FILTERING AND COMPRESSION OF STOCHASTIC SIGNALS UNDER CONSTRAINT OF VARIABLE FINITE MEMORY

Anatoli Torokhti and Stan Miklavcic

University of South Australia, Adelaide, Australia

Abstract: We study a new technique for optimal data compression subject to conditions of causality and different types of memory. The technique is based on the assumption that certain covariance matrices formed from observed data, reference signal and compressed signal are known or can be estimated. In particular, such an information can be obtained from the known solution of the associated problem with no constraints related to causality and memory. This allows us to consider two separate problems related to compression and de-compression subject to those constraints. Their solutions are given and the analysis of the associated errors is provided.

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. While the topics have been intensively studied (see e.g. (S. Friedland, 2006), (Jolliffe, 1986), (Hua and Nikpour, 1999), (Hua and Liu, 1998), (A. Torokhti, 2001), (Torokhti and Howlett, 2007), (T. Zhang, 2001)), a number of related fundamental questions are still open. One of them concerns specific restrictions associated with different types of causality and memory.

First Motivation: Causality and Memory. Data compression techniques mainly consist of three operations, compression itself, de-noising and de-compression (or reconstruction) of the compressed data. Each operation is implemented by a special filter. In reality, a value of the output of such a filter at time t_k is determined from a 'fragment' of its input defined at times $t_k, t_{k-1}, \ldots, t_{k-q}$. In other words, in practice both operations are subject to the conditions of causality and memory.

Our first motivation comes from a real-time signal processing. This implies that the filters we propose should be causal with variable finite memory.

Second Motivation: Reformulation of the Problem. Let (Ω, Σ, μ) be 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 $\mu(\Omega) = 1$.

In an informal way, the data compression problem we consider can be expressed as follows. 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, (a) the data $\mathbf{y}$ should be compressed to a 'shorter' vector $\mathbf{z} \in L^2(\Omega, \mathbb{R}^r)$[1] with $r < \min\{m, n\}$ and (b) $\mathbf{z}$ should be de-compressed (reconstructed) to a signal $\tilde{\mathbf{x}} \in L^2(\Omega, \mathbb{R}^m)$ that is 'close' to $\mathbf{x}$ in some appropriate sense. Both operations should be *causal* and have *variable finite memory*. In this paper, the term 'close' is used with respect to the minimum of the norm (2) of the difference between $\mathbf{x}$ and $\tilde{\mathbf{x}}$.

The problem can be formulated in several alternate ways.

The first way is as follows. Let $\mathcal{B} : L^2(\Omega, \mathbb{R}^n) \to L^2(\Omega, \mathbb{R}^r)$ signify compression so that $\mathbf{z} = \mathcal{B}(\mathbf{y})$ and let $\mathcal{A} : L^2(\Omega, \mathbb{R}^r) \to L^2(\Omega, \mathbb{R}^m)$ 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

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

where $B \in \mathbb{R}^{n \times r}$ and $A \in \mathbb{R}^{r \times m}$. In the remainder of

---

[1]Components of $\mathbf{z}$ are often called *principal components* (Jolliffe, 1986).

this paper we shall use the same symbol to represent both the linear operator acting on a random vector and its associated matrix.

We define the norm to be

$$\|\mathbf{x}\|_{\Omega}^2 = \int_{\Omega} \|\mathbf{x}(\omega)\|_2^2 d\mu(\omega) \qquad (2)$$

where $\|\mathbf{x}(\omega)\|_2$ is the Euclidean norm of $\mathbf{x}(\omega)$. Let us denote by $J(A,B)$, the norm of the difference between $\mathbf{x}$ and $\tilde{\mathbf{x}}$, constructed by A and B:

$$J(A,B) = \|\mathbf{x} - (A \circ B)(\mathbf{y})\|_{\Omega}^2. \qquad (3)$$

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

$$J(A^0, B^0) = \min_{A,B} J(A,B) \qquad (4)$$

subject to conditions of causality and variable finite memory for A and B. The problem consists of two unknowns, A and B.

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

$$[\mathcal{F}(\mathbf{y})](\omega) = F[\mathbf{y}(\omega)] \qquad (5)$$

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

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

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

$$J(F^0) = \min_{F} J(F) \qquad (6)$$

subject to

$$\text{rank } F \le \min\{m, n\} \qquad (7)$$

and conditions of causality and variable finite memory for F. Unlike (4), the problem (6)–(7) has only one unknown.

2 STATEMENT OF THE PROBLEM

The basic idea of our approach is as follows.

Let $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$, $\mathbf{y} \in L^2(\Omega, \mathbb{R}^n)$ and $\mathbf{z} \in L^2(\Omega, \mathbb{R}^r)$, and let A and B be defined as (1) below. Here, $\mathbf{z}$ is a compressed version of $\mathbf{x}$. 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 problem (6)-(7) *with no constraints associated with causality and memory.*

In this paper, the data compression problem *subject to conditions of causality and memory* is stated in the form of two separate problems, (8) and (10) formulated below.

We use the following notation: $\mathcal{M}(r, n, \eta_B)$ is a set of causal $r \times n$ matrices B with a so-called *complete* variable finite memory η_B. The notation $\mathcal{M}(m, r, \eta_A)$ is similar.

Consider

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

Let B^0 be such that

$$J_1(B^0) = \min_{B} J_1(B) \quad \text{subject to } B \in \mathcal{M}(r, n, \eta_B).$$
$$(8)$$

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

$$J_2(A) = \|\mathbf{x} - A(\mathbf{z}^0)\|_{\Omega}^2 \qquad (9)$$

and let A^0 be such that

$$J_2(A^0) = \min_{A} J_2(A) \quad \text{subject to } A \in \mathcal{M}(m, r, \eta_A).$$
$$(10)$$

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 (8) and (10), respectively.

The major differences between the above statement of the problem and the statements considered below are as follows.

First, A and B should be causal with variable finite memory.

Second, it is assumed that certain covariance matrices formed from $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are known or can be estimated. In particular, such information can be obtained from the known solution (Torokhti and Howlett, 2007) of problem (6)-(7) with no constraints associated with causality and memory. We note that such an assumption does not look too restrictive in comparison with the assumptions used in the associated methods (Hua and Nikpour, 1999)–(Torokhti and Howlett, 2007).

Consequently and thirdly, we represent the initial problem in the form of a concatenation of two new separate problems (8) and (10).

3 MAIN RESULTS

Let $\tau_1 < \tau_2 < \cdots < \tau_n$ be time instants and $\alpha, \beta, \vartheta : \mathbb{R} \to L^2(\Omega, \mathbb{R})$ be continuous functions. Sup-

pose $\alpha_k = \alpha(\tau_k)$, $\beta_k = \beta(\tau_k)$ and $\vartheta_k = \vartheta(\tau_k)$ are real-valued random variables having finite second moments. We write $\mathbf{x} = [\alpha_1, \alpha_2, \ldots, \alpha_m]^T$ $\mathbf{y} = [\beta_1, \beta_2, \ldots, \beta_n]^T$ and $\mathbf{z} = [\vartheta_1, \ldots, \vartheta_r]^T$.

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

In many applications[2], to obtain $\tilde{\vartheta}_k$ for $k = 1, \ldots, r$, it is necessary for B to use only a limited number of input components, $\eta_{B_k} = 1, \ldots, r$. A number of such input components η_{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 $1 \leq p \leq n$ and $n - r + 2 \leq g \leq n$.

Definition 1. *The vector $\eta_B = [\eta_{B_1}, \ldots, \eta_{B_r}]^T \in \mathbb{R}^r$ is called* a variable memory *of the compressor B. In particular, η_B is called a* complete *variable memory if $\eta_{B_1} = g$ and $\eta_{B_k} = n$ when $k = n - g + 1, \ldots, n$. 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 $\eta_A = [\eta_{A_1}, \ldots, \eta_{A_m}]^T \in \mathbb{R}^m$ has a similar meaning for the de-compressor A, *i.e.*, η_A is *a variable memory* of the de-compressor A. Here, η_{A_j} is the *jth local memory* of A.

The parameters q and s, which are positive integers such that $1 \leq q \leq r$ and $2 \leq s \leq m$, are used below to define two types of memory for A.

Definition 2. *Vector η_A is called a* complete *variable memory of the de-compressor A if $\eta_{A_1} = q$ and $\eta_{A_j} = r$ when $j = s + r - 1, \ldots, m$. 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

$$b_{i,j} = 0$$
$$\text{if} \quad j = p - r + i + 1, \ldots, n,$$
$$\text{for} \quad \begin{cases} p = r, \ldots, n-1, \\ i = 1, \ldots, r \end{cases} \text{and} \begin{cases} p = n, \\ i = 1, \ldots, r-1, \end{cases}$$

[2]Examples include computer medical diagnostics (Gimeno, 1987) and problems of bio-informatics (H. Kim, 2005).

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

$$b_{i,j} = 0$$
$$\text{if} \begin{cases} i = 1, \ldots, r-p, \\ j = 1, \ldots, n, \end{cases} \text{and} \begin{cases} i = r - p + 1, \ldots, r, \\ j = i - r + p + 1, \ldots, n. \end{cases}$$

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

$$a_{i,j} = 0 \tag{11}$$
$$\text{if } j = q + i, \ldots, r \text{ for } q = 1, \ldots, r - 1, i = 1, \ldots, r - q,$$

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

$$a_{i,j} = 0$$
$$\text{if } j = s + i, \ldots, r \text{ for } s = 1, \ldots, m, i = 1, \ldots, s + r - 1,$$

The above conditions imply the following definitions.

Definition 3. *A matrix B satisfying the constraint (11)–(11) is said to be a causal operator with the* complete *variable memory $\eta_B = [g, g+1, \ldots, n]^T$. Here, $\eta_{B_k} = n$ when $k = n - g + 1, \ldots, n$. The set of such matrices is denoted by $\mathcal{M}_C(r, n, \eta_B)$.*

Definition 4. *A matrix A satisfying the constraint (11)–(11) is said to be a causal operator with the* complete *variable memory $\eta_A = [r - q + 1, \ldots, r]^T$. Here, $\eta_{A_j} = r$ when $j = q, \ldots, m$. The set of such matrices is denoted by $\mathcal{M}_C(m, r, \eta_A)$.*

3.1 Solution of Problems (8) and (10)

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

$$\langle \alpha_i, \beta_j \rangle = \int_\Omega \alpha_i(\omega) \beta_j(\omega) d\mu(\omega) < \infty, \tag{12}$$

$$E_{xy} = \{\langle \alpha_i, \beta_j \rangle\}_{i,j=1}^{m,n} \in \mathbb{R}^{m \times n},$$

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

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

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

Lemma 1. (Torokhti and Howlett, 2007) *If we define $\mathbf{w}_1 = \mathbf{y}_1$ and $\mathbf{w}_2 = \mathbf{y}_2 - P_y \mathbf{y}_1$ where*

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

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

Let us first consider problem (8) when B has the complete variable memory $\eta_B = [g, g+1, \ldots, n]^T$ (see Definition 3).

Let us partition B into four blocks K_B, L_B, S_{B1} and S_{B2} so that $B = \begin{bmatrix} K_B & L_B \\ S_{B1} & S_{B2} \end{bmatrix}$, where

$K_B = \{k_{ij}\} \in \mathbb{R}^{n_b \times (g-1)}$ is a rectangular matrix,

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

$S_{B1} = \{s_{ij}^{(1)}\} \in \mathbb{R}^{(r-n_b) \times (g-1)}$,

$S_{B2} = \{s_{kl}^{(2)}\} \in \mathbb{R}^{(r-n_b) \times n_b}$

are rectangular matrices, and $\quad n_b = n - g + 1$.

We have $B(\mathbf{y}) = \begin{bmatrix} T_B(\mathbf{w}_1) + L_B(\mathbf{w}_2) \\ S_B(\mathbf{w}_1) + S_{B2}(\mathbf{w}_2) \end{bmatrix}$, where $T_B = K_B + L_B P_y$ and $S_B = S_{B1} + S_{B2} P_y$. Then

$$J_1(B) = J^{(1)}(T_B, L_B) + J^{(2)}(S_B, S_{B2}), \qquad (16)$$

where $J^{(1)}(T_B, L_B) = \|\mathbf{z}_1 - [T_B(\mathbf{w}_1) + L_B(\mathbf{w}_2)]\|_\Omega^2$, $J^{(2)}(S_B, S_{B2}) = \|\mathbf{z}_2 - [S_B(\mathbf{w}_1) + S_{B2}(\mathbf{w}_2)]\|_\Omega^2$. By analogy with Lemma 37 in (Torokhti and Howlett, 2007),

$$\min_{B \in \mathcal{M}(r,n,\eta_B)} J_1(B) = \min_{T_B, L_B} J^{(1)}(T_B, L_B) + \min_{S_B, S_{B2}} J^{(2)}(S_B, S_{B2}).$$

Therefore, problem (8) is reduced to finding matrices T_B^0, L_B^0, S_B^0 and S_{B2}^0 such that

$$J^{(1)}(T_B^0, L_B^0) = \min_{T_B, L_B} J^{(1)}(T_B, L_B) \qquad (17)$$

and

$$J^{(2)}(S_B^0, S_{B2}^0) = \min_{S_B, S_{B2}} J^{(2)}(S_B, S_{B2}). \qquad (18)$$

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_{B2}^0 are given by

$$S_B^0 = E_{z_2} E_{w_1 w_1}^\dagger + H_B(I - E_{w_1 w_1} E_{w_1 w_1}^\dagger) \qquad (19)$$

and

$$S_{B2}^0 = E_{z_2} E_{w_2 w_2}^\dagger + H_{B2}(I - E_{w_2 w_2} E_{w_2 w_2}^\dagger), \qquad (20)$$

where H_B and H_{B2} are arbitrary matrices.

Next, to find T_B^0 and L_B^0 we use the following notation.

For $r = 1, 2, \ldots, \ell$, let ρ 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

$$E_{w_2 w_2}^{1/2} = Q_{w,\rho} R_{w,\rho} \qquad (21)$$

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 ρ. We write $G_{w,\rho} = R_{w,\rho}^T$ and

use the notation $G_{w,\rho} = [g_1, \ldots, g_\rho] \in \mathbb{R}^{n_2 \times \rho}$ where $g_j \in \mathbb{R}^{n_2}$ denotes the j-th column of $G_{w,\rho}$. We also write $G_{w,s} = [g_1, \ldots, g_s] \in \mathbb{R}^{n_2 \times s}$ for $s \leq \rho$ to denote the matrix consisting of the first s columns of $G_{w,\rho}$. For simplicity, let us denote this $G_s := G_{w,s}$. Next, let $\mathbf{e}_1^T = [1, 0, 0, 0, \ldots], \quad \mathbf{e}_2^T = [0, 1, 0, 0, \ldots], \quad \mathbf{e}_3^T = [0, 0, 1, 0, \ldots],$ etc. denote the unit row vectors irrespective of the dimension of the space.

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

Theorem 1. *Let $B \in \mathcal{M}_C(r, n, \eta_B)$, i.e., the compressor B is causal and has the complete variable memory $\eta_B = [g, g+1, \ldots, n]^T$. Then the solution to problem (8) is provided by the matrix B^0, which has the form $B^0 = \begin{bmatrix} K_B^0 & L_B^0 \\ S_{B1}^0 & S_{B2}^0 \end{bmatrix}$, where the blocks $K_B^0 \in \mathbb{R}^{n_b \times (g-1)}$, $S_{B1}^0 \in \mathbb{R}^{(r-n_b) \times (g-1)}$ and $S_{B2}^0 \in \mathbb{R}^{(r-n_b) \times n_b}$ 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*

$$K_B^0 = T_B^0 - L_B^0 P_y \qquad (22)$$

with

$$T_B^0 = E_{z_1 w_1} E_{w_1 w_1}^\dagger + N_{B1}(I - E_{w_1 w_1} E_{w_1 w_1}^\dagger) \qquad (23)$$

where N_{B1} is an arbitrary matrix. The block $L_B^0 = \begin{bmatrix} \lambda_1^0 \\ \vdots \\ \lambda_{n_b}^0 \end{bmatrix}$, for each $s = 1, 2, \ldots, n_2$, is defined by its rows

$$\lambda_s^0 = \mathbf{e}_s^T E_{z_1 w_2} E_{w_2 w_2}^\dagger G_s G_s^\dagger + f_s^T (I - G_s G_s^\dagger) \quad (24)$$

with $f_s^T \in \mathbb{R}^{1 \times n_2}$ arbitrary. The blocks S_{B1}^0 and S_{B2}^0 are given by

$$S_{B1}^0 = S_B^0 - S_{B2}^0 P_y \qquad (25)$$

and (20), respectively. In (25), S_B^0 is presented by (19). The error associated with the compressor B^0 is given by

$$\|\mathbf{z} - B^0 \mathbf{y}\|_\Omega^2 = \sum_{s=1}^\rho \sum_{j=s+1}^{n_2} |e_s^T E_{z_1 w_2} E_{w_2 w_2}^\dagger g_j|^2$$

$$+ \sum_{j=1}^2 \|E_{z_j z_j}^{1/2}\|_F^2 - \sum_{i=1}^2 \sum_{j=1}^2 \|E_{z_i w_i} E_{w_j w_j}^{\dagger 1/2}\|_F^2. \quad (26)$$

Let us now consider problem (10) when the decompressor A has the complete variable memory $\eta_A = [r - q + 1, \ldots, r]^T$ (see Definition 4).

In analogy with our partitioning of matrix B, we partition matrix A in four blocks 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

$S_{A1} = \{s_{ij}^{(1)}\} \in \mathbb{R}^{(m-q) \times (r-q)}$,

$S_{A2} = \{s_{kl}^{(2)}\} \in \mathbb{R}^{(m-q) \times q}$

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

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

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.

We write $G_{v,s} = [g_1, \ldots, g_s] \in \mathbb{R}^{q \times s}$ 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 $A \in \mathcal{M}_C(m, r, \eta_A)$, i.e. the de-compressor A is causal and has the complete variable memory $\eta_A = [r-q+1, \ldots, r]^T$. Then the solution to problem* (10) *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*

$$K_A^0 = T_A^0 - L_A^0 P \tag{27}$$

with

$$T_A^0 = E_{x_1 v_1} E_{v_1 v_1}^\dagger + N_{A1}(I - E_{v_1 v_1} E_{v_1 v_1}^\dagger) \tag{28}$$

where N_1 is an arbitrary matrix. The block $L_A^0 = \begin{bmatrix} \lambda_1^0 \\ \vdots \\ \lambda_q^0 \end{bmatrix}$, for each $s = 1, 2, \ldots, q$, is defined by its rows

$$\lambda_s^0 = \mathbf{e}_s^T E_{x_1 v_2} E_{v_2 v_2}^\dagger G_s G_s^\dagger + f_s^T(I - G_s G_s^\dagger) \tag{29}$$

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

$$S_{A1}^0 = S_A^0 - S_{A2}^0 P, \quad S_{A2}^0 = E_{x_2} E_{v_2 v_2}^\dagger + H_{A2}(I - E_{v_2 v_2} E_{v_2 v_2}^\dagger), \tag{30}$$

where

$$S_A^0 = E_{x_2} E_{v_1 v_1}^\dagger + H_A(I - E_{v_1 v_1} E_{v_1 v_1}^\dagger) \tag{31}$$

and H_{A2} and H_A are arbitrary matrices.

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

$$\|\mathbf{x} - A^0 \mathbf{z}^0\|_\Omega^2 = \sum_{s=1}^\rho \sum_{j=s+1}^q |e_s^T E_{x_1 v_2} E_{v_2 v_2}^\dagger g_j|^2 \tag{32}$$

$$+ \sum_{j=1}^2 \|E_{x_j x_j}^{1/2}\|_F^2 - \sum_{i=1}^2 \sum_{j=1}^2 \|E_{x_i v_i} E_{v_j v_j}^{\dagger 1/2}\|_F^2. \tag{33}$$

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[3] presented in Fig. 1.

We divide X into 128 sub-matrices $X_{ij} \in \mathbb{R}^{m \times q}$ with $i = 1, \ldots, 16$, $j = 1, \ldots, 8$, $m = 16$ 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 $i = 1, \ldots, 16$ and $j = 1, \ldots, 8$, 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 16×32 matrix whose randomly-chosen elements are uniformly distributed in the interval $(0, 1)$.

The proposed filter F^0 has been applied to each pair $\{X_{ij}, Y_{ij}\}$. Each pair $\{X_{ij}, Y_{ij}\}$ was processed by compressors and de-compressors with the complete variable memory. 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

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

---

[3]The database is available in http://sipi.usc.edu/services/database/Database.html.

(a) Given reference signals.
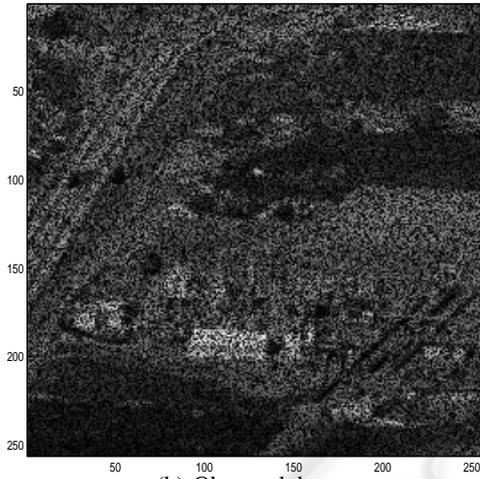
(b) Observed data.
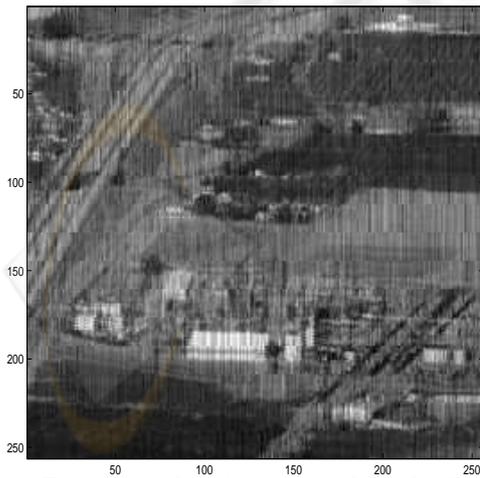
(c) Estimates of the reference signals by the filter F_C^0 with the complete variable memory.

Figure 1: Illustration of simulation results.

where $n = m = 16$, $r = 8$, $\eta_B = \{\eta_{Bk}\}_{k=1}^{16}$ with $\eta_{Bk} =$

$$\begin{cases} 12 + k - 1, & \text{if } k = 1, \ldots, 4, \\ 16, & \text{if } k = 5, \ldots, 16 \end{cases},$$

and $\eta_A = \{\eta_{Aj}\}_{j=1}^{16}$ with $\eta_{Aj} =$

$$\begin{cases} 6 + j - 1, & \text{if } j = 1, 2, \\ 8, & \text{if } k = 3, \ldots, 16 \end{cases}.$$ 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 $i = 1, \ldots, 16$ and $j = 1, \ldots, 8$. The compression ratio was $c = 1/2$. We obtained $J_C^0 = 3.3123e + 005$.

The results of simulations a are presented in Fig. 1 (a) - (c).

REFERENCES

A. Torokhti, P. H. (2001). Optimal fixed rank transform of the second degree. *IEEE Trans. Circuits & Syst., II, Analog & Digit. Signal Processing*, 48(3):309–315.

Gimeno, V. (1987). Obtaining the eeg envelope in real time: a practical method based on homomorphic filtering. *Neuropsychobiology*, 18:110–112.

Golub, G. and Loan, C. V. (1996). *Matrix Computation*. Johns Hopkins Univ. Press.

H. Kim, G.H. Golub, H. P. (2005). Missing value estimation for dna microarray gene expression data: local least squares imputation. *Bioinformatics*, 21:211–218.

Hua, Y. and Liu, W. Q. (1998). Generalized karhunen-loève transform. *IEEE Signal Process Letters*, 5:141–143.

Hua, Y. and Nikpour, M. (1999). Computing the reduced rank wiener filter by iqmd. *IEEE Signal Processing Letters*, 6(9):240–242.

Jolliffe, I. (1986). *Principal Component Analysis*. Springer Verlag.

S. Friedland, A. Niknejad, M. K. H. Z. (2006). Fast monte-carlo low rank approximations for matrices. *Proc. IEEE Conference SoSE*, pages 218–223.

T. Zhang, G. G. (2001). Rank-one approximation to high order tensors. *SIAM J. Matrix Anal. Appl.*, 23.

Torokhti, A. and Howlett, P. (2007). *Computational Methods for Modelling of Nonlinear Systems*. Elsevier.