

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Data Compression and De-compression by Causal Filters with Variable Finite Memory

A. Torokhti and S. Miklavcic  
University of South Australia  
Australia

## 1. Introduction

A study of data compression methods is motivated by the necessity to reduce expenditures incurred with the transmission, processing and storage of large data arrays (Scharf., 1991; Hua et al., 2001). Such methods have also been applied successfully to the solution of problems related, e.g., to clustering (Fukunaga, 1990), feature selection (Jolliffe, 1986; Gimeno, 1987), forecasting (Kim et al., 2005; Common and Golub, 1990) and estimating the medium from transmission data (Kraut et al., 2004).

Data compression techniques are often performed on the basis of the Karhunen-Loève transform (KLT)<sup>1</sup>, which is closely related to the Principal Component Analysis (PCA). A basic theory for the KLT-PCA can be found, for example, in (Scharf., 1991; Jolliffe, 1986; Torokhti and Howlett, 2007). In short, the KLT-PCA produces a linear operator of a given rank that minimizes an associated error over all linear operators of the same rank. In a so-called standard KLT-PCA (e.g., presented in (Jolliffe, 1986)), an observable signal and a reference signals are the same. In other words, the standard KLT-PCA provides data compression only and no noise filtering.

Scharf (Scharf., 1991; Scharf, 1991) presented an extension of the PCA-KLT<sup>2</sup> for the case when an observable signal  $\mathbf{y}$  and a reference signal  $\mathbf{x}$  are different and no explicit analytical representation of  $\mathbf{y}$  in terms of  $\mathbf{x}$  is known. In particular,  $\mathbf{y}$  can be a noisy version of  $\mathbf{x}$ . The method (Scharf., 1991; Scharf, 1991) assumes that covariance matrix  $E_{yy}$ , formed from  $\mathbf{y}$ , is nonsingular. Yamashita and Ogawa (Yamashita and Ogawa, 1996) proposed and justified a version of the PCA-KLT for the case where the covariance matrix  $E_{yy}$  may be singular and  $\mathbf{y} = \mathbf{x} + \mathbf{w}$  with  $\mathbf{w}$  an additive noise. Hua and Liu (Hua and Liu, 1998) considered an extension of the PCA-KLT to the case when  $\mathbf{y}$  and  $\mathbf{x}$  are different as in (Scharf., 1991; Scharf, 1991), to guarantee its existence when  $E_{yy}$  is singular. Torokhti and Friedland (Torokhti and Friedland, 2009) studied a weighted version of the the PCA-KLT. Torokhti and Howlett (Torokhti and Howlett, 2006; 2009) extended the PCA-KLT to the case of optimal non-linear data compression. Torokhti and Manton (Torokhti and Manton, 2009) further advanced results in (Torokhti and Friedland, 2009; Torokhti and Howlett, 2006; 2009) to the so-called generic weighted filtering of stochastic signals. Advanced computational aspects of the PCA-KLT were provided, in particular, by Hua, Nikpour and Stoica (Hua et al., 2001), Hua and Nikpour (Hua and Nikpour, 1999), Stoica and Viberg (Stoica and Viberg, 1996), and Zhang and Golub (Zhang and Golub,

<sup>1</sup>The KLT is also known as Hotelling Transform and Eigenvector Transform.

<sup>2</sup>List of references related to the PCA-KLT is very long. For example, a Google search for 'Karhunen-Loève transform and Principal Component Analysis' gives 9230 items. Here, we mention only the most relevant references to the problem under consideration.

2001). Other relevant references can be found, e.g. in the bibliographies of the books by Scharf (Scharf., 1991), and Torokhti and Howlett (Torokhti and Howlett, 2007).

While the topics of data compression have been intensively studied (in particular, in the references mentioned above), a number of related fundamental questions remain open. One of them concerns real-time data processing. In this paper, the real-time aspect of the data compression problem is the dominant motivation for considering specific restrictions associated with causality and memory. Similar observations, but not in context of data compression, motivated studies of Winer-like filtering in works by Fomin and Ruzhansky (Fomin and Ruzhansky, 2000), Torokhti and Howlett (Torokhti and Howlett, 2006), and Howlett, Torokhti and Pearce (Howlett et al., 2007).

We note that conditions of causality and memory make the problem very specific and difficult. To the best of our knowledge, such a problem is considered here for the first time. In Section 3.2, we provide a new approach to the problem solution and give an analysis of the associated error. In more detail, motivations to consider the problem are as follows.

FIRST MOTIVATION: CAUSALITY AND MEMORY. Data compression techniques mainly consist of two operations, data compression itself and a subsequent data de-compression (or reconstruction). In real time, the compressor and de-compressor are causal and may be performed with a memory.

A causality constraint follows from the observation that in practice, the present value of the output of a filter<sup>3</sup> is not affected by future values of the input (De Santis, 1976). To determine the output signal at time  $t_k$  with  $q_k = 1, \dots, k$ , to be defined below, the *causal filter* should “remember” the input signal up to time  $t_k$ , i.e. at times  $t_k, t_{k-1}, \dots, t_1$ .

A memory constraint is motivated as follows. The output of the compressor and/or de-compressor at time  $t_k$  with  $k = 0, 1, \dots, m$ , may only be determined from a ‘fragment’ of the input defined at times  $t_k, t_{k-1}, \dots, t_{k-(q_k-1)}$  with  $q_k = 1, \dots, k$ . In other words, compressor and de-compressor should ‘remember’ that fragment of the input. The ‘length’ of the fragment for a given  $k$ , i.e. the number  $q_k$ , is called a local memory. The local memory could be different for different  $k$ , therefore, we also say that  $q_k$  is a local *variable* memory. A collection of the local memories,  $\{q_1, \dots, q_m\}$ , is called the *variable finite memory* or simply *variable memory*. A formalization of these concepts is given in Section 3.1 below. Matrices that form filter models with a variable memory possess special structure. Some related examples are given in Section 3.1.

Thus, our first motivation to consider the problem in the form presented in Section 2.3 below comes from the observation that the compressor and de-compressor used in real time should be causal with variable finite memory.

SECOND MOTIVATION: SPECIFIC FORMULATION OF THE PROBLEM. In reality, the compression and de-compression are separated in time. Therefore, it is natural to pose optimization problems for them separately, one specific problem for each operation, compression and de-compression. Associated optimization criteria could be formulated in many ways. Some of them are discussed in Section 6, and we show that those criteria lead to significant difficulties. To avoid the difficulties considered in Section 6, a new approach to the solution of the data compression problem is presented here. The approach is based on a specific formulation of two related problems given in Section 2 below. Solutions of those problems represent an associated optimal compressor and optimal de-compressor, respectively. It is shown in Section 3.1 that the optimal compressor and de-compressor satisfying conditions of causality and variable finite memory must have special forms. This implies that signals processed by these

<sup>3</sup>Below, when a context is clear, we use the term ‘filter’ for both compressor and de-compressor.

operators should be presented in special forms as well. In Sections 3.1 and 3.2 this issue is discussed in detail.

Next, traditionally, the data compression problem is studied in terms of linear operators, mainly due to a simplicity of their implementation. See, for example, (Scharf, 1991)–(Jolliffe, 1986) and (Scharf, 1991)–(Torokhti and Friedland, 2009) and references herein. Here, we extend the approaches of linear data compression proposed in (Scharf, 1991)–(Jolliffe, 1986) and (Scharf, 1991)–(Torokhti and Friedland, 2009). A case of non-linear compression and de-compression with causality and memory is more complicated, and it can be studied on the basis of results obtained below combined, e.g., with the approaches to optimal non-linear filtering presented in (Torokhti and Howlett, 2007; 2006; 2009; Torokhti and Manton, 2009; Howlett et al., 2007).

## 2. Basic idea and statement of the problem

### 2.1 Informal statement of problem

In an informal way, the data compression problem we consider can be expressed as follows.

Let  $(\Omega, \Sigma, \mu)$  be a probability space, where  $\Omega = \{\omega\}$  is the set of outcomes,  $\Sigma$  a  $\sigma$ -field of measurable subsets in  $\Omega$  and  $\mu : \Sigma \rightarrow [0, 1]$  an associated probability measure on  $\Sigma$  with  $\mu(\Omega) = 1$ . Let  $\mathbf{y} \in L^2(\Omega, \mathbb{R}^n)$  be observable data and  $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$  be a reference signal that is to be estimated from  $\mathbf{y}$  in such a way that

- (i) first, data  $\mathbf{y}$  should be compressed to a shorter vector  $\mathbf{z} \in L^2(\Omega, \mathbb{R}^r)^4$  with  $r < \min\{m, n\}$ ,
- (ii) then  $\mathbf{z}$  should be decompressed (reconstructed) to a signal  $\tilde{\mathbf{x}} \in L^2(\Omega, \mathbb{R}^m)$  so that  $\tilde{\mathbf{x}}$  is 'close' to  $\mathbf{x}$  in some appropriate sense, and
- (iii) both operations, compression and de-compression, should be *causal* and should have *variable memory*.

The problem is to determine a compressor and de-compressor so that associated errors are minimal. A rigorous statement of the problem is given in Section 2.3 below.

In an intuitive way,  $\mathbf{y}$  can be regarded as a noise-corrupted version of  $\mathbf{x}$ . For example,  $\mathbf{y}$  can be interpreted as  $\mathbf{y} = \mathbf{x} + \mathbf{n}$  where  $\mathbf{n}$  is white noise. Thus, the above two operations, (i) and (ii), perform the noise filtering as well. We do not restrict ourselves to this simplest version of  $\mathbf{y}$  and assume that the dependence of  $\mathbf{y}$  on  $\mathbf{x}$  and  $\mathbf{n}$  is arbitrary.

### 2.2 Basic idea

Let  $\mathcal{B} : L^2(\Omega, \mathbb{R}^n) \rightarrow L^2(\Omega, \mathbb{R}^r)$  signify compression so that  $\mathbf{z} = \mathcal{B}(\mathbf{y})$  and let  $\mathcal{A} : L^2(\Omega, \mathbb{R}^r) \rightarrow L^2(\Omega, \mathbb{R}^m)$  designate data de-compression, i.e.,  $\tilde{\mathbf{x}} = \mathcal{A}(\mathbf{z})$ . We suppose that  $\mathcal{B}$  and  $\mathcal{A}$  are linear operators defined by the relationships

$$[\mathcal{B}(\mathbf{y})](\omega) = B[\mathbf{y}(\omega)] \quad \text{and} \quad [\mathcal{A}(\mathbf{z})](\omega) = A[\mathbf{z}(\omega)] \quad (1)$$

where  $B \in \mathbb{R}^{n \times r}$  and  $A \in \mathbb{R}^{r \times m}$ . In the remainder of this paper we shall use the same symbol to represent both the linear operator acting on a random vector and its associated matrix.

We assume that information about vector  $\mathbf{z}$  in the form of associated covariance matrices can be obtained, in particular, from the known solution (Torokhti and Howlett, 2007) of the data compression problem *with no constraints associated with causality and memory*. This allows us to state the problem, subject to conditions of causality and memory, in the form of two separate problems (2) and (3) formulated below. This is the basic idea of our approach.

<sup>4</sup>Components of  $\mathbf{z}$  are often called *principal components* (Jolliffe, 1986).

### 2.3 The problem formalization

**SPECIAL NOTATION.** We need to use some special notation to be defined in Section 3.1 below. The notation follows from six specific Definitions 1–6 of causal compressors and de-compressors with different types of memories given in Section 3.1. To shorten the way forward for the problem formalization, we describe briefly here that notation and refer the reader to Section 3.1 for more detail.

We write  $\mathcal{M}_c(r, n, \eta_B)$  for the set of causal  $r \times n$  compressors  $B$  with so-called *complete* variable finite memory  $\eta_B$ , and  $\mathcal{M}_T(r, n, \eta_B)$  for the set of causal  $r \times n$  compressors  $B$  with so-called *truncated* variable finite memory  $\eta_B$ . The definitions  $\mathcal{M}_c(m, r, \eta_A)$  and  $\mathcal{M}_T(m, r, \eta_A)$  for sets of de-compressors  $A$  are similar. Related Definitions 1–6 are in Section 3.1.

**FORMULATION OF THE PROBLEM.** Now, we are in a position to formulate the problem rigorously. We define the norm to be  $\|\mathbf{x}\|_\Omega^2 = \int_\Omega \|\mathbf{x}(\omega)\|_2^2 d\mu(\omega)$  where  $\|\mathbf{x}(\omega)\|_2$  is the Euclidean norm of  $\mathbf{x}(\omega)$ . Consider

$$J_1(B) = \|\mathbf{z} - B(\mathbf{y})\|_\Omega^2.$$

Let  $B^0$  be such that

$$J_1(B^0) = \min_B J_1(B) \quad \text{subject to } B \in \mathcal{M}_c(r, n, \eta_B) \text{ or } B \in \mathcal{M}_T(r, n, \eta_B). \quad (2)$$

We write  $\mathbf{z}^0 = B^0(\mathbf{y})$ . Next, let

$$J_2(A) = \|\mathbf{x} - A(\mathbf{z}^0)\|_\Omega^2$$

and let  $A^0$  be such that

$$J_2(A^0) = \min_A J_2(A) \quad \text{subject to } A \in \mathcal{M}(m, r, \eta_A) \text{ or } A \in \mathcal{M}_T(m, r, \eta_A). \quad (3)$$

We denote  $\mathbf{x}^0 = A^0(\mathbf{z}^0)$ .

The problem considered in this paper is to find operators  $B^0$  and  $A^0$  that satisfy minimization criteria (2) and (3), respectively.

Operator  $B^0$  provides a compression of the signal  $\mathbf{y}$  to the form  $\mathbf{z}^0$ . Further,  $A^0$  reconstructs then compressed signal  $\mathbf{z}^0$  to the form  $\mathbf{x}^0$  so that  $\mathbf{x}^0$  is an optimal representation of  $\mathbf{x}$  in the sense of the constrained minimization problem (3).

### 2.4 Differences of our statement of the problem

The major differences between our statement of the problem (2)–(3) and the known statements considered, for example, in (Scharf, 1991)–(Jolliffe, 1986) and (Scharf, 1991)–(Torokhti and Friedland, 2009) are as follows. Firstly, de-compressor  $A$  and compressor  $B$  should be causal with variable finite memory. Secondly, we represent the data compression problem in the form of a concatenation of two new separate problems (2) and (3). Some related arguments for considering the real-time data compression problem in the form (2)–(3) are presented in Section 6.

## 3. Main results

### 3.1 Formalization of the concept of variable memory

Let  $\tau_1 < \tau_2 < \dots < \tau_n$  be time instants and  $\alpha, \beta, \vartheta : \mathbb{R} \rightarrow L^2(\Omega, \mathbb{R})$  be continuous functions. Suppose  $\alpha_k = \alpha(\tau_k)$ ,  $\beta_k = \beta(\tau_k)$  and  $\vartheta_k = \vartheta(\tau_k)$  are real-valued, random variables having finite second moments. We write

$$\mathbf{x} = [\alpha_1, \alpha_2, \dots, \alpha_m]^T \quad \mathbf{y} = [\beta_1, \beta_2, \dots, \beta_n]^T \quad \text{and} \quad \mathbf{z} = [\vartheta_1, \dots, \vartheta_r]^T.$$

Let  $\tilde{\mathbf{z}}$  be a compressed form of data  $\mathbf{y}$  defined by  $\tilde{\mathbf{z}} = B(\mathbf{y})$  with  $\tilde{\mathbf{z}} = [\tilde{\theta}_1, \dots, \tilde{\theta}_r]^T$ , and  $\tilde{\mathbf{x}}$  be a de-compression of  $\tilde{\mathbf{z}}$  defined by  $\tilde{\mathbf{x}} = A(\tilde{\mathbf{z}})$  with  $\tilde{\mathbf{x}} = [\tilde{\alpha}_1, \dots, \tilde{\alpha}_m]^T$ .

In many applications<sup>5</sup>, to obtain  $\tilde{\theta}_k$  for  $k = 1, \dots, r$ , it is necessary for the compressor  $B$  to use only a limited number of input components,  $\eta_{B_k} = 1, \dots, r$ . A number of such input components  $\eta_{B_k}$  is here called a *kth local memory* for  $B$ .

To define a notation of memory for the compressor  $B$ , we use parameters  $p$  and  $g$  which are positive integers such that

$$1 \leq p \leq n \quad \text{and} \quad n - r + 2 \leq g \leq n.$$

**Definition 1** The vector  $\eta_B = [\eta_{B_1}, \dots, \eta_{B_r}]$  is called a *variable memory* of the compressor  $B$ . In particular,  $\eta_B$  is called a *complete variable memory* if  $\eta_{B_1} = g$  and  $\eta_{B_k} = n$  when  $k = n - g + 1, \dots, n$ . Vector  $\eta_B$  is called a *truncated variable memory* of  $B$  if, for  $r \leq p \leq n$ ,  $\eta_{B_1} = p - r + 1$  and  $\eta_{B_r} = p$ . Here,  $p$  relates to the last possible nonzero entry in the bottom row of  $B$  and  $g$  relates to the last possible nonzero entry in the first row.

The notation  $\eta_A = [\eta_{A_1}, \dots, \eta_{A_m}]$  has a similar meaning for the de-compressor  $A$ . Here,  $\eta_{A_j}$  is the *jth local memory* of  $A$ . In other words,  $\eta_{A_j}$  is the number of input components used by the de-compressor  $A$  to obtain the estimate  $\tilde{\alpha}_j$  with  $j = 1, \dots, m$ .

The parameters  $q$  and  $s$  which are positive integers such that

$$1 \leq q \leq r \quad \text{and} \quad 2 \leq s \leq m$$

are used below to define two types of memory for  $A$ .

**Definition 2** Vector  $\eta_A$  is called a *complete variable memory* of the de-compressor  $A$  if  $\eta_{A_1} = q$  and  $\eta_{A_j} = r$  when  $j = s + r - 1, \dots, m$ . Vector  $\eta_A$  is called a *truncated variable memory* of  $A$  if  $\eta_{A_j} = 0$  for  $j = 1, \dots, s - 1$ ,  $\eta_{A_s} = s$  and  $\eta_{A_j} = r$  when  $j = s + r - 1, \dots, m$ . Here,  $q$  relates to the first possible nonzero entry in the last column of  $A$  and  $s$  relates to the first possible nonzero entry in the first column.

The memory constraints described above imply that certain elements of the matrices  $B = \{b_{ij}\}_{i,j=1}^{r,n}$  and  $A = \{a_{ij}\}_{i,j=1}^{m,r}$  must be set equal to zero. In this regard, for matrix  $B$  with  $r \leq p \leq n$ , we require that

$$b_{i,j} = 0 \tag{4}$$

$$\text{if } j = p - r + i + 1, \dots, n, \quad \text{for } \begin{cases} p = r, \dots, n - 1, \\ i = 1, \dots, r \end{cases} \quad \text{and} \quad \begin{cases} p = n, \\ i = 1, \dots, r - 1, \end{cases} \tag{5}$$

and, for  $1 \leq p \leq r - 1$ , it is required that

$$b_{i,j} = 0 \tag{6}$$

$$\text{if } \begin{cases} i = 1, \dots, r - p, \\ j = 1, \dots, n, \end{cases} \quad \text{and} \quad \text{if } \begin{cases} i = r - p + 1, \dots, r, \\ j = i - r + p + 1, \dots, n. \end{cases} \tag{7}$$

See Examples 1 and 2 below.

<sup>5</sup>Examples include computer medical diagnostics (Gimeno, 1987) and problems of bio-informatics (Kim et al., 2005).



For matrix  $A$  with  $r \leq p \leq n$ , we require

$$a_{i,j} = 0 \quad (8)$$

$$\text{if } j = q + i, \dots, r \text{ for } q = 1, \dots, r - 1 \text{ and } i = 1, \dots, r - q, \quad (9)$$

and, for  $2 \leq s \leq m$ , it is required that

$$a_{i,j} = 0 \quad (10)$$

$$\text{if } j = s + i, \dots, r \text{ for } s = 1, \dots, m \text{ and } i = 1, \dots, s + r - 1, \quad (11)$$

See Examples 3 and 4 below.

The above conditions imply the following definitions.

**Definition 3** A matrix  $B$  satisfying the constraint (4)–(5) is said to be a causal operator with the truncated variable memory  $\eta_B = [p - r + 1, \dots, p]$ . The set of such matrices is denoted by  $\mathcal{M}_T(r, n, \eta_B)$ .

**Example 1** Let  $n = 8$ ,  $r = 3$  and  $p = 7$ . If the symbol  $\bullet$  denotes an entry that may be non-zero, then  $B$  is of the form

$$B = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \bullet & 0 & \overset{b_{1,p-r+1}}{\bullet} & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & 0 & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & 0 \end{bmatrix}_p$$

is a causal operator with the truncated variable memory  $\eta_B = [5, 6, 7]$ .

**Definition 4** A matrix  $B$  satisfying the constraint (6)–(7) is said to be a causal operator with the complete variable memory  $\eta_B = [g, g + 1, \dots, n]$ . Here,  $\eta_{B_k} = n$  when  $k = n - g + 1, \dots, n$ . The set of such matrices is denoted by  $\mathcal{M}_C(r, n, \eta_B)$ .

**Example 2** Let  $n = 6$ ,  $r = 4$  and  $g = 4$ . Then  $B$  with

$$B = \begin{bmatrix} \bullet & \bullet & \bullet & \bullet & 0 & \overset{g}{\bullet} \\ \bullet & \bullet & \bullet & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \end{bmatrix}_{n-g+1}$$

is the causal operator with the complete variable memory  $\eta_B = [4, 5, 6, 6]$ .

**Definition 5** A matrix  $A$  satisfying the constraint (8)–(9) is said to be a causal operator with the complete variable memory  $\eta_A = [r - q + 1, \dots, r]$ . Here,  $\eta_{A_j} = r$  when  $j = q, \dots, m$ . The set of such matrices is denoted by  $\mathcal{M}_C(m, r, \eta_A)$ .

**Example 3** Let  $m = 5$ ,  $r = 4$  and  $q = 3$ . Then  $A$  is of the form

$$A = \begin{bmatrix} \bullet & \bullet & 0 & \overset{a_{1,r-q+1}}{\bullet} \\ \bullet & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{bmatrix} \leftarrow a_{q,r}$$

and is a causal operator with the complete variable memory  $\eta_A = [2, 3, 4, 4, 4]$ .

**Definition 6** A matrix  $A$  satisfying the constraint (10)–(11) is said to be a causal operator with the truncated variable memory  $\eta_A = [0, \dots, 0, 1, \dots, r]$ . Here,  $\eta_{A_j} = 0$  when  $j = 1, \dots, s-1$ , and  $\eta_{A_j} = r$  when  $j = s+r-1, \dots, m$ . The set of such matrices is denoted by  $\mathcal{M}_T(m, r, \eta_A)$ .

**Example 4** Let  $m = 6$ ,  $r = 4$  and  $s = 3$ . Then matrix  $A$  of the form

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \bullet & 0 & 0 & 0 \\ \bullet & \bullet & 0 & 0 \\ \bullet & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & \bullet \end{bmatrix}$$

is a causal operator with the truncated variable memory  $\eta_A = [0, 0, 1, 2, 3, 4]$ .

### 3.2 Solution of problems (2) and (3)

To proceed any further we shall require some more notation. Let

$$\langle \alpha_i, \beta_j \rangle = \int_{\Omega} \alpha_i(\omega) \beta_j(\omega) d\mu(\omega) < \infty, \quad E_{xy} = \{ \langle \alpha_i, \beta_j \rangle \}_{i,j=1}^{m,n} \in \mathbb{R}^{m \times n}, \quad (12)$$

$$\mathbf{y}_1 = [\beta_1, \dots, \beta_{g-1}]^T, \quad \mathbf{y}_2 = [\beta_g, \dots, \beta_n]^T, \quad (13)$$

$$\mathbf{z}_1 = [\vartheta_1, \dots, \vartheta_{g-1}]^T \quad \text{and} \quad \mathbf{z}_2 = [\vartheta_g, \dots, \vartheta_n]^T. \quad (14)$$

The pseudo-inverse matrix (Golub and Van Loan, 1996) for any matrix  $M$  is denoted by  $M^\dagger$ . The symbol  $\mathbf{O}$  designates the zero matrix.

**Definition 7** (Torokhti and Howlett, 2007) Two random vectors  $\mathbf{u}$  and  $\mathbf{w}$  are said to be mutually orthogonal if  $E_{uw} = \mathbf{O}$ .

**Lemma 1** (Torokhti and Howlett, 2007) If we define

$$\mathbf{w}_1 = \mathbf{y}_1 \quad \text{and} \quad \mathbf{w}_2 = \mathbf{y}_2 - P_y \mathbf{y}_1$$

where

$$P_y = E_{y_1 y_2} E_{y_1 y_1}^\dagger + D_y (I - E_{y_1 y_1} E_{y_1 y_1}^\dagger) \quad (15)$$

with  $D_y$  an arbitrary matrix, then  $\mathbf{w}_1$  and  $\mathbf{w}_2$  are mutually orthogonal random vectors.

#### 3.2.1 Solution of problem (2). The case of complete variable memory

Let us first consider problem (2) when  $B$  has the complete variable memory  $\eta_B = [g, g+1, \dots, n]$  (see Definition 4).

Let us partition  $B$  in four matrices  $K_B, L_B, S_{B_1}$  and  $S_{B_2}$  so that

$$B = \begin{bmatrix} K_B & L_B \\ S_{B_1} & S_{B_2} \end{bmatrix},$$



where

$K_B = \{k_{ij}\} \in \mathbb{R}^{n_b \times (g-1)}$  is a rectangular matrix with  $n_b = n - g + 1$ ,  
 $L_B = \{\ell_{ij}\} \in \mathbb{R}^{n_b \times n_b}$  is a lower triangular matrix, and  
 $S_{B_1} = \{s_{ij}^{(1)}\} \in \mathbb{R}^{(r-n_b) \times (g-1)}$  and  $S_{B_2} = \{s_{kl}^{(2)}\} \in \mathbb{R}^{(r-n_b) \times n_b}$  are rectangular matrices.

We have

$$\begin{aligned} B(\mathbf{y}) &= \begin{bmatrix} K_B & L_B \\ S_{B_1} & S_{B_2} \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} K_B(\mathbf{y}_1) + L_B(\mathbf{y}_2) \\ S_{B_1}(\mathbf{y}_1) + S_{B_2}(\mathbf{y}_2) \end{bmatrix} \\ &= \begin{bmatrix} K_B(\mathbf{w}_1) + L_B(\mathbf{w}_2 + P_y(\mathbf{w}_1)) \\ S_{B_1}(\mathbf{w}_1) + S_{B_2}(\mathbf{w}_2 + P_y(\mathbf{w}_1)) \end{bmatrix} = \begin{bmatrix} T_B(\mathbf{w}_1) + L_B(\mathbf{w}_2) \\ S_B(\mathbf{w}_1) + S_{B_2}(\mathbf{w}_2) \end{bmatrix}, \end{aligned} \quad (16)$$

where

$$T_B = K_B + L_B P_y \quad \text{and} \quad S_B = S_{B_1} + S_{B_2} P_y. \quad (17)$$

Then

$$\begin{aligned} J_1(B) &= \|\mathbf{z} - B(\mathbf{y})\|_{\Omega}^2 \\ &= \left\| \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} - \begin{bmatrix} T_B & L_B \\ S_B & S_{B_2} \end{bmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \end{bmatrix} \right\|_{\Omega}^2 \\ &= J^{(1)}(T_B, L_B) + J^{(2)}(S_B, S_{B_2}), \end{aligned} \quad (18)$$

where

$$J^{(1)}(T, L) = \|\mathbf{z}_1 - [T_B(\mathbf{w}_1) + L_B(\mathbf{w}_2)]\|_{\Omega}^2 \quad \text{and} \quad J^{(2)}(S_B, S_{B_2}) = \|\mathbf{z}_2 - [S_B(\mathbf{w}_1) + S_{B_2}(\mathbf{w}_2)]\|_{\Omega}^2.$$

By analogy with Lemma 37 in (Torokhti and Howlett, 2007),

$$\min_{B \in \mathcal{M}(r, n, \eta_B)} J_1(B) = \min_{T_B, L_B} J^{(1)}(T_B, L_B) + \min_{S_B, S_{B_2}} J^{(2)}(S_B, S_{B_2}). \quad (19)$$

Therefore, problem (2) is reduced to finding matrices  $T_B^0, L_B^0, S_B^0$  and  $S_{B_2}^0$  such that

$$J^{(1)}(T_B^0, L_B^0) = \min_{T_B, L_B} J^{(1)}(T_B, L_B) \quad (20)$$

and

$$J^{(2)}(S_B^0, S_{B_2}^0) = \min_{S_B, S_{B_2}} J^{(2)}(S_B, S_{B_2}). \quad (21)$$

Taking into account the orthogonality of vectors  $\mathbf{w}_1$  and  $\mathbf{w}_2$ , and working in analogy with the argument on pp. 348–352 in (Torokhti and Howlett, 2007), it follows that matrices  $S_B^0$  and  $S_{B_2}^0$  are given by

$$S_B^0 = E_{z_2 w_1} E_{w_1 w_1}^{\dagger} + H_B (I - E_{w_1 w_1} E_{w_1 w_1}^{\dagger}) \quad (22)$$

and

$$S_{B_2}^0 = E_{z_2 w_2} E_{w_2 w_2}^{\dagger} + H_{B_2} (I - E_{w_2 w_2} E_{w_2 w_2}^{\dagger}), \quad (23)$$

where  $H_B$  and  $H_{B_2}$  are arbitrary matrices. See Section 7 for more details.  
 Next, to find  $T_B^0$  and  $L_B^0$  we use the following notation.

For  $r = 1, 2, \dots, \ell$ , let  $\rho$  be the rank of the matrix  $E_{w_2 w_2} \in \mathbb{R}^{n_2 \times n_2}$  with  $n_b = n - g + 1$ , and let

$$E_{w_2 w_2}^{1/2} = Q_{w,\rho} R_{w,\rho}$$

be the QR-decomposition for  $E_{w_2 w_2}^{1/2}$  where  $Q_{w,\rho} \in \mathbb{R}^{n_2 \times \rho}$  and  $Q_{w,\rho}^T Q_{w,\rho} = I$  and  $R_{w,\rho} \in \mathbb{R}^{\rho \times n_2}$  is upper trapezoidal with rank  $\rho$ . We write  $G_{w,\rho} = R_{w,\rho}^T$  and use the notation

$$G_{w,\rho} = [g_1, \dots, g_\rho] \in \mathbb{R}^{n_2 \times \rho}$$

where  $g_j \in \mathbb{R}^{n_2}$  denotes the  $j$ -th column of  $G_{w,\rho}$ . We also write

$$G_{w,s} = [g_1, \dots, g_s] \in \mathbb{R}^{n_2 \times s} \quad (24)$$

for  $s \leq \rho$  to denote the matrix consisting of the first  $s$  columns of  $G_{w,\rho}$ . For the sake of simplicity, let us set

$$G_s := G_{w,s}. \quad (25)$$

Next,

$$\mathbf{e}_1^T = [1, 0, 0, 0, \dots], \quad \mathbf{e}_2^T = [0, 1, 0, 0, \dots], \quad \mathbf{e}_3^T = [0, 0, 1, 0, \dots], \quad \text{etc.}$$

denote the unit row vectors whatever the dimension of the space.

Finally, any square matrix  $M$  can be written as  $M = M_\Delta + M_\nabla$  where  $M_\Delta$  is lower triangular and  $M_\nabla$  is strictly upper triangular. We write  $\|\cdot\|_F$  for the Frobenius norm.

**Lemma 2** Let matrices  $K$ ,  $N$  and  $C$  be such that  $K = N + C$  and  $n_{ij}c_{ij} = 0$  for all  $i, j$ , where  $n_{ij}$  and  $c_{ij}$  are entries of  $N$  and  $C$ , respectively. Then

$$\|K\|_F^2 = \|N\|_F^2 + \|C\|_F^2.$$

**Proof.** The proof is obvious. □

**Theorem 1** Let  $B$  have the complete variable memory  $\eta_B = [g, g+1, \dots, n]$ . Then the solution to problem (2) is provided by the matrix  $B^0$ , which has the form

$$B^0 = \begin{bmatrix} K_B^0 & L_B^0 \\ S_{B_1}^0 & S_{B_2}^0 \end{bmatrix},$$

where the blocks  $K_B^0 \in \mathbb{R}^{n_b \times (g-1)}$ ,  $S_{B_1}^0 \in \mathbb{R}^{(r-n_b) \times (g-1)}$  and  $S_{B_2}^0 \in \mathbb{R}^{(r-n_b) \times n_b}$  are rectangular, and the block  $L_B^0 \in \mathbb{R}^{n_b \times n_b}$  is lower triangular. These blocks are given as follows. The block  $K_B^0$  is given by

$$K_B^0 = T_B^0 - L_B^0 P_y \quad (26)$$

with

$$T_B^0 = E_{z_1 w_1} E_{w_1 w_1}^+ + N_{B_1} (I - E_{w_1 w_1} E_{w_1 w_1}^+) \quad (27)$$

where  $N_{B_1}$  is an arbitrary matrix. The block  $L_B^0 = \begin{bmatrix} \lambda_1^0 \\ \vdots \\ \lambda_{n_b}^0 \end{bmatrix}$ , for each  $s = 1, 2, \dots, n_2$ , is defined by its rows

$$\lambda_s^0 = \mathbf{e}_s^T E_{z_1 w_2} E_{w_2 w_2}^+ G_s G_s^+ + f_s^T (I - G_s G_s^+) \quad (28)$$

with  $f_s^T \in \mathbb{R}^{1 \times n_2}$  arbitrary. The blocks  $S_{B_1}^0$  and  $S_{B_2}^0$  are given by

$$S_{B_1}^0 = S_B^0 - S_{B_2}^0 P_y \quad (29)$$

and (23), respectively. In (29),  $S_B^0$  is presented by (22).

The error associated with the matrix  $B^0$  is given by

$$\begin{aligned} \|\mathbf{z} - B^0 \mathbf{y}\|_{\Omega}^2 &= \sum_{s=1}^{\rho} \sum_{j=s+1}^{n_2} |e_s^T E_{z_1 w_2} E_{w_2 w_2}^{\dagger} g_j|^2 \\ &\quad + \sum_{j=1}^2 \|E_{z_j z_j}^{1/2}\|_F^2 - \sum_{i=1}^2 \sum_{j=1}^2 \|E_{z_i w_i} E_{w_j w_j}^{\dagger 1/2}\|_F^2. \end{aligned} \quad (30)$$

**Proof.** Since  $E_{w_2 w_2} = G_{w, \rho} G_{w, \rho}^T$ , we have

$$\begin{aligned} J^{(1)}(T, L) &= \|\mathbf{z}_1 - [T(\mathbf{w}_1) + L(\mathbf{w}_2)]\|_{\Omega}^2 \\ &= \text{tr} \left\{ E_{z_1 z_1} - E_{z_1 w_1} T^T - E_{z_1 w_2} L^T - T E_{w_1 z_1} \right. \\ &\quad \left. + T E_{w_1 w_1} T^T + T E_{w_1 w_2} L^T - L E_{w_2 z_1} \right. \\ &\quad \left. + L E_{w_2 w_1} T^T + L E_{w_2 w_2} L^T \right\} \\ &= \text{tr} \left\{ E_{z_1 z_1} - E_{z_1 w_1} T^T - E_{z_1 w_2} L^T - T E_{w_1 z_1} \right. \\ &\quad \left. + T E_{w_1 w_1} T^T - L E_{w_2 z_1} + L E_{w_2 w_2} L^T \right\} \\ &= \text{tr} \left\{ (T - E_{z_1 w_1} E_{w_1 w_1}^{\dagger}) E_{w_1 w_1} (T^T - E_{w_1 w_1}^{\dagger} E_{w_1 z_1}) \right. \\ &\quad \left. + (L - E_{z_1 w_2} E_{w_2 w_2}^{\dagger}) G_{\rho} G_{\rho}^T (L^T - E_{w_2 w_2}^{\dagger} E_{w_2 z_1}) \right. \\ &\quad \left. + E_{z_1 z_1} - E_{z_1 w_1} E_{w_1 w_1}^{\dagger} E_{w_1 z_1} - E_{z_1 w_2} E_{w_2 w_2}^{\dagger} E_{w_2 z_1} \right\} \\ &= \|(T - E_{z_1 w_1} E_{w_1 w_1}^{\dagger}) E_{w_1 w_1}^{1/2}\|_F^2 \\ &\quad + \|(L - E_{z_1 w_2} E_{w_2 w_2}^{\dagger}) G_{\rho}\|_F^2 + \|E_{z_1 z_1}^{1/2}\|_F^2 \\ &\quad - \|E_{z_1 w_1} E_{w_1 w_1}^{\dagger 1/2}\|_F^2 - \|E_{z_1 w_2} E_{w_2 w_2}^{\dagger 1/2}\|_F^2. \end{aligned} \quad (31)$$

On the basis of Lemma 2 and the fact that the matrix  $LG_\rho$  is lower triangular we note that

$$\begin{aligned}
 & \| (L - E_{z_1 w_2} E_{w_2 w_2}^\dagger) G_\rho \|_F^2 \\
 &= \| (LG_\rho - E_{z_1 w_2} E_{w_2 w_2}^\dagger G_\rho)_\Delta - (E_{z_1 w_2} E_{w_2 w_2}^\dagger G_\rho)_\nabla \|_F^2 \\
 &= \sum_{s=1}^{\rho} \sum_{j=1}^s |(\lambda_s g_j - e_s^T E_{z_1 w_2} E_{w_2 w_2}^\dagger g_j)|^2 \\
 &\quad + \sum_{s=1}^{\rho} \sum_{j=s+1}^{n_b} |e_s^T E_{z_1 w_2} E_{w_2 w_2}^\dagger g_j|^2 \\
 &= \sum_{s=1}^{\rho} \| \lambda_s G_s - e_s^T E_{z_1 w_2} E_{w_2 w_2}^\dagger G_s \|_2^2 \\
 &\quad + \sum_{s=1}^{\rho} \sum_{j=s+1}^{n_b} |e_s^T E_{z_1 w_2} E_{w_2 w_2}^\dagger g_j|^2. \tag{32}
 \end{aligned}$$

The first sum in (32) calculates the contribution from all elements with  $j \leq s$  that are on or below the leading diagonal of the matrix  $(L - E_{z_1 w_2} E_{w_2 w_2}^\dagger) G_\rho$  and the second sum calculates the contribution from all elements with  $j > s$  that are strictly above the leading diagonal. To minimize the overall expression (32) it would be sufficient to set the terms in the first sum to zero. Thus we wish to solve the matrix equation

$$\lambda_s G_s - e_s^T E_{z_1 w_2} E_{w_2 w_2}^\dagger G_s = 0.$$

This equation is a system of  $(2n_b - \rho + 1)\rho_2/2$  equations in  $(n_b + 1)n_b/2$  unknowns. Hence, there is always at least one solution. Indeed, by applying similar arguments to those used in proving Lemma 26 of (Torokhti and Howlett, 2007), it can be seen that the general solution is given by (28).

Next, it follows from (31) that the minimum of

$$\| (T_B - E_{z_1 w_1} E_{w_1 w_1}^\dagger) E_{w_1 w_1}^{1/2} \|_F^2$$

is attained if

$$(T_B - E_{z_1 w_1} E_{w_1 w_1}^\dagger) E_{w_1 w_1}^{1/2} = \mathbf{O}. \tag{33}$$

By Lemma 26 of (Torokhti and Howlett, 2007), this equation is equivalent to the equation

$$T_B E_{w_1 w_1} - E_{z_1 w_1} = \mathbf{O}. \tag{34}$$

See Section 7 for more details. The general solution (Ben-Israel and Greville, 1974) to (34) is given by (43). Then (26) follows from (17). Equality (29) also follows from (17).

To obtain the error representation (30), we write

$$\begin{aligned}
 \| \mathbf{z} - B^0 \mathbf{y} \|_\Omega^2 &= J^{(1)}(T_B^0, L_B^0) + J^{(2)}(S_B^0, S_{B_2}^0) \\
 &= \| \mathbf{z}_1 - [T_B^0(\mathbf{w}_1) + L_B^0(\mathbf{w}_2)] \|_\Omega^2 + \| \mathbf{z}_2 - [S_B^0(\mathbf{w}_1) + S_{B_2}^0(\mathbf{w}_2)] \|_\Omega^2.
 \end{aligned}$$

Here,

$$J^{(1)}(T_B^0, L_B^0) = \sum_{s=1}^{\rho} \sum_{j=s+1}^{n_2} |e_s^T E_{z_1 w_2} E_{w_2 w_2}^\dagger g_j|^2 \\ + \|E_{z_1 z_1}^{1/2}\|_F^2 - \sum_{j=1}^2 \|E_{z_1 w_j} E_{w_j w_j}^\dagger\|_F^2$$

which follows from (31) by substituting (43) and (28). Similarly, the term  $J^{(2)}(S_B^0, S_{B_2}^0)$  is represented by (see Corollary 20 (p. 351) in (Torokhti and Howlett, 2007))

$$J^{(2)}(S_B^0, S_{B_2}^0) = \|E_{z_2 z_2}^{1/2}\|_F^2 - \sum_{j=1}^2 \|E_{z_2 w_j} E_{w_j w_j}^\dagger\|_F^2.$$

Hence, (30) follows.  $\square$

### 3.2.2 Solution of problem (3). The case of complete variable memory

Let us now consider problem (3) when the de-compressor  $A$  has the complete variable memory  $\eta_A = [r - q + 1, \dots, r]$  (see Definition 5).

In analogy with our partitioning of matrix  $B$ , we partition matrix  $A$  in four matrices  $K_A, L_A, S_{A1}$  and  $S_{A2}$  so that

$$A = \begin{bmatrix} K_A & L_A \\ S_{A1} & S_{A2} \end{bmatrix},$$

where

$K_A = \{k_{ij}\} \in \mathbb{R}^{q \times (r-q)}$  is a rectangular matrix,

$L_A = \{\ell_{ij}\} \in \mathbb{R}^{q \times q}$  is a lower triangular matrix, and

$S_{A1} = \{s_{ij}^{(1)}\} \in \mathbb{R}^{(m-q) \times (r-q)}$  and  $S_{A2} = \{s_{kl}^{(2)}\} \in \mathbb{R}^{(m-q) \times q}$  are rectangular matrices.

Let us partition  $\mathbf{z}^0$  so that

$$\mathbf{z}^0 = \begin{bmatrix} \mathbf{z}_1^0 \\ \mathbf{z}_2^0 \end{bmatrix}$$

with  $\mathbf{z}_1^0 \in L^2(\Omega, \mathbb{R}^{r-q})$  and  $\mathbf{z}_2^0 \in L^2(\Omega, \mathbb{R}^q)$ . We also write

$$\mathbf{x}_1 = [\alpha_1, \dots, \alpha_{r-q}]^T \quad \text{and} \quad \mathbf{x}_2 = [\alpha_{r-q+1}, \dots, \alpha_m]^T,$$

and denote by  $\mathbf{v}_1 \in L^2(\Omega, \mathbb{R}^{r-q})$  and  $\mathbf{v}_2 \in L^2(\Omega, \mathbb{R}^q)$ , orthogonal vectors according to Lemma 1 as

$$\mathbf{v}_1 = \mathbf{z}_1^0 \quad \text{and} \quad \mathbf{v}_2 = \mathbf{z}_2^0 - P_z \mathbf{z}_1^0,$$

where  $P_z = E_{z_1 z_2} E_{z_1 z_1}^\dagger + D_z (I - E_{z_1 z_1} E_{z_1 z_1}^\dagger)$  with  $D_z$  an arbitrary matrix.

By analogy with (24)–(25), we write

$$G_{v,s} = [g_1, \dots, g_s] \in \mathbb{R}^{q \times s}$$

where  $G_{v,s}$  is constructed from a QR-decomposition of  $E_{v_2 v_2}^{1/2}$ , in a manner similar to the construction of matrix  $G_{w,s}$ .

Furthermore, we shall define  $G_s := G_{v,s}$ .

**Theorem 2** Let  $A$  have the complete variable memory  $\eta_A = [r - q + 1, \dots, r]$ . Then the solution to problem (3) is provided by the matrix  $A^0$ , which has the form

$$A^0 = \begin{bmatrix} K_A^0 & L_A^0 \\ S_{A1}^0 & S_{A2}^0 \end{bmatrix},$$

where the blocks  $K_A^0 \in \mathbb{R}^{q \times (r-q)}$ ,  $S_{A1}^0 \in \mathbb{R}^{(m-q) \times (r-q)}$  and  $S_{A2}^0 \in \mathbb{R}^{(m-q) \times q}$  are rectangular, and the block  $L_A^0 \in \mathbb{R}^{q \times q}$  is lower triangular. These blocks are given as follows. The block  $K_A^0$  is given by

$$K_A^0 = T_A^0 - L_A^0 P \quad (35)$$

with

$$T_A^0 = E_{x_1 v_1} E_{v_1 v_1}^+ + N_{A1} (I - E_{v_1 v_1} E_{v_1 v_1}^+) \quad (36)$$

where  $N_{A1}$  is an arbitrary matrix. The block  $L_A^0 = \begin{bmatrix} \lambda_1^0 \\ \vdots \\ \lambda_q^0 \end{bmatrix}$  for each  $s = 1, 2, \dots, q$ , is defined by its rows

$$\lambda_s^0 = \mathbf{e}_s^T E_{x_1 v_2} E_{v_2 v_2}^+ G_s G_s^+ + f_s^T (I - G_s G_s^+) \quad (37)$$

with  $f_s^T \in \mathbb{R}^{1 \times q}$  arbitrary. The blocks  $S_{A1}^0$  and  $S_{A2}^0$  are given by

$$S_{A1}^0 = S_A^0 - S_{A2}^0 P \quad \text{and} \quad S_{A2}^0 = E_{x_2 v_2} E_{v_2 v_2}^+ + H_{A2} (I - E_{v_2 v_2} E_{v_2 v_2}^+), \quad (38)$$

where

$$S_A^0 = E_{x_2 v_1} E_{v_1 v_1}^+ + H_A (I - E_{v_1 v_1} E_{v_1 v_1}^+) \quad (39)$$

and  $H_{A2}$  and  $H_A$  are arbitrary matrices.

The error associated with the de-compressor  $A^0$  is given by

$$\begin{aligned} \|\mathbf{x} - A^0 \mathbf{z}^0\|_\Omega^2 &= \sum_{s=1}^p \sum_{j=s+1}^q |e_s^T E_{x_1 v_2} E_{v_2 v_2}^+ g_j|^2 \\ &\quad + \sum_{j=1}^2 \|E_{x_j x_j}^{1/2}\|_F^2 - \sum_{i=1}^2 \sum_{j=1}^2 \|E_{x_i v_i} E_{v_j v_j}^{+1/2}\|_F^2. \end{aligned} \quad (40)$$

**Proof.** The proof follows in analogy with the proof of Theorem 1.  $\square$

### 3.2.3 Solution of problem (2): the case of truncated variable memory

Let us now consider a solution to the problem (2) when the compressor  $B$  has the truncated variable memory  $\eta_B = [p - r + 1, \dots, p]^T$  (see Definition 3).

To this end, let us partition  $B$  in three blocks  $K_B, L_B$  and  $Z_B$  so that

$$B = [K_B \quad L_B \quad Z_B],$$

where

$K_B = \{k_{ij}\} \in \mathbb{R}^{r \times (p-r)}$  is a rectangular matrix,

$L_B = \{\ell_{ij}\} \in \mathbb{R}^{r \times r}$  is a lower triangular matrix, and

$Z_B = \{z_{ij}\} \in \mathbb{R}^{r \times (n-p)}$  is the zero rectangular matrix.



Let us write

$$\mathbf{y}_1 = [\beta_1, \dots, \beta_{p-r}]^T, \quad \mathbf{y}_2 = [\beta_{p-r+1}, \dots, \beta_p]^T \quad \text{and} \quad \mathbf{y}_3 = [\beta_{p+1}, \dots, \beta_n]^T.$$

Therefore

$$B(\mathbf{y}) = \begin{bmatrix} K_B & L_B & Z_B \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = K_B(\mathbf{w}_1) + L_B(\mathbf{w}_2 + P_y(\mathbf{w}_1)) = T_B(\mathbf{w}_1) + L_B(\mathbf{w}_2),$$

where  $T_B$  is given by (17). Then

$$\begin{aligned} J_1(B) &= \|\mathbf{z} - B(\mathbf{y})\|_{\Omega}^2 \\ &= \|\mathbf{z} - [T_B(\mathbf{w}_1) + L_B(\mathbf{w}_2)]\|_{\Omega}^2 \\ &= J^{(1)}(T_B, L_B). \end{aligned} \quad (41)$$

Similarly to (24)–(25), we write

$$G_{w,s} = [g_1, \dots, g_s] \in \mathbb{R}^{r \times s}$$

where  $G_{w,s}$  is constructed from a QR-decomposition of  $E_{w_2 w_2}^{1/2}$ , in a manner similar to the construction of matrix  $G_{w,s}$ .

Furthermore, we shall define  $G_s := G_{w,s}$ .

A comparison of (41) with (18) and Theorem 1 shows that the solution to the problem under consideration follows from Theorem 1 as its particular case as follows.

**Corollary 1** Let  $B \in \mathcal{M}_T(r, n, \eta_B)$ , i.e. the compressor  $B$  is causal and has the truncated variable memory  $\eta_B = [p - r + 1, \dots, p]^T$ . Then the solution to problem (2) is provided by the matrix  $B^0$ , which has the form

$$B^0 = \begin{bmatrix} K_B^0 & L_B^0 & Z_B \end{bmatrix}.$$

Here, the block  $K_B^0 \in \mathbb{R}^{r \times (p-r)}$  is given by

$$K_B^0 = T_B^0 - L_B^0 P_y \quad (42)$$

with  $P_y$  determined by (15) and

$$T_B^0 = E_{z_1 w_1} E_{w_1 w_1}^+ + N_B (I - E_{w_1 w_1} E_{w_1 w_1}^+) \quad (43)$$

where  $N_B$  is an arbitrary matrix. The lower triangular block  $L_B^0 = \begin{bmatrix} \lambda_1^0 \\ \vdots \\ \lambda_r^0 \end{bmatrix} \in \mathbb{R}^{r \times r}$ , for each  $s = 1, 2, \dots, r$ , is defined by its rows

$$\lambda_s^0 = \mathbf{e}_s^T E_{z w_2} E_{w_2 w_2}^+ G_s G_s^+ + f_s^T (I - G_s G_s^+), \quad (44)$$

where  $G_s$  determined by (24)–(25) with  $E_{w_2 w_2} \in \mathbb{R}^{r \times r}$ , and  $f_s^T \in \mathbb{R}^{1 \times r}$  arbitrary.

Let  $\rho$  be the rank of the matrix  $E_{w_2 w_2} \in \mathbb{R}^{r \times r}$ . The error associated with the compressor  $B^0$  is given by

$$J^{(1)}(T_B^0, L_B^0) = \sum_{s=1}^{\rho} \sum_{j=s+1}^r |e_s^T E_{z w_2} E_{w_2 w_2}^\dagger g_j|^2 + \|E_{zz}^{1/2}\|_F^2 - \sum_{j=1}^2 \|E_{zw_j} E_{w_j w_j}^\dagger\|_F^2. \quad (45)$$

**Proof.** (42)–(44) follow in analogy with (26)–(28), and (45) is similar to (30).  $\square$

### 3.2.4 Solution of problem (3): the case of truncated variable memory

A solution to the problem (3) when the de-compressor  $A$  has the truncated variable memory  $\eta_A = [0, \dots, 0, 1, \dots, r]^T$  (see Definition 6) is obtained in a similar manner to the solution of the problem (3) given in Section 3.2.2 above.

We partition matrix  $A$  and vector  $\mathbf{z}^0$  in three blocks  $Z_A$ ,  $L_A$  and  $K_A$ , and three sub-vectors  $\mathbf{z}_1^0$ ,  $\mathbf{z}_2^0$  and  $\mathbf{z}_3^0$ , respectively, so that

$$A = \begin{bmatrix} Z_A \\ L_A \\ K_A \end{bmatrix} \quad \text{and} \quad \mathbf{z}^0 = \begin{bmatrix} \mathbf{z}_1^0 \\ \mathbf{z}_2^0 \\ \mathbf{z}_3^0 \end{bmatrix},$$

where  $Z_A \in \mathbb{R}^{(s-1) \times r}$  is the zero matrix,  $L_A \in \mathbb{R}^{r \times r}$  is a lower triangular matrix,  $K_A \in \mathbb{R}^{(m-r) \times r}$  is a rectangular matrix, and  $\mathbf{z}_1^0 \in L^2(\Omega, \mathbb{R}^{s-1})$ ,  $\mathbf{z}_2^0 \in L^2(\Omega, \mathbb{R}^{r-s+1})$  and  $\mathbf{z}_3^0 \in L^2(\Omega, \mathbb{R}^{m-r})$ .

We also write

$$\mathbf{x}_1 = [\alpha_1, \dots, \alpha_{s-1}]^T, \quad \mathbf{x}_2 = [\alpha_s, \dots, \alpha_{s+r-1}]^T \quad \text{and} \quad \mathbf{x}_3 = [\alpha_{s+r}, \dots, \alpha_m]^T.$$

Therefore,

$$\begin{aligned} J_2(A) &= \left\| \mathbf{x} - \begin{bmatrix} Z_A \\ L_A \\ K_A \end{bmatrix} \mathbf{z}^0 \right\|_{\Omega}^2 \\ &= \left\| \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 - L_A(\mathbf{z}_2^0) \\ \mathbf{x}_3 - K_A(\mathbf{z}_3^0) \end{bmatrix} \right\|_{\Omega}^2 \\ &= \|\mathbf{x}_1\|_{\Omega}^2 + J_2^{(1)}(L_A) + J_2^{(1)}(K_A), \end{aligned} \quad (46)$$

where

$$J_2^{(1)}(L_A) = \|\mathbf{x}_2 - L_A(\mathbf{z}_2^0)\|_{\Omega}^2 \quad \text{and} \quad J_2^{(1)}(K_A) = \|\mathbf{x}_3 - K_A(\mathbf{z}_3^0)\|_{\Omega}^2. \quad (47)$$

By analogy with (24)–(25), we write  $G_s := G_{z_2^0, s}$  with

$$G_{z_2^0, s} = [g_1, \dots, g_s] \in \mathbb{R}^{r_s \times s},$$

where  $r_s = r - s + 1$  and  $G_{z_2^0, s}$  is constructed from a QR-decomposition of  $E_{z_2^0 z_2^0}^{1/2}$ , in a manner similar to the construction of matrix  $G_{w, s}$ .

The solution of the problem under consideration is as follows.

**Corollary 2** Let a de-compressor  $A$  have the truncated variable memory  $\eta_A = [0, \dots, 0, 1, \dots, r]^T$ . Then the solution to problem (3) is provided by the matrix  $A^0$  which has the form

$$A^0 = \begin{bmatrix} Z_A \\ L_A^0 \\ K_A^0 \end{bmatrix}.$$

Here,  $Z_A \in \mathbb{R}^{(s-1) \times r}$  is the zero matrix and  $L_A^0 = \begin{bmatrix} \lambda_1^0 \\ \vdots \\ \lambda_r^0 \end{bmatrix} \in \mathbb{R}^{r \times r}$  is a lower triangular matrix, for each  $s = 1, 2, \dots, r$ , defined by its rows

$$\lambda_s^0 = \mathbf{e}_s^T E_{x_2 z_2^0} E_{z_2^0 z_2^0}^\dagger G_s G_s^\dagger + f_s^T (I - G_s G_s^\dagger), \quad (48)$$

with  $f_s^T \in \mathbb{R}^{1 \times r}$  arbitrary. The block  $K_A^0 \in \mathbb{R}^{(m-r) \times r}$  is given by

$$K_A^0 = E_{x_3 z_3^0} E_{z_3^0 z_3^0}^\dagger + H_A (I - E_{z_3^0 z_3^0} E_{z_3^0 z_3^0}^\dagger), \quad (49)$$

where  $H_A$  is an arbitrary matrix.

Let  $\rho$  be the rank of the matrix  $E_{z_2^0 z_2^0} \in \mathbb{R}^{r_s \times r_s}$  where  $r_s = r - s + 1$ . The error associated with the de-compressor  $A^0$  is given by

$$\|\mathbf{x} - A^0(\mathbf{z}^0)\|_\Omega^2 = \|\mathbf{x}_1\|_\Omega^2 + \sum_{s=1}^{\rho} \sum_{j=s+1}^{r_s} |e_s^T E_{x_2 z_2^0} E_{z_2^0 z_2^0}^\dagger g_j|^2 + \|E_{x_3 x_3}^{+1/2}\|_F^2 - \|E_{x_3 z_3^0} E_{z_3^0 z_3^0}^{+1/2}\|_F^2. \quad (50)$$

**Proof.** By analogy with (31)) and (32), we have

$$J_2^{(1)}(L_A) = \|(L_A - E_{x_2 z_2^0} E_{z_2^0 z_2^0}^\dagger) G_{w, \rho}\|_F^2 + \|E_{x_2 x_2}^{1/2}\|_F^2 - \|E_{x_2 z_2^0} E_{z_2^0 z_2^0}^{+1/2}\|_F^2$$

and

$$\begin{aligned} \|(L_A - E_{x_2 z_2^0} E_{z_2^0 z_2^0}^\dagger) G_{w, \rho}\|_F^2 &= \sum_{s=1}^{\rho} \|\lambda_s G_s - e_s^T E_{x_2 z_2^0} E_{z_2^0 z_2^0}^\dagger G_s\|_2^2 \\ &\quad + \sum_{s=1}^{\rho} \sum_{j=s+1}^{r_s} |e_s^T E_{x_2 z_2^0} E_{z_2^0 z_2^0}^\dagger g_j|^2, \end{aligned} \quad (51)$$

respectively. Then (48) follows similarly to (28).

The equality (49) is determined from minimizing  $J_2^{(1)}(K_A)$  in (47) as it has been done in (22)–(23) and on pp. 348–352 in (Torokhti and Howlett, 2007).

To derive the error representation (50), we write

$$\begin{aligned} J_2(A^0) &= \|\mathbf{x} - A^0(\mathbf{z}^0)\|_\Omega^2 = J_2^{(1)}(L_A^0) + J_2^{(2)}(K_A^0) \\ &= \|\mathbf{x}_1\|_\Omega^2 + \sum_{s=1}^{\rho} \sum_{j=s+1}^{r_s} |e_s^T E_{x_2 z_2^0} E_{z_2^0 z_2^0}^\dagger g_j|^2 + \|E_{x_3 x_3}^{+1/2}\|_F^2 - \|E_{x_3 z_3^0} E_{z_3^0 z_3^0}^{+1/2}\|_F^2. \end{aligned}$$

Thus, (48)–(50) are true.  $\square$

### 3.3 Device for data compression and de-compression

The proposed filter  $F^0$  for data compression, filtering and de-compression consists of two devices, the compressor  $B^0$  and the de-compressor  $A^0$ , so that  $F^0 = A^0 B^0$ . The device  $B^0$  compresses observed data  $\mathbf{y} \in L^2(\Omega, \mathbb{R}^n)$  to a 'shorter' vector  $\mathbf{z}^0 \in L^2(\Omega, \mathbb{R}^r)$  with  $r < \min\{m, n\}$  where  $m$  is a dimension of the reference signal  $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$ . The de-compressor  $A^0$  restores  $\mathbf{z}^0$  to a signal  $\tilde{\mathbf{x}} \in L^2(\Omega, \mathbb{R}^m)$  so that this procedure minimizes the cost functional  $J_2(A)$  given by (3). The compression ratio associated with such devices is given by

$$c = \frac{r}{\min\{m, n\}}. \quad (52)$$

## 4. Simulations

The following simulations and numerical results illustrate the performance of the proposed approach.

Our filter  $F^0 = A^0 B^0$  has been applied to compression, filtering and subsequent restoration of the reference signals given by the matrix  $X \in \mathbb{R}^{256 \times 256}$ . The matrix  $X$  represents the data obtained from an aerial digital photograph of a plant<sup>6</sup> presented in Fig. 1.

We divide  $X$  into 128 sub-matrices  $X_{ij} \in \mathbb{R}^{m \times q}$  with  $i = 1, \dots, 16$ ,  $j = 1, \dots, 8$ ,  $m = 16$  and  $q = 32$  so that  $X = \{X_{ij}\}$ . By assumption, the sub-matrix  $X_{ij}$  is interpreted as  $q$  realizations of a random vector  $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$  with each column representing a realization. For each  $i = 1, \dots, 16$  and  $j = 1, \dots, 8$ , observed data  $Y_{ij}$  were modelled from  $X_{ij}$  in the form

$$Y_{ij} = X_{ij} \bullet \text{rand}(16, 32)_{(ij)}.$$

Here,  $\bullet$  means the Hadamard product and  $\text{rand}(16, 32)_{(ij)}$  is a  $16 \times 32$  matrix whose elements are uniformly distributed in the interval  $(0, 1)$ .

The proposed filter  $F^0$  has been applied to each pair  $\{X_{ij}, Y_{ij}\}$ .

In these simulations, we are mainly concerned with the implementation of conditions of causality and variable finite memory in the filter  $F^0$  and their influence on the filter performance. To this end, we considered compressors  $B^0$  and de-compressors  $A^0$  with different types of memory studied above.

First, each pair  $\{X_{ij}, Y_{ij}\}$  has been processed by compressors and de-compressors with the complete variable memory. In this regard, see Definitions 4 and 5. We denote  $B_c^0 = B^0$  and  $A_c^0 = A^0$  for such a compressor and de-compressor determined by Theorems 1 and 2, respectively, so that

$$B_c^0 \in \mathcal{M}_T(r, n, \eta_B) \quad \text{and} \quad A_c^0 \in \mathcal{M}_C(m, r, \eta_A)$$

where  $n = m = 16$ ,  $r = 8$ ,  $\eta_B = \{\eta_{Bk}\}_{k=1}^{16}$  with  $\eta_{Bk} = \begin{cases} 12 + k - 1, & \text{if } k = 1, \dots, 4, \\ 16, & \text{if } k = 5, \dots, 16 \end{cases}$ , and

$\eta_A = \{\eta_{Aj}\}_{j=1}^{16}$  with  $\eta_{Aj} = \begin{cases} 6 + j - 1, & \text{if } j = 1, 2, \\ 8, & \text{if } j = 3, \dots, 16 \end{cases}$ . In this case, the optimal filter  $F^0$  is denoted by  $F_c^0$  so that  $F_c^0 = A_c^0 B_c^0$ . We write

$$J_c^0 = \max_{ij} \|X_{ij} - F_c^0 Y_{ij}\|^2$$

for a maximal error associated with the filter  $F_c^0$  over all  $i = 1, \dots, 16$  and  $j = 1, \dots, 8$ .

<sup>6</sup>The database is available in <http://sipi.usc.edu/services/database/Database.html>.

Second, each pair  $\{X_{ij}, Y_{ij}\}$  has been processed by compressors and de-compressors with the truncated variable memory defined by Definitions 3 and 6, respectively. Such a compressor and de-compressor are denoted by  $B_T^0 = B^0$  and  $A_T^0 = A^0$  with  $B^0$  and  $A^0$  determined by Corollaries 1 and 2. Here,

$$B_T^0 \in \mathcal{M}_T(r, n, \eta_B) \quad \text{and} \quad A_T^0 \in \mathcal{M}_T(m, r, \eta_A)$$

where  $n = m = 16, r = 8, \eta_B = \{\eta_{Bk}\}_{k=1}^{16}$  with  $\eta_{Bk} = 9 + k - 1$  for  $k = 1, \dots, 8$ , and  $\eta_A = \{\eta_{Aj}\}_{j=1}^{16}$  with  $\eta_{Aj} = \begin{cases} j, & \text{if } j = 1, \dots, 7, \\ 8, & \text{if } j = 8, \dots, 16 \end{cases}$ . The filter  $F^0$  composed from  $B_T^0$  and  $A_T^0$  is denoted by  $F_C^0$  so that  $F_C^0 = A_T^0 B_T^0$ . The maximal error associated with the filter  $F_T^0$  over all  $i = 1, \dots, 16$  and  $j = 1, \dots, 8$  is

$$J_T^0 = \max_{ij} \|X_{ij} - F_T^0 Y_{ij}\|^2.$$

The compression ratio for both cases above was  $c = 1/2$ .

We assumed that covariance matrices associated with a compressed signal and used to compute  $B_C^0, A_C^0, B_T^0$  and  $A_T^0$  could be determined from the Generic Karhunen-Loève Transform (GKLT) (Torokhti and Howlett, 2007) applied to each pair  $\{X_{ij}, Y_{ij}\}$ . We remind that the GKLT is not restricted by the conditions of causality and memory.

To the best of our knowledge, a method for the data compression subject to the conditions of causality and memory is presented here for the first time. Therefore, we could not compare our results with similar methods. However, to have an idea of some sort of comparison, we also computed the maximal error associated with the GKLT  $F_{GKLT}$  over all  $i = 1, \dots, 16$  and  $j = 1, \dots, 8$  as

$$J_{GKLT} = \max_{ij} \|X_{ij} - F_{GKLT} Y_{ij}\|^2.$$

The results of simulations are presented in Table 1 and Figures 1 (a) - (d).

Table 1.

$J_C^0$	$J_T^0$	$J_{GKLT}$
$3.3123e + 005$	$4.3649e + 005$	$3.0001e + 005$

As it has been shown above, the conditions of causality and memory imply the specific structure of the filter  $F^0 = A^0 B^0$  such that matrices  $A^0$  and  $B^0$  must have special zero entries. The more zeros  $A^0$  and  $B^0$  have the worse an associated error is. Matrices  $B_C^0$  and  $A_C^0$  have more non-zero entries than  $B_T^0$  and  $A_T^0$ , therefore,  $J_C^0$  is lesser than  $J_T^0$ . The error  $J_{GKLT}$  associated with the GKLT is naturally lesser than  $J_C^0$  and  $J_T^0$  because the conditions of causality and memory are not imposed on  $F_{GKLT}$  and, therefore, it is not required that  $F_{GKLT}$  must have specific zero entries.

## 5. Conclusion

The new results obtained in the paper are summarized as follows.

We have presented a new approach to the data processing consisting from compression, de-compression and filtering of observed stochastic data subject to the conditions of causality and variable memory. The approach is based on the assumption that certain covariance matrices formed from observed data, reference signal and compressed signal are known or can be estimated. This allowed us to consider two separate problem related to compression and



(a) Given reference signals.



(b) Observed data.



(c) Estimates of the reference signals by the filter  $F_C^0$  with the complete variable memory.



(d) Estimates of the reference signals by the filter  $F_T^0$  with the truncated variable memory.

Fig. 1. Illustration of simulation results.

de-compression subject to those constraints. It has been shown that the structure of the filter should have a special form due to constraints of causality and variable memory. The latter has implied the new method for the filter determination presented in Sections 3.2.1–3.2.4. The analysis of the associated errors has also been provided.



## 6. Appendix 1: Difficulties associated with some possible formalizations of the real-time data compression problem

In addition to the motivations presented in Section 1, we provide here some further, related arguments for considering the real-time data compression problem in the form (2)–(3).

Let us consider two seemingly natural ways to state the real-time data compression problem and show that they lead to significant difficulties. This observation, in particular, has motivated us to state the problem as it has been done in Section 2.3 above.

*The first way* is as follows. Let us denote by  $J(A, B)$ , the norm of the difference between the reference signal  $\mathbf{x}$  and its estimate  $\tilde{\mathbf{x}} = (A \circ B)(\mathbf{y})$ , constructed by de-compressor  $A : L^2(\Omega, \mathbb{R}^r) \rightarrow L^2(\Omega, \mathbb{R}^m)$  and compressor  $B : L^2(\Omega, \mathbb{R}^n) \rightarrow L^2(\Omega, \mathbb{R}^r)$  from the observed signal  $\mathbf{y}$ :

$$J(A, B) = \|\mathbf{x} - (A \circ B)(\mathbf{y})\|_{\Omega}^2. \quad (53)$$

The problem is to find  $B^0 : L^2(\Omega, \mathbb{R}^n) \rightarrow L^2(\Omega, \mathbb{R}^r)$  and  $A^0 : L^2(\Omega, \mathbb{R}^r) \rightarrow L^2(\Omega, \mathbb{R}^m)$  such that

$$J(A^0, B^0) = \min_{A, B} J(A, B) \quad (54)$$

*subject to conditions of causality and variable finite memory* for  $A$  and  $B$ .

The problem (54), with no constraints of causality and variable finite memory, has been considered, in particular, in (Hua and Nikpour, 1999).

*A second way* to formulate the problem is as follows. Let  $\mathcal{F} : L^2(\Omega, \mathbb{R}^n) \rightarrow L^2(\Omega, \mathbb{R}^m)$  be a linear operator defined by

$$[\mathcal{F}(\mathbf{y})](\omega) = F[\mathbf{y}(\omega)] \quad (55)$$

where  $F \in \mathbb{R}^{n \times m}$ . Let  $\text{rank } F = r$  and

$$J(F) = \|\mathbf{x} - \mathcal{F}(\mathbf{y})\|_{\Omega}^2.$$

Find  $\mathcal{F}^0 : L^2(\Omega, \mathbb{R}^n) \rightarrow L^2(\Omega, \mathbb{R}^m)$  such that

$$J(F^0) = \min_F J(F) \quad (56)$$

subject to

$$\text{rank } F \leq \min\{m, n\} \quad (57)$$

and *conditions of causality and variable finite memory* for  $F$ .

The constraint (57) implies that  $F$  can be written as a product of two matrices,  $A$  and  $B$ , representing a de-compressor and compressor, respectively.

If the conditions of causality and variable finite memory for  $F$  are not imposed then the solution of problem (56)–(57) is known and provided, for example, in (Torokhti and Howlett, 2007), Section 7.4.

If there are no constraints associated with causality and variable finite memory, then solutions of problems (54) and (56)–(57) are based on the assumption that certain covariance matrices are known. In this regard, see, for example, (Scharf, 1991)–(Jolliffe, 1986) and (Scharf, 1991)–(Torokhti and Friedland, 2009). In particular, the solution of problem (54), with no constraints of causality and variable finite memory, provided in (Hua and Nikpour, 1999) follows from an iterative scheme that requires knowledge of two covariance matrices at each step of the scheme. Thus, if the method proposed in (Hua and Nikpour, 1999) requires  $p$  iterative steps, then it requires a knowledge of  $2p$  covariance matrices.

In the case of no constraints of causality and variable finite memory, the known solution of problem (56)-(57) requires knowledge of two covariance matrices only (see, e.g., (Scharf, 1991)–(Jolliffe, 1986) and (Scharf, 1991)–(Torokhti and Friedland, 2009)).

A special difficulty with solving problem (54) is that it involves two unknowns,  $A$  and  $B$ , but only one functional to minimize. An iterative approach to its *approximate* solution based on the method presented in (Hua and Nikpour, 1999) requires knowledge of a number of covariance matrices. Another difficulty is that  $A$  and  $B$  must keep their special form associated with causality and variable finite memory. Related definitions are given in Section 3.1.

Unlike (54), problem (56)–(57) has only one unknown. Nevertheless, the main difficulty associated with problem (56)–(57) is similar: an implementation of the conditions of causality and variable finite memory into a structure of  $F$  implies a representation of  $F$  as a product of two factors, each with a specific shape related to causality and memory, respectively. An approach to its *exact* solution based on the methodology presented in (Scharf, 1991)–(Jolliffe, 1986) and (Scharf, 1991)–(Torokhti and Friedland, 2009) would require knowledge of two covariance matrices only, but implies constraints related to a special shape of each factor in the decomposition of  $F$  into a product of two factors. Therefore, as with problem (54), the difficulty again relates to dealing with two unknowns with only one functional to minimize.

To avoid the difficulties discussed above, we proposed the approach to the solution of the real-time data compression problem presented in Sections 2 and 3. We note that the problem (2)–(3) requires knowledge of covariance matrices used in Theorems 1 and 2, and Corollaries 1 and 2, i.e., such an assumption is similar to the assumptions used in (Scharf, 1991)–(Jolliffe, 1986) and (Scharf, 1991)–(Torokhti and Friedland, 2009) for the solution of problems (54) and (56)–(57) with no constraints associated with causality and variable finite memory.

## 7. Appendix 2: Some derivations used in Section 3.2

### 7.1 Proof of relationships (22) and (23)

Let us show that (22) and (23) are true. For simplicity, we denote

$$\mathbf{z} := \mathbf{z}_2, \quad S_1 := S_B, \quad S_2 := S_{B_2} \quad \text{and} \quad J(S_1, S_2) := J^{(2)}(S_B, S_{B_2}). \quad (58)$$

Let

$$J_0 = \|E_{zz}^{1/2}\|_F^2 - \sum_{k=1}^2 \|E_{zw_k}(E_{w_k w_k}^{1/2})^\dagger\|_F^2 \quad \text{and} \quad J_1 = \sum_{k=1}^2 \|S_k E_{w_k w_k}^{1/2} - E_{zw_k}(E_{w_k w_k}^{1/2})^\dagger\|_F^2.$$

**Lemma 3** Let  $J(S_1, S_2) = \|\mathbf{z} - \sum_{k=1}^2 S_k(\mathbf{w}_k)\|_\Omega^2$ . Then  $\min_{S_1, S_2} J(S_1, S_2)$  is achieved when

$$S_k = E_{zw_k} E_{w_k w_k}^\dagger + H_k(I - E_{w_k w_k} E_{w_k w_k}^\dagger) \quad \text{for } k = 1, 2, \quad (59)$$

where  $H_k$  is an arbitrary matrix.

**Proof.** First, let us prove that

$$J(S_1, S_2) = J_0 + J_1. \quad (60)$$

We note that  $\|M\|_F^2 = \text{tr}[MM^T]$ , and write

$$J_0 = \text{tr}[E_{zz}] - \sum_{k=1}^2 \text{tr}[E_{zw_k} E_{w_k w_k}^\dagger E_{w_k z}] \quad (61)$$

and

$$\begin{aligned} J_1 &= \sum_{k=1}^2 \text{tr}[(S_k - E_{zw_k} E_{w_k w_k}^\dagger) E_{w_k w_k} (S_k^T - E_{w_k w_k}^\dagger E_{w_k z})] \\ &= \sum_{k=1}^2 \text{tr}[(S_k E_{w_k w_k} S_k^T - S_k E_{w_k z} - E_{zw_k} S_k^T + E_{zw_k} E_{w_k w_k}^\dagger E_{w_k z})]. \end{aligned} \quad (62)$$

The latter is due to the relationship  $E_{w_k w_k}^\dagger E_{w_k w_k}^{1/2} = (E_{w_k w_k}^{1/2})^\dagger$  and orthogonality of vectors  $\mathbf{w}_1$  and  $\mathbf{w}_2$ .

Next, since for any random vector  $\mathbf{z}$ ,  $\|\mathbf{z}\|_\Omega^2 = \text{tr}[E_{zz}]$ , we have

$$J(S_1, S_2) = \|\mathbf{z} - \sum_{k=1}^2 S_k(\mathbf{w}_k)\|_\Omega^2 = \text{tr}[E_{zz} - \sum_{k=1}^2 (E_{zw_k} S_k^T - S_k E_{w_k z} + S_k E_{w_k w_k} S_k^T)]. \quad (63)$$

Then (60) follows from (61)–(63).

In (60), the only term depending on  $S_1$  and  $S_2$ , is  $J_1$ . Therefore,  $\min_{S_1, S_2} J(S_1, S_2)$  is achieved when

$$S_k E_{w_k w_k}^{1/2} - E_{zw_k} (E_{w_k w_k}^{1/2})^\dagger = \mathbf{O} \quad \text{for } k = 1, 2. \quad (64)$$

By (Torokhti and Howlett, 2007) (see p. 352), matrix equation (64) has its general solutions in the form (59).  $\square$

As a result, relationships (22) and (23) follow from (58) and (59).

## 7.2 Proof of equivalency of equations (33) and (34)

In Lemma 4 below,  $T_B$  is any matrix of the size that makes sense of the difference  $T_B - E_{z_1 w_1} E_{w_1 w_1}^\dagger$ . In particular,  $T_B$  can be the matrix defined by (17).

**Lemma 4** *The equations*

$$(T_B - E_{z_1 w_1} E_{w_1 w_1}^\dagger) E_{w_1 w_1}^{1/2} = \mathbf{O}. \quad (65)$$

and

$$T_B E_{w_1 w_1} - E_{z_1 w_1} = \mathbf{O}. \quad (66)$$

are equivalent.

**Proof.** Let us suppose that (65) is true. Multiplying on the right by  $E_{w_1 w_1}^{1/2}$  gives

$$T_B E_{w_1 w_1} - E_{z_1 w_1} E_{w_1 w_1}^\dagger E_{w_1 w_1} = \mathbf{O}.$$

Then  $T_B E_{w_1 w_1} - E_{z_1 w_1} = \mathbf{O}$  follows on the basis of  $E_{z_1 w_1} E_{w_1 w_1}^\dagger E_{w_1 w_1} = E_{z_1 w_1}$  (see Lemma 24 on p. 168 in (Torokhti and Howlett, 2007)).

On the other hand, if  $T_B E_{w_1 w_1} - E_{z_1 w_1} E_{w_1 w_1}^\dagger E_{w_1 w_1} = \mathbf{O}$  then multiplying on the right by  $E_{w_1 w_1}^\dagger$  gives

$$T_B E_{w_1 w_1} E_{w_1 w_1}^\dagger - E_{z_1 w_1} E_{w_1 w_1}^\dagger = \mathbf{O}. \quad (67)$$

Here,  $E_{w_1 w_1}^{1/2} E_{w_1 w_1}^{1/2\dagger} = E_{w_1 w_1} E_{w_1 w_1}^\dagger$  (see p. 170 in (Torokhti and Howlett, 2007)). Hence, equation (67) can be rewritten as

$$T_B E_{w_1 w_1}^{1/2} E_{w_1 w_1}^{1/2\dagger} - E_{z_1 w_1} E_{w_1 w_1}^\dagger = \mathbf{O}.$$

Multiplying on the right by  $E_{w_1 w_1}^{1/2}$  gives the required result (66).  $\square$

In other words, (33) and (34) are equivalent.

## 8. References

- L. L. Scharf, *Statistical Signal Processing*, Reading, MA: Addison-Wesley, 1991.
- Y. Hua, M. Nikpour and P. Stoica, Optimal reduced rank estimation and filtering, *IEEE Trans. Signal Processing*, Vol. 49, No. 3, pp. 457-469, 2001.
- K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Boston: Academic Press, 1990.
- I.T. Jolliffe, *Principal Component Analysis*, Springer Verlag, New York, 1986.
- V. Gimeno, Obtaining the EEG envelope in real time: a practical method based on homomorphic filtering, *Neuropsychobiology*, Vol. 18, No. 2, pp. 110-112, 1987.
- H. Kim, G.H. Golub, H. Park, Missing value estimation for DNA microarray gene expression data: local least squares imputation, *Bioinformatics*, Vol. 21, No. 2, pp. 187-198 2005.
- P. Common, G.H. Golub, Tracking a few extreme singular values and vectors in signal processing, *Proc. IEEE*, Vol. 78, No. 8, pp.1327-1343, 1990.
- S. Kraut, Member, R. H. Anderson, and J. L. Krolik, A Generalized KarhunenLoeve Basis for Efficient Estimation of Tropospheric Refractivity Using Radar Clutter, *IEEE Trans. Signal Processing*, Vol. 52, No. 1, pp. 48 -60, 2004.
- A. Torokhti and P. Howlett, *Computational Methods for Modelling of Nonlinear Systems*, Elsevier, 2007.
- L. L. Scharf, The SVD and reduced rank signal processing, *Signal Processing*, Vol. 25, pp. 113 - 133, 1991.
- Y. Yamashita and H. Ogawa, Relative Karhunen-Loève transform, *IEEE Trans. on Signal Processing*, Vol. 44, No. 2, pp. 371-378, 1996.
- Y. Hua and W. Q. Liu, Generalized Karhunen-Loève transform, *IEEE Signal Processing Letters*, Vol. 5, No. 6, pp. 141-143, 1998.
- A. Torokhti and S. Friedland, Towards theory of generic Principal Component Analysis, *J. Multivariate Analysis*, Vol. 100, No. 4, pp. 661-669, 2009.
- A. Torokhti and P. Howlett, Optimal Transform Formed by a Combination of Nonlinear Operators: The Case of Data Dimensionality Reduction, *IEEE Trans. on Signal Processing*, Vol. 54, No. 4, 2006.
- A. Torokhti and P. Howlett, Filtering and Compression for Infinite Sets of Stochastic Signals, *Signal Processing*, Vol. 89, Issue 3, pp. 291-304, 2009.
- A. Torokhti and J. Manton, Generic Weighted Filtering of Stochastic Signals, *IEEE Trans. Signal Processing* (accepted 23 January 2009).
- Y. Hua and M. Nikpour, Computing the reduced rank Wiener filter by IQMD, *IEEE Signal Processing Letters*, Vol. 6, No. 9, pp. 240-242, 1999.
- P. Stoica and M. Viberg, Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions, *IEEE Trans. Signal Processing*, Vol. 44, No. 12, pp. 3069-3078, 1996.
- T. Zhang, G. Golub, Rank-One Approximation to High Order Tensors, *SIAM J. Matrix Anal. Appl.*, Vol. 23, Issue 2, pp. 534-550, 2001.
- V.N. Fomin and M.V. Ruzhansky, Abstract optimal linear filtering, *SIAM J. Control and Optimization*, 2 Vol. 38, pp. 1334 - 1352, 2000.
- A. Torokhti and P. Howlett, Best approximation of identity mapping: the case of variable memory, *J. Approx. Theory*, Vol. 143, No. 1, pp. 111-123, 2006.
- P.G. Howlett, A. Torokhti, and C.E.M. Pearce, Optimal multilinear estimation of a random vector under constraints of causality and limited memory, *Computational Statistics & Data Analysis*, Vol. 52, Issue 2, pp. 869-878, 2007.

- R. M. De Santis, Causality Theory in Systems Analysis, *Proc. of IEEE*, Vol. 64, Issue 1, pp. 36–44, 1976.
- G. H. Golub and C. F. Van Loan, *Matrix Computations*, Baltimore, MD: Johns Hopkins University Press, 1996.
- A. Ben-Israel and T. N. E. Greville, *Generalized Inverses: Theory and Applications*, John Wiley & Sons, New York, 1974.

IntechOpen

IntechOpen



## **Advanced Technologies**

Edited by Kankesu Jayanthakumaran

ISBN 978-953-307-009-4

Hard cover, 698 pages

**Publisher** InTech

**Published online** 01, October, 2009

**Published in print edition** October, 2009

This book, edited by the Intech committee, combines several hotly debated topics in science, engineering, medicine, information technology, environment, economics and management, and provides a scholarly contribution to its further development. In view of the topical importance of, and the great emphasis placed by the emerging needs of the changing world, it was decided to have this special book publication comprise thirty six chapters which focus on multi-disciplinary and inter-disciplinary topics. The inter-disciplinary works were limited in their capacity so a more coherent and constructive alternative was needed. Our expectation is that this book will help fill this gap because it has crossed the disciplinary divide to incorporate contributions from scientists and other specialists. The Intech committee hopes that its book chapters, journal articles, and other activities will help increase knowledge across disciplines and around the world. To that end the committee invites readers to contribute ideas on how best this objective could be accomplished.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

A. Torokhti and S. Miklavcic (2009). Data Compression and De-compression by Causal Filters with Variable Finite Memory, Advanced Technologies, Kankesu Jayanthakumaran (Ed.), ISBN: 978-953-307-009-4, InTech, Available from: <http://www.intechopen.com/books/advanced-technologies/data-compression-and-de-compression-by-causal-filters-with-variable-finite-memory>

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821



© 2009 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen