{ij} = \mathbb{O}$ *in* (18) *for* $(i,j) \neq (1,1)$ *and it is given by*

$$X_0 = B^\dagger (P_{B,L} A P_{C,R})_k C^\dagger. \tag{32}$$

The solution is unique if and only if either (15) *or* (16) *is true.*

A related solution to the problem (4)–(5) is as follows.

Corollary 2 *The solution to the problem* (4)–(5) *having the minimal* $||F^0||$ *is given by*

$$F^0 = \tilde{B}^\dagger (P_{\tilde{B},L} \tilde{A} P_{\tilde{C},R})_k \tilde{C}^\dagger, \tag{33}$$

and it is unique if and only if either

$$k \geq \text{rank } (P_{\tilde{B},L} \tilde{A} P_{\tilde{C},R}) \tag{34}$$

or

$$1 \leq k < \text{rank } (P_{\tilde{B},L} \tilde{A} P_{\tilde{C},R}) \quad \textit{and} \quad \sigma_k(P_{\tilde{B},L} \tilde{A} P_{\tilde{C},R}) > \sigma_{k+1}(P_{\tilde{B},L} \tilde{A} P_{\tilde{C},R}). \tag{35}$$

Proof. The proof follows when Theorem 3 [5] is applied to the term (31) in (30). ■

5.5 A particular case: generalized PCA

Now, we wish to represent a solution of the particular case of the problem (4)–(5) when $q = n$, $s = p = m$ and $W_x = W_F = W_y = I$, i.e. when the problem is to find $\bar{F}^0$ such that

$$J(\bar{F}^0) = \min_{\text{rank } F \leq k \leq \min\{m,n\}} J(F) \tag{36}$$

where $J(F)$ has the form

$$J(F) = \|\mathbf{x} - \mathcal{F}(\mathbf{y})\|_E^2. \tag{37}$$

The problem (36)–(37) relates to the generalized PCA considered, for example, in [11, 33]. A particular associated feature is that the generalized PCA is constructed in terms of the pseudo-inverse matrix, i.e. it always exists. Here, we show that the solution to the problem (36)–(37) is given by (40) (see Theorem 4 below), and its related analysis requires Theorem 1 and the following Lemma 1.

Lemma 1 *Let* $\bar{A} = E_{xy} E_{yy}^{1/2\dagger}$ *and* $\bar{A}_k$ *still denote the truncated SVD defined similarly to* (8). *Then*

$$\bar{A}_k (E_{yy}^{1/2})^\dagger E_{yy}^{1/2} = \bar{A}_k. \tag{38}$$

Proof. As an extension of the technique presented in proving Lemmas 1 and 2 in [27], it can be shown that for any matrices $Q_1, Q_2 \in \mathbb{R}^{m\times n}$,

$$\mathcal{N}(Q_1) \subseteq \mathcal{N}(Q_2) \quad \Rightarrow \quad Q_2(I - Q_1^\dagger Q_1) = \mathbb{O}, \tag{39}$$

where $\mathcal{N}(Q_i)$ is the null space of Q_i for $i = 1, 2$. In regard of the equation under consideration, $\mathcal{N}([E_{yy}^{1/2}]^\dagger) \subseteq \mathcal{N}(E_{xy}[E_{yy}^{1/2}]^\dagger)$. The definition of $\bar{A}_k$ implies that

$$\mathcal{N}(E_{xy}[E_{yy}^{1/2}]^\dagger) \subseteq \mathcal{N}(\bar{A}_k) \quad \text{and} \quad \mathcal{N}([E_{yy}^{1/2}]^\dagger) \subseteq \mathcal{N}(\bar{A}_k).$$

On the basis of (39), the latter implies

$$\bar{A}_k[I - (E_{yy}^{1/2})^\dagger E_{yy}^{1/2}] = \mathbb{O},$$

i.e. (38) is true. ∎

Now, we are in the position to give a rigorous justification of the generalized PCA. In this regard, also see Remark 4 below.

Theorem 4 *Let $\bar{A} = E_{xy}E_{yy}^{1/2\dagger}$ and $\bar{C} = E_{yy}^{1/2}$, and let $\bar{L} = \bar{T}(I_n - P_{\bar{C},L})$ where $\bar{T} \in \mathbb{R}^{n\times n}$ any matrix. Then the matrix*

$$\bar{F}^0 = \bar{A}_k\bar{C}^\dagger(I_n + \bar{L}) \tag{40}$$

provides the generalized PCA, i.e. is a minimizing matrix for the problem (36)–(37)*. The error associated with the generalized PCA is given by*

$$\|\mathbf{x} - \bar{\mathcal{F}}^0(\mathbf{y})\|_E^2 = \|E_{xx}^{1/2}\|^2 - \sum_{j=1}^{k} \sigma_j^2(\bar{A}). \tag{41}$$

Proof. Similarly to (30),

$$\|\mathbf{x} - \mathcal{F}(\mathbf{y})\|_E^2 = \|E_{xx}^{1/2}\|^2 - \|\bar{A}\|^2 + \|\bar{A} - F\bar{C}\|^2. \tag{42}$$

Then (40) follows when Theorem 1 is applied to $\|\bar{A} - F\bar{C}\|^2$. We note that $\bar{A}P_{\bar{C},R} = \bar{A}$.

Next, let rank $\bar{A} = \ell$. Then on the basis of Lemma 1 and the relation $P_{\bar{C},L}\bar{C} = \bar{C}$, we have $\|E_{xy}(E_{yy}^{1/2})^\dagger - \bar{F}^0E_{yy}^{1/2}\|^2 = \|E_{xy}(E_{yy}^{1/2})^\dagger - \bar{A}_k\|^2 = \sum_{j=k+1}^{l}\sigma_j^2(\bar{A})$. Since $\|E_{xy}(E_{yy}^{1/2})^\dagger\|^2 = \sum_{j=1}^{l}\sigma_j^2(\bar{A})$, then (42) implies (41). ∎

We note that the generalized PCA given by (40) is not unique because $\bar{F}^0$ depends on the arbitrary matrix $\bar{T}$.

Remark 3 *The expression* (41) *justifies a natural observation that the accuracy of the vector* $\mathbf{x}$ *estimation increases if k increases.*

Corollary 3 *The minimum norm generalized PCA $\tilde{F}^0$ is unique and it is given by*

$$\tilde{F}^0 = (E_{xy}E_{yy}^{1/2\dagger})_k E_{yy}^{1/2\dagger}. \tag{43}$$

The error associated with $\tilde{F}^0$ is still presented by (41)*.*

Proof. The proof of (43) follows from the proof of Corollary 2 when $\tilde{B}$ is the identity matrix, $\tilde{A} = \bar{A}$ and $\tilde{C} = \bar{C}$. The error representation follows directly from the proof of Theorem 4. ∎

Remark 4 *In references* [11, 33] *where the solution to the problem* (36)–(37) *in the form* (43) *has been outlined, it is proposed to determine rank-constrained minimum of the term* (42) *from the Eckart-Young theorem* (*see* [4, 8] *and* (12)). *Nevertheless, the Eckart-Young theorem can be applied to the term* (42) *only if* $\text{rank}\,(FE_{yy}^{1/2}) = \text{rank}\,F$. *The latter is true if* $E_{yy}^{1/2}$ *is nonsingular which is not a case. Thus, Theorem* 1 *is needed for establishing the solution. Besides, the justification of the related error requires Lemma* 1 *above.*

Remark 5 *To make the solution of the problem* (36)–(37) *unique, an alternative condition could seemingly be a minimization of* $\|\bar{\mathcal{F}}^0(\mathbf{y})\|_E^2$ *with* $\bar{\mathcal{F}}^0$ *given by* (40). *Nevertheless, a minimization of* $\|\bar{\mathcal{F}}^0(\mathbf{y})\|_E^2$ *does not affect the uniqueness issue because* $\|\bar{\mathcal{F}}^0(\mathbf{y})\|_E^2$ *does not depend on an arbitrary matrix* $\bar{T}$. *Indeed, we have*

$$\begin{aligned}
\|\bar{\mathcal{F}}^0(\mathbf{y})\|_E^2 &= \text{tr}\; E[\bar{A}_k\bar{C}_k^\dagger(I_n+\bar{L})\mathbf{y}\mathbf{y}^T(I_n+\bar{L}^T)\bar{C}_k^{\dagger T}\bar{A}_k^T] \\
&= \text{tr}\;[\bar{A}_k\bar{C}_k^\dagger E_{yy}\bar{C}_k^{\dagger T}\bar{A}_k^T + \bar{A}_k\bar{C}_k^\dagger L^T\bar{C}_k^{\dagger T}\bar{A}_k^T + \bar{A}_k\bar{C}_k^\dagger L\bar{C}_k^{\dagger T}\bar{A}_k^T + \bar{A}_k\bar{C}_k^\dagger LL^T\bar{C}_k^{\dagger T}\bar{A}_k^T] \\
&= \|\bar{A}_k(E_{yy}^{1/2})^\dagger E_{yy}^{1/2}\|^2 = \|\bar{A}_k\|^2
\end{aligned}$$

where $\bar{C}_k^\dagger L^T = \mathbb{O}$ *because of* $\bar{C}_k^\dagger P_{\bar{C},L}^T = \bar{C}_k^\dagger$.

References

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

[2] K.I. Diamantaras and S.Y. Kung, *Principal Component Neural Networks. Theory and Applications,* John Wiley & Sons, Inc., New York, 1996.

[3] K.-L. Du and M.N.S. Swamy, *Neural Networks in a Softcomputing Framework,* Springer, London, 2006.

[4] C. Eckart and G. Young, The approximation of one matrix by another of lower rank, *Psychometrika,* 1, pp. 211-218, 1936.

[5] S. Friedland and A. P. Torokhti, Generalized rank-constrained matrix approximations, *SIAM J. Matrix Anal. Appl.,* 29, issue 2, pp. 656-659, 2007.

[6] A. Frieze, R. Kannan, and S. Vempala, Fast Monte-Carlo Algorithms for Finding Low-Rank Approximations, *Journal of the ACM,* 51, No. 6, 2004.

[7] V. Gimeno, Obtaining the EEG envelope in real time: a practical method based on homomorphic filtering, *Neuropsychobiol.,* 18, pp. 110–112, 1987.

[8] G.H. Golub and C.F. Van Loan, *Matrix Computation,* Johns Hopkins Univ. Press, 3rd Ed., 1996.

[9] T. Gustafsson and B.D. Rao, Statistical analysis of subspace-based estimation of reduced-rank linear regression, *IEEE Trans. Signal Processing,* vol. 50, no. 1, pp. 151-159, 2002.

[10] H. Hotelling, Analysis of a Complex of Statistical Variables into Principal Components, *J. Educ. Psychol.,* 24, pp. 417-441 and 498-520, 1933.

[11] Y. Hua and W. Q. Liu, Generalized Karhunen–Loève transform, *IEEE Signal Proc. Lett.* vol. 5 (6), 141–142, 1998.

[12] I.T. Jolliffe, *Principal Component Analysis,* Springer Verlag, New York, 1986.

[13] K. Karhunen, Über Lineare Methoden in der Wahrscheinlichkeitsrechnung, *Ann. Acad. Sci. Fennicae,* Ser. A137, 1947.

[14] T. G. Kolda, A Counterexample to the Possibility of an Extension of the Eckart-Young low-rank Approximation Theorem for the Orthogonal Rank Tensor Decomposition, *SIAM J. Matrix Anal. Appl.,* 24, No. 3, pp. 762-767, 2003.

[15] D.N. Lawley and A.E. Maxwell, *Factor Analysis as a Statistical Method,* Butterworth & Co Ltd, London, 1971.

[16] M. Loève, Fonctions aléatoires de second ordre, In *P. Lévy, Processus Stochastiques et Mouvement Brownien,* Hermann, Paris, 1948.

[17] W.-S. Lu, P.-H. Wang, and A. Antoniou, Design of two-dimensional digital filters using singular-value decomposition and balanced approximation method, *IEEE Trans. Signal Process.,* vool. 39, 2253–2262, 1991.

[18] W.-S. Lu, S.-C. Pei and P.-H. Wang, Weighted Low-Rank Approximation of General Complex Matrices and Its Application in the Design of 2-D Digital Filters, *IEEE Trans. on Circuits and Systems-I: Fundamental Theory & Appl.,* 44, No. 7, pp. 650-655, 1997.

[19] J.H. Manton and Y. Hua, Convolutive reduced rank Wiener filtering, *Proc. of ICASSP'01,* vol. 6, pp. 4001-4004, 2001.

[20] M. Nicoli and O. Simeone, Multislot estimation of fast-varying spaace-time communication channels, *IEEE Trans. Signal Process.*, vol. 511, no. 5, pp. 1184-1195, 2003.

[21] F. A. Ocaña, A. M. Aguilera, and M. J. Valderrama, Functional Principal Componenets Analysis by Choice of Norm, *J. Multivariate Analysis,* 71, pp. 262 -276, 1999.

[22] K. Pearson, On lines and planes of closest fit to systems of points in space. *Phil. Mag.,* 6 No. 2, pp. 559 - 572, 1901.

[23] J.C. Principe, N.R. Eulano and W.C. Lefebvre, *Neural and Adaptive Systems. Fundamentals through Simulations,* John Wiley & Sons, Inc., New York, 2000.

[24] L. L. Scharf, The SVD and reduced rank signal processing, *Signal Processing*, vol. 25, pp. 113 - 133, 1991.

[25] P. Stoica and M. Jansson, Mimo system identification: State-space and subspace approximation versus transfer function and instrumental variables, *IEEE Trans. Signal Process.,* vol. 48, no. 11, pp. 3087-3099, 2000.

[26] S. Treitel and J.L. Shanks, The design of multistage separable planar filters, *IEEE Trans. Geosci. Electron* GE-9, 10–27, 1971.

[27] A. Torokhti and P. Howlett, Optimal Fixed Rank Transform of the Second Degree, *IEEE Trans. on Circuits & Systems. Part II, Analog & Digital Signal Processing,* vol. 48, pp. 309–315, 2001.

[28] A. Torokhti and P. Howlett, Constructing Fixed Rank Optimal Estimators with Method of Recurrent Best Approximations, *J. Multivariate Analysis,* 86, 2, pp. 293-309, 2003.

[29] A. Torokhti and P. Howlett, Optimal Transform Formed by a Combination of Nonlinear Operators: The Case of Data Dimensionality Reduction, *IEEE Trans. Signal Process.*, vol. 54, 4, pp. 1431-1444, 2006.

[30] A. Torokhti and P. Howlett, *Computational Methods for Modelling of Nonlinear Systems,* Elsevier, 2007.

[31] M. Viberg, P. Stoica and B. Ottersen, Maximum likelihood array processing in spatially correlated noise fields using parameterized signals, *IEEE Trans. Signal Process.,* vol. 45, no. 4, pp. 996-1064, 1997.

[32] K. Werner and M. Jansson, Reduced rank linear regression and weighted low rank approximation, *IEEE Trans. Signal Processing,* vol. 54, no. 6, pp. 2063-2075, 2006.

[33] Y. Yamashita and H. Ogawa, Relative Karhunen–Loève transform, *IEEE Trans. Signal Process.,* vol 44, pp. 371–378, 1996.

[34] K. Yang and J. Trewn, *Multyvariate Statistical Methods in Quality Management,* McGraw-Hill, New York, 2004.