

# Optimal Data Compression and Filtering: the Case of Infinite Signal Sets

Anatoli Torokhti and Phil Howlett

**Abstract**—We present a theory for optimal filtering of infinite sets of random signals. There are several new distinctive features of the proposed approach. First, we provide a *single* optimal filter for processing any signal from a given *infinite* signal set. Second, the filter is presented in the special form of a sum with  $p$  terms where each term is represented as a combination of three operations. Each operation is a special stage of the filtering aimed at facilitating the associated numerical work. Third, an iterative scheme is implemented into the filter structure to provide an improvement in the filter performance at each step of the scheme. The final step of the concerns signal compression and decompression. This step is based on the solution of a new rank-constrained matrix approximation problem. The solution to the matrix problem is described in this paper. A rigorous error analysis is given for the new filter.

**Keywords**—stochastic signals, optimization problems in signal processing.

## I. INTRODUCTION

### A. Motivation

**I**N this paper, we consider extensions of known approaches to optimal filtering based on the Wiener idea<sup>1</sup>. We present a theory for a new nonlinear filter which processes *infinite* sets of random signals. The filter is constructed via an iterative scheme that provides a signal processing improvement with each step. The filter provides simultaneous signal filtering and compression and the subsequent decompression (reconstruction).

There has been significant attention in the literature to filters that process finite sets of random signals but it seems that a filter which is able to process *infinite* sets of random signals has not been developed. The filter presented in this paper is designed specifically to process infinite sets of random signals. For the case of *finite* sets of random signals, we show that our filter leads to a lower computational load and better accuracy than the known filters; the improved accuracy is due to the special iteration procedure incorporated into the filter structure (see Section III-E2).

There are three motivations for the proposed method which we now describe.

1) *First motivation: infinite sets of signals:* Most of the literature on Wiener-like filtering provides an optimal filter for an *individual* input signal given by a *finite* random vector<sup>2</sup>. This means that if we wish to transform an infinite set  $Y = \{\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(N)}, \dots\}$  of input vector signal into an

infinite set  $X = \{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(N)}, \dots\}$  of output vector signals using a Wiener-like approach then we have to find a set of corresponding Wiener filters  $\{\mathcal{F}_{(1)}, \mathcal{F}_{(2)}, \dots, \mathcal{F}_{(N)}, \dots\}$  so that each representative  $\mathcal{F}_i$  of the filter set relates to a representative  $\mathbf{y}_{(i)}$  of the signal-vector set  $Y$ . Therefore such a filter cannot be applied if  $X$  and  $Y$  are infinite sets of signals. Moreover, in some situations, a *recognizer* must be used that will determine to which of the filters  $\{\mathcal{F}_{(1)}, \mathcal{F}_{(2)}, \dots, \mathcal{F}_{(N)}, \dots\}$  each component from  $Y$  should be directed.

Note that even in the case when  $Y$  and  $X$  are finite sets,  $Y = \{\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(N)}\}$  and  $X = \{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(N)}\}$  where  $Y$  and  $X$  can be represented as finite vectors, the Wiener approach leads to computation of large covariance matrices. Indeed, if each  $\mathbf{y}_i$  has  $n$  components and each  $\mathbf{x}_i$  has  $m$  components then the Wiener approach leads to computation of a product of an  $mN \times nN$  matrix and an  $nN \times nN$  matrix and computation of an  $nN \times nN$  pseudo-inverse matrix [18]. This requires  $O(2mn^2N^3)$  and  $O(22n^3N^3)$  flops, respectively [7]. If  $m$ ,  $n$  and  $N$  are sufficiently large then the computational work associated with this approach becomes unreasonably hard.

To avoid such drawbacks here, we here study an approach that allows us to use *only one* filter to process any signal from the infinite set  $Y$ .

*The first question* we address in the paper is as follows. Let  $X$  and  $Y$  be *infinite* sets of signals. How should we construct a *single* optimal filter  $\mathcal{F} : Y \rightarrow X$  which can be applied to each pair of signals  $(\mathbf{x}, \mathbf{y}) \in X \times Y$  and which, moreover, transforms each  $\mathbf{y}$  to a corresponding  $\mathbf{x}$  with associated minimal error?

Surprisingly, perhaps, the answer is based firstly, on an equivalent alternative signal representation in a different space and secondly, on the use of a special norm (3) in the statement of the problem. The dual representation means that  $\mathbf{x}$  is considered as a single signal in one representation, and on the other hand, as an infinite set of signals in the other, original, representation. A detailed explanation is given in Section VI. Examples of different special cases of the norm (3) used in our statement of the problem are presented in Section VI.

The answer for the first question is provided in Sections II-A, II-C, in Theorems 3 and 4, and in Section III-E. The special norm is given by (3) below.

2) *Second motivation: improvement in the filter performance:* The performance of filters used for data filtering, compression and subsequent reconstruction, is characterized by the accuracy, the compression ratio and the related computational load. The Karhunen-Loève filter (KLF) [13], [14],

Anatoli Torokhti and Phil Howlett are with the School of Mathematics and Statistics, University of South Australia, Mawson Lakes, SA 5095, Australia, email: anatoli.torokhti@unisa.edu.au and phil.howlett@unisa.edu.au.

<sup>1</sup>Some references on Wiener-like filtering can be found in [8], [12], [16], [18], [19].

<sup>2</sup>We say a random vector  $\mathbf{x}$  is finite if each realization  $\mathbf{x} = \mathbf{x}(\omega)$  has a finite number of scalar components.

[18]<sup>3</sup> is known to be the *optimal* filter that minimizes the reconstruction error over the class of all *linear* data compression-reconstruction filters. The KLF model is based on the solution of a rank-constrained matrix approximation problem. Nevertheless, it may happen that the accuracy and compression ratio associated with the KLF are still not satisfactory in some circumstances.

The *second question* we address in this paper is as follows. Is there a filter that will have greater accuracy and better compression ratio than the KLF? An obvious but not constructive answer is that such a filter must be *nonlinear*. We propose a constructive determination of a nonlinear filter in the form of a sum with  $p$  terms given by (1) below, where each term is represented as a combination of three operations  $\mathcal{G}_k$ ,  $\mathcal{H}_k$  and  $\mathcal{P}_k$  for each  $k = 1, \dots, p$ . The operator  $\mathcal{P}_k$  is a non-linear operator that allows us to incorporate additional information, the operator  $\mathcal{H}_k$  is a generalized Gram-Schmidt orthogonalization that decouples the additional information and the operator  $\mathcal{G}_k$  minimizes the estimation error. The orthogonalization is used primarily to reduce the associated numerical load.

The *prime idea* is to determine  $\mathcal{G}_k$  from an iterative scheme aimed at improving the filter performance with each step. At the final step of the scheme,  $\mathcal{G}_k$  is determined subject to the prescribed rank restrictions. The scheme is described in Section II-C. Moreover, due to the orthogonal structure created by the  $\mathcal{H}_k$  it turns out that  $\mathcal{G}_k$  is determined by solution of a *separate* minimization problem for each  $k = 1, \dots, p$ . See Theorems 3 and 4 in this regard. The nonlinear operators  $\mathcal{P}_k$  imply *non-linearity* of the filter. In the case  $\mathcal{P}_k = I$  for each  $k = 1, \dots, p$ , where  $I$  is the identity mapping, the filter is linear and the method presented in Sections II-C and III-B below becomes degenerate. Examples of possible nonlinear operators  $\mathcal{P}_k$  are given in Section III-C.

The operations  $\mathcal{P}_k$  and  $\mathcal{H}_k$  are auxiliary operations related to finding  $\mathcal{G}_k$ , and they are described in Sections II-C and III-C, and Lemma 1, respectively.

3) *Third motivation: generalized rank-constrained matrix approximation:* Data compression filters are often based on the solution of a best rank-constrained matrix approximation problem. It has been shown in [6] that for such problems the solution given in [6] should be used. This solution has been obtained for the case of a minimal norm matrix. In Section III-A below, we extend the solution to a more general case and exploit it in our filter derivation.

## B. Contribution

We provide a theory for the new filter which processes infinite sets of random signals. The filter is presented in the form (1) and is based on a new iterative scheme (Section II-C) that ensures the filter accuracy is improved with each step. The final step provides signal compression and decompression (Section III-D). This step is based on the solution of a new rank-constrained matrix approximation problem (Section

<sup>3</sup>The Karhunen-Loève filter is often called the rank-reduced Wiener filter [14]. The analytic procedure is described by statisticians as Principal Component Analysis [10].

III-A) following the methodology used in [6]. A rigorous error analysis for the filter is also given (Section III-B).

## II. METHOD DESCRIPTION AND STATEMENT OF THE PROBLEM

### A. Filter structure

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  $\mathbf{x} = \{\mathbf{x}(\cdot, \alpha) \in L^2(\Omega, \mathbb{R}^m) \mid \alpha \in K\}$  where  $K$  is a measurable set [9] in some appropriate sigma field with a measure  $\lambda(\alpha)$ . Thus  $\mathbf{x} \in L^2(\Omega \times K, \mathbb{R}^m)$ . We observe an interesting duality in the representation of  $\mathbf{x}$  in that  $\mathbf{x}$  is a single signal in the space  $L^2(\Omega \times K, \mathbb{R}^m)$  and is an infinite set of signals  $\{\mathbf{x}(\cdot, \alpha) \in L^2(\Omega, \mathbb{R}^m) \mid \alpha \in K\}$  in the space  $L^2(\Omega, \mathbb{R}^m)$ . In similar fashion we write  $\mathbf{y} = \{\mathbf{y}(\cdot, \alpha) \in L^2(\Omega, \mathbb{R}^m) \mid \alpha \in K\}$  and  $\mathbf{y} \in L^2(\Omega \times K, \mathbb{R}^m)$ . We interpret  $\mathbf{x}$  as a signal that should be estimated and  $\mathbf{y}$  as an observed signal that can be used as an input to the filter  $\mathcal{F}_p$  studied below. In an intuitive way  $\mathbf{y}$  can be regarded as a noise-corrupted version of  $\mathbf{x}$ . Let  $\mathbf{x}_1 = \mathbf{y}$  and let  $\mathbf{x}_1, \dots, \mathbf{x}_p$  be a sequence of estimates of  $\mathbf{x}$  obtained by the method described in Sections II-C, III-A and III-B below. The filter  $\mathcal{F}_p$  is presented in the form

$$\mathcal{F}_p(\mathbf{y}) = \sum_{k=1}^p \mathcal{G}_k \mathcal{H}_k \mathcal{P}_k(\mathbf{x}_1, \dots, \mathbf{x}_k), \quad (1)$$

where  $\mathcal{P}_k : L^2((\Omega \times K)^k, \mathbb{R}^n) \rightarrow L^2(\Omega \times K, \mathbb{R}^n)$ ,  $\mathcal{H}_k : L^2(\Omega \times K, \mathbb{R}^n) \rightarrow L^2(\Omega \times K, \mathbb{R}^n)$  and  $\mathcal{G}_k : L^2(\Omega \times K, \mathbb{R}^n) \rightarrow L^2(\Omega \times K, \mathbb{R}^m)$ .

The purpose of the components  $\mathcal{G}_k$ ,  $\mathcal{H}_k$  and  $\mathcal{P}_k$  has been described above in Section I-A2. The components  $\mathcal{P}_k$  are defined in Sections II-C and III-C, and components the  $\mathcal{G}_k$  and  $\mathcal{H}_k$  are determined in Section III-B.

### B. Basic ideas

The input to the filter  $\mathcal{F}_p$  is a stochastic signal – a random vector  $\mathbf{y} \in L^2(\Omega \times K, \mathbb{R}^m)$ . We wish to filter and compress  $\mathbf{y}$ , and to reconstruct a compressed vector that is close to  $\mathbf{x}$ .

The first idea is to represent  $\mathbf{y}$  as a generalized sequence of signals

$$\mathbf{y} = \{\mathbf{y}(\cdot, \alpha) \in L^2(\Omega, \mathbb{R}^m) \mid \alpha \in K\}$$

where (typically)  $K \subset \mathbb{R}$  and to use the special norm (3) in the statement of the problems in Sections II-C and II-D. The filter processes each  $\mathbf{y}(\cdot, \alpha)$  separately but the filter components are determined from the entire sets  $\mathbf{x}$  and  $\mathbf{y}$ .

The filter is best illustrated for a particular case when the sets  $\mathbf{x}$  and  $\mathbf{y}$  are finite as in Example 4 of Section VI where we have, in (1),  $p = 1$ ,  $\mathcal{H}_1$  and  $\mathcal{P}_1$  are each the identity mapping, and  $\mathcal{G}_1$  is determined from problem (4). In such a case, we set  $\mathbf{y} = \{\mathbf{y}(\cdot, t_k) \in L^2(\Omega, \mathbb{R}^m) \mid t_k \in \mathbb{R} \forall k = 1, \dots, N\}$  and  $\mathbf{x} = \{\mathbf{x}(\cdot, t_k) \in L^2(\Omega, \mathbb{R}^m) \mid t_k \in \mathbb{R} \forall k = 1, \dots, N\}$ . The estimate  $\tilde{\mathbf{x}}$  of  $\mathbf{x}$  by the proposed filter is given by

$$\tilde{\mathbf{x}} = [\tilde{\mathbf{x}}(\cdot, t_1), \dots, \tilde{\mathbf{x}}(\cdot, t_N)] = \tilde{\mathcal{G}}_1[\mathbf{y}(\cdot, t_1), \dots, \mathbf{y}(\cdot, t_N)]$$

where  $\tilde{\mathcal{G}}_1$  is provided by Theorem 3 of Section III-B and, for each  $k = 1, \dots, N$ , the estimate  $\tilde{\mathbf{x}}(\cdot, t_k) = \tilde{\mathcal{G}}_1 \mathbf{y}(\cdot, t_k)$  is determined separately.

The second idea is based on the following observation. *It is natural to expect that the filter performance would be better if the input to  $\mathcal{F}_p$  was  $\mathbf{x}$ , and not  $\mathbf{y}$ .* Indeed, in this case, no transformation need be done and in particular, no noise filtering is necessary. The filter would provide only compression and subsequent decompression. Therefore, our next idea is to construct an iterative scheme such that, at each step, the scheme allows us:

- (a) to find a new input for  $\mathcal{F}_p$  which is closer to  $\mathbf{x}$  than those determined in preceding steps,
- (b) to determine operators  $\mathcal{G}_1, \dots, \mathcal{G}_k$  that minimize the associated error at each step, and
- (c) to define operators  $\mathcal{H}_1, \dots, \mathcal{H}_k$  so that the desired operators  $\mathcal{G}_1, \dots, \mathcal{G}_k$  are determined from a numerically simple scheme (see Section III-E1 for more detail).

This idea is realized below in Sections II-C and III-B in such a way that  $\mathcal{F}_p$  is determined from a sequence of associated error minimization problems.

### C. Basic device

Let the model for the filter  $\mathcal{F}_p$  be given by (1). The proposed device for determining operators  $\mathcal{H}_k$ ,  $\mathcal{G}_k$  and  $\mathcal{P}_k$  in (1) is as follows. The operators  $\mathcal{H}_k$  in (1) are defined by Lemma 1 of Section III-B below. The recommended procedure for determining the operators  $\mathcal{G}_k$  and  $\mathcal{P}_k$  consists of  $p$  successive steps based on the following idea. If  $\mathbf{x}_1 = \mathbf{y}$  and the subsequent estimates of  $\mathbf{x}$  denoted by  $\mathbf{x}_2, \dots, \mathbf{x}_{k-1}$  have been determined from successive steps then for the  $k$ th step, we define

$$\mathbf{x}_k = \sum_{j=1}^{k-1} \tilde{\mathcal{G}}_j \mathcal{H}_j(\mathbf{y}_1, \dots, \mathbf{y}_j) \quad (2)$$

where  $\mathbf{y}_j = \mathcal{P}_j(\mathbf{x}_1, \dots, \mathbf{x}_j)$  and  $\mathcal{P}_j$  is conceptually defined in (1) and where  $\tilde{\mathcal{G}}_1, \dots, \tilde{\mathcal{G}}_k$  are determined from the proposed error minimization criteria. We propose the following scheme.

First, we introduce the norm  $\|\cdot\|_{K,\Omega}$  defined by

$$\|\mathbf{x}\|_{K,\Omega}^2 = \frac{1}{\lambda(K)} \iint_{\Omega \times K} \|\mathbf{x}(\omega, \alpha)\|_2^2 d(\mu(\omega), \lambda(\alpha)), \quad (3)$$

where  $\lambda(K) = \int_K d\lambda(\alpha)$  is the measure of  $K$  and where  $\|\mathbf{x}(\omega, \alpha)\|_2$  is the Euclidean norm of  $\mathbf{x}(\omega, \alpha) \in \mathbb{R}^m$ . More explanation is given in Section VI.

The proposed scheme consists of a sequence of steps.

**Step 1 The initial step.** For  $k = 1$ , let  $\mathbf{x}_1 := \mathbf{y}$  and let  $\mathcal{P}_1 = I$  and  $\mathcal{H}_1 = I$  so that  $\mathbf{y}_1 := \mathbf{x}_1$  and find  $\tilde{\mathcal{G}}_1$  that satisfies

$$\|\mathbf{x} - \tilde{\mathcal{G}}_1(\mathbf{y}_1)\|_{K,\Omega}^2 = \min_{\mathcal{G}_1} \|\mathbf{x} - \mathcal{G}_1(\mathbf{y}_1)\|_{K,\Omega}^2. \quad (4)$$

Set  $\mathbf{x}_2 = \tilde{\mathcal{G}}_1(\mathbf{x}_1)$ .

**Step  $\ell - 1$  for  $\ell = 3, \dots, p$ . The filtering steps.** Let  $\mathbf{x}_2, \dots, \mathbf{x}_{\ell-1}$  be known from the preceding  $\ell - 2$  steps. For  $k = \ell - 1$  in (2) define  $\mathcal{P}_{\ell-1}$  so that  $\mathbf{y}_{\ell-1} = \mathcal{P}_{\ell-1}(\mathbf{x}_1, \dots, \mathbf{x}_{\ell-1})$

and find  $\tilde{\mathcal{G}}_1, \tilde{\mathcal{G}}_2, \dots, \tilde{\mathcal{G}}_{\ell-1}$  that satisfy

$$\begin{aligned} & \left\| \mathbf{x} - \sum_{j=1}^{\ell-1} \tilde{\mathcal{G}}_j \mathcal{H}_j(\mathbf{y}_1, \dots, \mathbf{y}_j) \right\|_{K,\Omega}^2 \\ &= \min_{\mathcal{G}_1, \dots, \mathcal{G}_{\ell-1}} \left\| \mathbf{x} - \sum_{j=1}^{\ell-1} \mathcal{G}_j \mathcal{H}_j(\mathbf{y}_1, \dots, \mathbf{y}_j) \right\|_{K,\Omega}^2. \end{aligned} \quad (5)$$

Set  $\mathbf{x}_\ell = \sum_{j=1}^{\ell-1} \tilde{\mathcal{G}}_j \mathcal{H}_j(\mathbf{y}_1, \dots, \mathbf{y}_j)$ .

**Step  $p$ . The reconstruction step.** At the final step, for  $k = p$  in (2) define  $\mathcal{P}_p$  so that  $\mathbf{y}_p = \mathcal{P}_p(\mathbf{x}_1, \dots, \mathbf{x}_p)$  and find  $\tilde{\mathcal{G}}_1, \tilde{\mathcal{G}}_2, \dots, \tilde{\mathcal{G}}_p$  that satisfy

$$\begin{aligned} & \left\| \mathbf{x} - \sum_{j=1}^p \tilde{\mathcal{G}}_j \mathcal{H}_j(\mathbf{y}_1, \dots, \mathbf{y}_j) \right\|_{K,\Omega}^2 \\ &= \min_{\mathcal{G}_1, \dots, \mathcal{G}_p} \left\| \mathbf{x} - \sum_{j=1}^p \mathcal{G}_j \mathcal{H}_j(\mathbf{y}_1, \dots, \mathbf{y}_j) \right\|_{K,\Omega}^2. \end{aligned} \quad (6)$$

subject to

$$\sum_{j=1}^p \text{rank } \tilde{\mathcal{G}}_j \leq r \leq \min\{m, n\} \quad \text{and} \quad (7)$$

$$\text{rank } \tilde{\mathcal{G}}_j \leq r_j \quad \text{with} \quad \sum_{j=1}^p r_j = r. \quad (8)$$

We set  $\mathbf{x}_{p+1} = \sum_{i=1}^p \tilde{\mathcal{G}}_i \mathcal{H}_i(\mathbf{y}_1, \dots, \mathbf{y}_i)$ .

Due to condition (7)–(8), this step provides data compression–reconstruction as it is shown in Section III-D below.

### D. Statement of the problem

The problem we solve follows from the basic structure presented above. Let  $\mathbf{x}$  and  $\mathbf{y}$  be random signals where  $\mathbf{x}$  is an unobservable signal and  $\mathbf{y}$  is an observable input signal. Both  $\mathbf{x}$  and  $\mathbf{y}$  are analytically unknown and the only available information is given by covariance matrices formed from  $\mathbf{x}$  and  $\mathbf{y}$ . We wish to find a filter of the form (1) with  $\mathcal{G}_1, \dots, \mathcal{G}_p$  determined from the error minimization problems (4)–(7). The filter must satisfy the requirements (a), (b) and (c) of the Section II-B.

**The main differences** of the problem given by (1), (4)–(7) from known problems are as follows. First, we are looking for the optimal filter that minimizes the associated error for any signal from an infinite signal set. This is achieved due to the new norm given by (3). The second major difference is the special structure of the filter (1)–(7). Each step of the suggested procedure represents the solution of a best approximation problem that minimizes an associated error. If the final stage **Step  $p$**  is omitted the scheme performs filtering only. If **Step  $p$**  is included then the scheme provides simultaneous signal filtering and compression and the subsequent reconstruction.

Solutions to problems (4)–(7) are given below under assumptions that certain covariance matrices are known. The assumptions are specified in Theorems 3 and 4 below. It is important to note that for new signals constructed during the phase of iterative improvement the required covariance matrices can be constructed from the original data.

E. Advantages of the filter

The advantages of the proposed filter are discussed in Section III-E.

III. MAIN RESULTS

The solution to the problem (6)–(7) is based on the results presented in Section III-A. The solution itself is given in Theorem 4 below.

A. Generic low-rank matrix approximation problem

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

$$A = U_A \Sigma_A V_A^*, \quad (9)$$

where  $U_A \in \mathbb{C}^{m \times m}$  and  $V_A \in \mathbb{C}^{n \times n}$  are unitary matrices and

$$\Sigma_A := \text{diag}(\sigma_1(A), \dots, \sigma_{\min(m,n)}(A)) \in \mathbb{C}^{m \times n}$$

is a generalized diagonal matrix with singular values  $\sigma_1(A) \geq \sigma_2(A) \geq \dots \geq 0$  on the main diagonal.

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

$$P_{A,L} := \sum_{i=1}^{\text{rank } A} u_i u_i^* \in \mathbb{C}^{m \times m} \text{ and } P_{A,R} := \sum_{i=1}^{\text{rank } A} v_i v_i^* \in \mathbb{C}^{n \times n} \quad (10)$$

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

$$A_k := (A)_k := \sum_{i=1}^k \sigma_i(A) u_i v_i^* = U_{A_k} \Sigma_{A_k} V_{A_k}^* \in \mathbb{C}^{m \times n} \quad (11)$$

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

$$U_{A_k} = [u_1 \ u_2 \ \dots \ u_k], \quad \Sigma_{A_k} = \text{diag}(\sigma_1(A), \dots, \sigma_k(A)) \quad (12)$$

$$\text{and } V_{A_k} = [v_1 \ v_2 \ \dots \ v_k]. \quad (13)$$

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

Henceforth  $\|\cdot\|_F$  denotes the Frobenius norm. We write  $A^\dagger$  for the Moore-Penrose pseudo-inverse of the matrix  $A$  [1]. Below, we provide generalizations of the classical minimization problem due to Eckart and Young [3]. First, we present the result obtained in [?].

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

$$X = B^\dagger (P_{B,L} A P_{C,R})_k C^\dagger \quad (14)$$

is a solution to the minimal problem

$$\min_{X \in \mathcal{R}(p,q,k)} \|A - BXC\|_F, \quad (15)$$

having the minimal  $\|X\|_F$ . This solution is unique if and only if either

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

or

$$1 \leq k < \text{rank} (P_{B,L} A P_{C,R}) \text{ and}$$

$$\sigma_k(P_{B,L} A P_{C,R}) > \sigma_{k+1}(P_{B,L} A P_{C,R}).$$

Let us denote  $B_1 := \text{diag}(\sigma_1(B), \dots, \sigma_s(B)) \in \mathbb{C}^{s \times s}$ ,  $C_1 := \text{diag}(\sigma_1(C), \dots, \sigma_t(C)) \in \mathbb{C}^{t \times t}$ ,  $s = \text{rank } B$ ,  $t = \text{rank } C$ ,  $\hat{A} = U_B^* A V_C$ ,  $\hat{A} = [A_{ij}]_{i,j=1}^{s,t}$  and  $A_{11} \in \mathbb{C}^{s \times t}$ .

It follows from the proof of Theorem 1 provided in [?] that a solution to the minimal problem (15) without the constraint of the minimal  $\|X\|_F$  is given by

$$X = V_B \hat{X} U_C^* \text{ with } \hat{X} = \begin{bmatrix} B_1^{-1} (A_{11})_k C_1^{-1} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \quad (16)$$

where  $X_{12}$ ,  $X_{21}$  and  $X_{22}$  should be chosen in such a way that  $\hat{X} \in \mathcal{R}(p, q, k)$ . In Theorem 2 below, we show how to choose  $X_{12}$ ,  $X_{21}$  and  $X_{22}$  (see (20)) to satisfy the condition  $\hat{X} \in \mathcal{R}(p, q, k)$ .

**Theorem 2:** Let  $P \in \mathbb{C}^{(p-s) \times s}$ ,  $Q \in \mathbb{C}^{t \times (q-t)}$  be arbitrary matrices,  $V_B = [V_1 \ V_2] \in \mathbb{C}^{p \times p}$ ,  $U_C = [U_1 \ U_2] \in \mathbb{C}^{q \times q}$  where  $V_1 \in \mathbb{C}^{p \times s}$  and  $U_1 \in \mathbb{C}^{q \times t}$ , and

$$K = [B^\dagger B, \ I - B^\dagger B] \begin{bmatrix} \mathbb{O} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \begin{bmatrix} CC^\dagger \\ I - CC^\dagger \end{bmatrix}, \quad (17)$$

where

$$K_{12} = V_1 X_{11} Q U_2^*, \quad K_{21} = V_2 P X_{11} U_1^*, \\ K_{22} = V_2 P X_{11} Q U_2^* \text{ and } X_{11} = B_1^{-1} (A_{11})_k C_1^{-1}. \quad (18)$$

If the constraint of the minimal  $\|X\|_F$  is omitted in the problem (15) then the solution is not unique. A family of solutions to the problem (15) without this constraint is given by

$$X = B^\dagger (P_{B,L} A P_{C,R})_k C^\dagger + K. \quad (19)$$

**Proof.** To preserve  $\text{rank } \hat{X} \leq k$ , we should choose  $X_{12}$ ,  $X_{21}$  and  $X_{22}$  in (16) a compatible form. In particular, the columns of  $X_{12}$  and the rows of  $X_{21}$  must be linear combinations of the columns of  $X_{11}$  and the rows of  $X_{11}$ , respectively. To this end, we need to show that there exist matrices  $P \in \mathbb{C}^{(p-s) \times s}$  and  $Q \in \mathbb{C}^{t \times (q-t)}$  such that, for  $(i, j) \neq (1, 1)$ ,  $X_{ij}$  can be written in the form

$$X_{12} = X_{11} Q, \quad X_{21} = P X_{11} \text{ and } X_{22} = P X_{11} Q. \quad (20)$$

The existence follows from the next observation. By Gaussian elimination, there are matrices  $P \in \mathbb{C}^{(p-s) \times s}$  and  $Q \in \mathbb{C}^{t \times (q-t)}$  such that  $\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \begin{bmatrix} I & -Q \\ \mathbb{O} & I \end{bmatrix} = \begin{bmatrix} X_{11} & \mathbb{O} \\ X_{21} & X_{22} - X_{21} Q \end{bmatrix}$  and  $\begin{bmatrix} I & \mathbb{O} \\ -P & I \end{bmatrix} \begin{bmatrix} X_{11} & \mathbb{O} \\ X_{21} & X_{22} - X_{21} Q \end{bmatrix} = \begin{bmatrix} X_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$  which is true if

$$X_{12} - X_{11} Q = \mathbb{O}, \quad X_{21} - P X_{11} = \mathbb{O} \text{ and } X_{22} - X_{21} Q = \mathbb{O}. \quad (21)$$

The last condition in (21) follows from the observation that the matrices  $\begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$  and  $\begin{bmatrix} X_{11} & \mathbb{O} \\ \mathbb{O} & X_{22} - X_{21} Q \end{bmatrix}$  have

the same rank. Therefore, to preserve rank  $\hat{X} \leq k$ , we should choose  $X_{22} - X_{21}Q = \mathbb{O}$ . Thus, (20) follows.

Next, let us write  $\bar{X} = \begin{bmatrix} B_1^{-1}(A_{11})_k C_1^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$ . To show that

$$\bar{X} = \Sigma_B^\dagger (P_{\Sigma_B, L} \tilde{A} P_{\Sigma_C, R})_k \Sigma_C^\dagger \quad (22)$$

we write

$$\begin{aligned} \bar{X} &= \Sigma_B^\dagger \begin{bmatrix} (A_{11})_k & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \Sigma_C^\dagger = \Sigma_B^\dagger \begin{bmatrix} A_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}_k \Sigma_C^\dagger \\ &= \Sigma_B^\dagger \left( \begin{bmatrix} I_{ss} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} I_{tt} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} \right)_k \Sigma_C^\dagger \end{aligned}$$

where  $I_{ss}$  is the  $s \times s$  identity matrix. The SVD for  $\Sigma_B = \begin{bmatrix} B_1 & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$  is given by  $\Sigma_B = U_{\Sigma_B} \begin{bmatrix} B_1 & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} V_{\Sigma_B}^*$ . Here  $U_{\Sigma_B} = [e_1, \dots, e_m]$  and  $V_{\Sigma_B} = [e_1, \dots, e_p]$  where  $e_j = [0, \dots, 0, 1, 0, \dots, 0]^T$  with 1 on the  $j$ th position. Therefore,  $P_{\Sigma_B, L} = \sum_{i=1}^s e_i e_i^T = \begin{bmatrix} I_{ss} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}$  and by analogy,

$$P_{\Sigma_C, R} = \sum_{i=1}^t e_i e_i^T = \begin{bmatrix} I_{tt} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix}. \text{ Thus, } \begin{bmatrix} A_{11} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} =$$

$P_{\Sigma_B, L} \tilde{A} P_{\Sigma_C, R}$ . Therefore, (22) follows.

Next,  $\hat{X} = \Sigma_B^\dagger (P_{\Sigma_B, L} \tilde{A} P_{\Sigma_C, R})_k \Sigma_C^\dagger + \begin{bmatrix} \mathbb{O} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}$ . As a result,

$$\begin{aligned} X &= V_B \hat{X} U_C^* = B^\dagger (P_{B, L} A P_{C, R})_k C^\dagger \\ &\quad + [V_1 \ V_2] \begin{bmatrix} \mathbb{O} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \begin{bmatrix} U_1^* \\ U_2^* \end{bmatrix} \\ &= B^\dagger (P_{B, L} A P_{C, R})_k C^\dagger + V_2 X_{21} U_1^* + V_1 X_{12} U_2^* \\ &\quad + V_2 X_{22} U_2^*. \end{aligned}$$

We note that  $V_1^* V_1 = I_{ss}$ ,  $V_2^* V_2 = I_{p-s, p-s}$ ,  $U_1^* U_1 = I_{tt}$ ,  $U_2^* U_2 = I_{q-t, q-t}$ ,  $V_1 V_1^* + V_2 V_2^* = I_{pp}$ ,  $U_1 U_1^* + U_2 U_2^* = I_{qq}$ , and  $B^\dagger B = V_1 V_1^*$  and  $CC^\dagger = U_1 U_1^*$ .

Then

$$\begin{aligned} X &= B^\dagger (P_{B, L} A P_{C, R})_k C^\dagger + (I - V_1 V_1^*) K_{21} U_1 U_1^* \\ &\quad + V_1 V_1^* K_{12} (I - U_1 U_1^*) + (I - V_1 V_1^*) K_{22} (I - U_1 U_1^*) \\ &= B^\dagger (P_{B, L} A P_{C, R})_k C^\dagger + (I - B^\dagger B) K_{21} C C^\dagger \\ &\quad + B^\dagger B K_{12} (I - C C^\dagger) + (I - B^\dagger B) K_{22} (I - C C^\dagger) \\ &= B^\dagger (P_{B, L} A P_{C, R})_k C^\dagger + K. \end{aligned} \quad (23)$$

Thus, (19) is true. ■

The following Corollary will be used in the next Section.

**Corollary 1:** Let  $p = m$ ,  $q = n$ ,  $B = I$  and  $\tilde{A} = A V_C = [\tilde{A}_1 \ \tilde{A}_2]$  where  $\tilde{A}_1 \in \mathbb{C}^{m \times t}$ . Let  $Q \in \mathbb{C}^{t \times (n-t)}$  be arbitrary. Then the solution to the problem

$$\min_{X \in \mathcal{R}(m, n, k)} \|A - XC\| \quad (24)$$

is given by

$$X = (A P_{C, R})_k V_C^* C^\dagger + (\tilde{A}_1)_k C_1^{-1} Q U_2^* (I - C C^\dagger). \quad (25)$$

**Proof.** Indeed, a solution to the problem

$$\min_{\tilde{X} \in \mathcal{R}(m, n, k)} \|\tilde{A} - \tilde{X} \Sigma_C\| \quad (26)$$

where  $\tilde{X} = [X_1 \ X_2]$ , is  $\hat{X} = [\hat{X}_1 \ \hat{X}_2]$  with  $\hat{X}_1 = (\tilde{A}_1)_k C_1^{-1}$  and  $\hat{X}_2 = \hat{X}_1 Q$ . Then  $X = \hat{X} U_C^*$  is the solution to the problem (24). We have  $\hat{X} = [\hat{X}_1 \ \mathbb{O}] + [\mathbb{O} \ \hat{X}_2]$  where

$$\hat{X}_1 = [(\tilde{A}_1)_k \ \mathbb{O}] \begin{bmatrix} C_1^{-1} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} = [(\tilde{A}_1)_k \ \mathbb{O}] \Sigma_C^\dagger = ([\tilde{A}_1 \ \mathbb{O}])_k \Sigma_C^\dagger$$

with  $[\tilde{A}_1 \ \mathbb{O}] = [\tilde{A}_1 \ \tilde{A}_2] \begin{bmatrix} I_{tt} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} \end{bmatrix} = \tilde{A} P_{\Sigma_C, R}$ . Therefore,

$$X = (\tilde{A} P_{\Sigma_C, R})_k \Sigma_C^\dagger U_C^* + [\mathbb{O}, \hat{X}_1 Q] U_C^*.$$

Here,

$$(\tilde{A} P_{\Sigma_C, R})_k \Sigma_C^\dagger U_C^* = (\tilde{A} P_{\Sigma_C, R})_k V_C^* C^\dagger \quad \text{and}$$

$$[\mathbb{O}, \hat{X}_1 Q] U_C^* = \hat{X}_1 Q U_2^*.$$

Next, let  $K_2 = X_2 U_2^*$ . Then  $X_2 U_2^* = K_2 U_2 U_2^* = K_2 (I - U_1 U_1^*) = K_2 (I - C C^\dagger)$ . Thus,  $X = (\tilde{A} P_{\Sigma_C, R})_k \Sigma_C^\dagger U_C^* + \hat{X}_1 Q U_2^* (I - C C^\dagger)$  and (25) follows. ■

**Remark 1:** The Eckart-Young theorem [3] follows from Theorem 2 as a particular case when  $p = m$ ,  $q = n$ ,  $B = I_m$  and  $C = I_n$ .

### B. Solution of the problems (4)–(5) and (6)–(7)

Let  $\mathbf{x} = \{\mathbf{x}(\cdot, \alpha) \in L^2(\Omega, \mathbb{R}^m) \mid \alpha \in K\} \in L^2(\Omega \times K, \mathbb{R}^m)$  and  $\mathbf{y} = \{\mathbf{y}(\cdot, \alpha) \in L^2(\Omega, \mathbb{R}^m) \mid \alpha \in K\} \in L^2(\Omega \times K, \mathbb{R}^m)$  where  $K$  is a measurable set with finite measure  $\lambda(K)$  in some sigma field of measurable sets [9].

We write  $\mathbf{x} = (x^{(1)}, \dots, x^{(m)})^T$  and  $\mathbf{y} = (y^{(1)}, \dots, y^{(n)})^T$  where  $x^{(i)} \in L^2(\Omega \times K, \mathbb{R})$  for  $i = 1, \dots, m$  and  $y^{(j)} \in L^2(\Omega \times K, \mathbb{R})$  for  $j = 1, \dots, n$  are real valued random variables. Let

$$\langle x^{(i)}, y^{(j)} \rangle = \frac{1}{\lambda(K)} \iint_{\Omega \times K} x^{(i)}(\omega, \alpha) y^{(j)}(\omega, \alpha) d(\mu(\omega), \lambda(\alpha)) \quad (27)$$

and

$$\langle y^{(i)}, y^{(j)} \rangle = \frac{1}{\lambda(K)} \iint_{\Omega \times K} y^{(i)}(\omega, \alpha) y^{(j)}(\omega, \alpha) d(\mu(\omega), \lambda(\alpha)) \quad (28)$$

and define covariance matrices

$$R[\mathbf{xy}^T] = [\langle x^{(i)}, y^{(j)} \rangle] \in \mathbb{R}^{m \times n} \quad (29)$$

and

$$R[\mathbf{yy}^T] = [\langle y^{(i)}, y^{(j)} \rangle] \in \mathbb{R}^{n \times n}. \quad (30)$$

If  $M \in \mathbb{R}^{m \times n}$  then we define a corresponding operator  $\mathcal{M} : L^2(\Omega \times K, \mathbb{R}^n) \rightarrow L^2(\Omega \times K, \mathbb{R}^m)$  by the formula

$$[\mathcal{M}(\mathbf{y})](\omega, \alpha) = M[\mathbf{y}(\omega, \alpha)]. \quad (31)$$

Henceforth all such operators will be denoted by a calligraphic letter.

We wish

- (i) to give explicit solutions to the problems (4)–(5) and (6)–(7), and
- (ii) to show that the error associated with the proposed method decreases as the number  $k$  of steps (4)–(7) increases.

First, we give a method for the determination of the operators  $\mathcal{H}_1, \dots, \mathcal{H}_k$  in the filter model (1).

**Definition 1:** The random vectors  $\{\mathbf{v}_i\}_{i=1, \dots, k} \subset L^2(\Omega \times K, \mathbb{R}^n)$  are called pairwise orthogonal if

$$R[\mathbf{v}_i \mathbf{v}_j^T] = \mathbf{0} \in \mathbb{R}^{n \times n} \quad (32)$$

for  $i \neq j$  with  $i, j = 1, \dots, k$ .

**Lemma 1:** Let  $\mathbf{y}_i = \mathcal{P}_i(\mathbf{x}_1, \dots, \mathbf{x}_i)$  and  $\mathbf{v}_i = \mathcal{H}_i(\mathbf{y}_1, \dots, \mathbf{y}_i)$  for each  $i = 1, 2, \dots, k$  where the linear operators  $\mathcal{H}_i : L^2((\Omega \times K)^i, \mathbb{R}^n) \mapsto L^2(\Omega \times K, \mathbb{R}^n)$  are defined inductively by the generalized Gram-Schmidt algorithm

$$\mathbf{v}_1 = \mathbf{y}_1 \quad \text{and} \quad \mathbf{v}_i = \mathbf{y}_i - \sum_{\ell=1}^{i-1} \mathcal{L}_{i\ell}(\mathbf{v}_\ell) \quad (33)$$

with  $\mathcal{L}_{i\ell} : L^2(\Omega \times K, \mathbb{R}^n) \rightarrow L^2(\Omega \times K, \mathbb{R}^n)$  defined by  $\mathcal{L}_{i\ell}(\mathbf{v}_\ell) = L_{i\ell} \mathbf{v}_\ell$  where

$$L_{i\ell} = R[\mathbf{y}_i \mathbf{v}_\ell^T] R[\mathbf{v}_\ell \mathbf{v}_\ell^T]^\dagger + M_{i\ell} (I - R[\mathbf{v}_\ell \mathbf{v}_\ell^T] R[\mathbf{v}_\ell \mathbf{v}_\ell^T]^\dagger) \quad (34)$$

and  $M_{i\ell} \in \mathbb{R}^{n \times n}$  is arbitrary for each  $i = 2, \dots, k$ . Then the vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  are pairwise orthogonal in  $L^2(\Omega \times K, \mathbb{R}^n)$ .

**Proof.** The proof follows directly from (33), (34) and Definition 1 on the basis of the relation  $R[\mathbf{y}_i \mathbf{v}_\ell^T] = R[\mathbf{y}_i \mathbf{v}_\ell^T] R[\mathbf{v}_\ell \mathbf{v}_\ell^T]^\dagger R[\mathbf{v}_\ell \mathbf{v}_\ell^T]$ . We refer the reader to Torokhti and Howlett [18], p. 168 for additional information. ■

**Remark 2:** In practice it is usual to take  $M_{i\ell} = \mathbf{0}$ .

**Remark 3:** The Gram-Schmidt formula can be rearranged to give

$$\mathbf{y}_1 = \mathbf{v}_1 \quad \text{and} \quad \mathbf{y}_i = \mathbf{v}_i + \sum_{\ell=1}^{i-1} L_{i\ell} \mathbf{v}_\ell$$

for each  $i = 2, \dots, k$ . Thus we can write

$$\mathbf{y}^{(i)} = (\mathbf{I}_{im} + \mathbf{L}_i) \mathbf{v}^{(i)}$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_i \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{L}_{31} & \mathbf{L}_{32} & \mathbf{I}_m & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{L}_{i1} & \mathbf{L}_{i2} & \mathbf{L}_{i3} & \cdots & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \mathbf{v}_3 \\ \vdots \\ \mathbf{v}_i \end{bmatrix}$$

from which it follows easily that

$$\begin{aligned} \mathbf{v}^{(i)} &= (\mathbf{I}_{im} + \mathbf{L}_i)^{-1} \mathbf{y}^{(i)} \\ &= \begin{bmatrix} \mathbf{I}_m & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{H}_{21} & \mathbf{I}_m & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{H}_{31} & \mathbf{H}_{32} & \mathbf{I}_m & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{i1} & \mathbf{H}_{i2} & \mathbf{H}_{i3} & \cdots & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \vdots \\ \mathbf{y}_i \end{bmatrix} \end{aligned}$$

is well defined. This formula shows clearly that  $\mathbf{v}_i = \mathcal{H}_i(\mathbf{y}_1, \dots, \mathbf{y}_i) = \mathbf{H}_{i1} \mathbf{y}_1 + \dots + \mathbf{H}_{i(i-1)} \mathbf{y}_{i-1} + \mathbf{y}_i$ .

The matrices  $R[\mathbf{x} \mathbf{v}_j^T]$ ,  $R[\mathbf{y}_i \mathbf{v}_j^T]$  and  $R[\mathbf{v}_j \mathbf{v}_j^T]$  are assumed known. Estimation of these matrices is discussed in Section 5.3 of the book by Torokhti and Howlett [18].

**Theorem 3:** For  $k = 1, \dots, p - 1$  the solution  $\tilde{G}_j$  to problems (4)–(5) is

$$\tilde{G}_j = R[\mathbf{x} \mathbf{v}_j^T] R^\dagger[\mathbf{v}_j \mathbf{v}_j^T] + M_j (I - R[\mathbf{v}_j \mathbf{v}_j^T] R^\dagger[\mathbf{v}_j \mathbf{v}_j^T]) \quad (35)$$

where  $M_j$  is arbitrary for each  $j = 1, \dots, k - 1$  and  $\mathbf{v}_j = \mathcal{H}_j(\mathbf{y}_1, \dots, \mathbf{y}_j)$  has been determined for each  $j = 1, 2, \dots, k$  by Lemma 1. The associated error is

$$\begin{aligned} &\|\mathbf{x} - \mathbf{x}_k\|_{K, \Omega}^2 \\ &= \text{tr} \left\{ R[\mathbf{x} \mathbf{x}^T] - \sum_{j=1}^{k-1} R[\mathbf{x} \mathbf{v}_j^T] R[\mathbf{v}_j \mathbf{v}_j^T]^\dagger R[\mathbf{v}_j \mathbf{x}^T] \right\} \quad (36) \end{aligned}$$

**Proof** We write  $\delta_k = \|\mathbf{x} - \mathbf{x}_k\|_{K, \Omega}^2$ . We have  $\|\mathbf{x}\|_{K, \Omega}^2 = \text{tr}\{R[\mathbf{x} \mathbf{x}^T]\}$ . Therefore, for  $k = 1, \dots, p$ ,

$$\begin{aligned} \delta_k &= \text{tr} \left\{ R \left[ \left( \mathbf{x} - \sum_{j=1}^{k-1} \mathcal{G}_j(\mathbf{v}_j) \right) \left( \mathbf{x} - \sum_{j=1}^{k-1} \mathcal{G}_j(\mathbf{v}_j) \right)^T \right] \right\} \\ &= \text{tr} \left\{ R[\mathbf{x} \mathbf{x}^T] - \sum_{j=1}^{k-1} (R[\mathbf{x} \mathbf{v}_j^T] G_j^T + G_j R[\mathbf{v}_j \mathbf{x}^T]) \right. \\ &\quad \left. + R \left[ \left( \sum_{j=1}^{k-1} \mathcal{G}_j(\mathbf{v}_j) \right) \left( \sum_{j=1}^{k-1} \mathcal{G}_j(\mathbf{v}_j) \right)^T \right] \right\}. \quad (37) \end{aligned}$$

In (37),

$$R \left[ \left( \sum_{j=1}^{k-1} \mathcal{G}_j(\mathbf{v}_j) \right) \left( \sum_{j=1}^{k-1} \mathcal{G}_j(\mathbf{v}_j) \right)^T \right] = \sum_{j=1}^k G_j R[\mathbf{v}_j \mathbf{v}_j^T] G_j^T$$

owing to the orthogonality of vectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ . Thus,

$$\begin{aligned} \delta_k &= \|R^{1/2}[\mathbf{x} \mathbf{x}^T]\|_F^2 - \sum_{j=1}^{k-1} \|R[\mathbf{x} \mathbf{v}_j^T] (R[\mathbf{v}_j \mathbf{v}_j^T]^{1/2})^\dagger\|_F^2 \\ &\quad + \sum_{j=1}^{k-1} J_j(G_j) \quad (39) \end{aligned}$$

where

$$J_j(G_j) = \|A_j - G_j C_j\|_F^2. \quad (40)$$

Because the terms  $J_1(G_1), \dots, J_k(G_k)$  in the sum  $\sum_{j=1}^k J_j(G_j)$  are determined independently we have

$$\min_{G_1, \dots, G_k} \sum_{j=1}^k J_j(G_j) = \sum_{j=1}^k \min_{G_j} J_j(G_j). \quad (41)$$

Therefore, for  $k = 1, \dots, p - 1$ , the minimum in (4)–(5) is achieved if

$$\begin{aligned} &A_j - G_j C_j = \mathbf{0} \\ \Leftrightarrow &(G_j R[\mathbf{v}_j \mathbf{v}_j^T] - R[\mathbf{x} \mathbf{v}_j^T]) (R[\mathbf{v}_j \mathbf{v}_j^T]^{1/2})^\dagger = \mathbf{0} \end{aligned}$$

and if we multiply on the right by  $(R[\mathbf{v}_j \mathbf{v}_j^T]^{1/2})^\dagger R[\mathbf{v}_j \mathbf{v}_j^T]$  we can see that

$$G_j R[\mathbf{v}_j \mathbf{v}_j^T] - R[\mathbf{x} \mathbf{v}_j^T] = \mathbf{0}.$$

According to [1] a general solution to this equation is given by (35) if and only if  $R[\mathbf{x} \mathbf{v}_j^T] = R[\mathbf{x} \mathbf{v}_j^T] R[\mathbf{v}_j \mathbf{v}_j^T]^\dagger R[\mathbf{v}_j \mathbf{v}_j^T]$ . This is clearly true. Therefore the optimal value  $\tilde{G}_j$  is given by

$$\tilde{G}_j = R[\mathbf{x} \mathbf{v}_j^T] R[\mathbf{v}_j \mathbf{v}_j^T]^\dagger + M_j (I - R[\mathbf{x} \mathbf{v}_j^T] R[\mathbf{v}_j \mathbf{v}_j^T]^\dagger)$$

where  $M_j$  is arbitrary for all  $j = 1, 2, \dots, k-1$ . ■

*Remark 4:* We remind the reader that there is no rank restriction on the matrix operators during the filtration procedure. Thus Theorem 3 solves a matrix optimization problem with no rank restriction.

Define

$$A_j = R[\mathbf{x}\mathbf{v}_j^T]R[\mathbf{v}_j\mathbf{v}_j^T]^{1/2\dagger} \quad \text{and} \quad C_j = R[\mathbf{v}_j\mathbf{v}_j^T]^{1/2}.$$

and let

$$A_j = U_{A_j}\Sigma_{A_j}V_{A_j}^T \quad \text{and} \quad C_j = U_{C_j}\Sigma_{C_j}V_{C_j}^T \quad (42)$$

be the singular value decompositions for  $A_j$  and  $C_j$  respectively. We use a similar notation to that used in (9). To derive a solution to the problem (6)–(7) define

$$U_{C_j} = [U_{C_j1} \ U_{C_j2}] \quad \text{and} \quad V_{C_j} = [V_{C_j1} \ V_{C_j2}]$$

and let  $\tilde{A}_j = A_jV_{C_j}$  and  $D_j = \text{diag}(\sigma_1(C_j), \dots, \sigma_{t_j}(C_j)) \in \mathbb{C}^{t_j \times t_j}$  with  $t_j = \text{rank } C_j$  and

$$K_j = (\tilde{A}_{j1})_{r_j} D_j^{-1} Q_j U_{j2}^T (I - C_j C_j^\dagger)$$

where  $Q_j$  is arbitrary. The singular values of matrix  $A_j$  are denoted by  $\sigma_k(A_j)$ . The matrix

$$(A_j)_{r_j} = U_{A_j r_j} \Sigma_{A_j r_j} V_{A_j r_j}^T \quad (43)$$

is determined similarly to  $(A)_k$  given by (11) with the replacement of  $A$  and  $k$  by  $A_1$  and  $r_1$ , respectively.

*Theorem 4:* The solution to problem (6)–(7) is given by

$$\hat{G}_j = (A_j)_{r_j} C_j^\dagger + K_j \quad (44)$$

and the associated error is

$$\|\mathbf{x} - \mathbf{x}_{p+1}\|_{K,\Omega}^2 = \|R^{1/2}[\mathbf{x}\mathbf{x}^T]\|^2 - \sum_{j=1}^p \sum_{k=1}^{r_j} [\sigma_k(A_j)]^2. \quad (45)$$

**Proof.** We write  $\epsilon = \|\mathbf{x} - \mathbf{x}_{p+1}\|_{K,\Omega}^2$ . By an argument analogous to that which established (38) and (41) we have

$$\begin{aligned} \epsilon &= \|R^{1/2}[\mathbf{x}\mathbf{x}^T]\|_F^2 - \sum_{j=1}^p \|R[\mathbf{x}\mathbf{v}_j^T](R[\mathbf{v}_j\mathbf{v}_j^T])^{1/2\dagger}\|_F^2 \quad (46) \\ &\quad + \sum_{j=1}^p J_j(G_j) \quad (47) \end{aligned}$$

and

$$\min_{\text{rank } G_j \leq r_j \forall j} \sum_{j=1}^p \|A_j - G_j C_j\|_F^2 \quad (48)$$

$$= \sum_{j=1}^p \min_{\text{rank } G_j \leq r_j} \|A_j - G_j C_j\|_F^2. \quad (49)$$

By Corollary 1 the solution to

$$\min_{\text{rank } G_j \leq r_j} \|A_j - G_j C_j\|_F^2$$

is attained when

$$G_j = (A_j P_{C_j, R})_{r_j} V_{C_j}^T C_j^\dagger + K_j = (A_j)_{r_j} C_j^\dagger + K_j = \hat{G}_j.$$

We refer to [6] for more details. If  $\hat{G}_j$  is substituted instead of  $G_j$  in (46), then we have

$$\begin{aligned} \epsilon &= \|R^{1/2}[\mathbf{x}\mathbf{x}^T]\|_F^2 - \sum_{j=1}^p \|R[\mathbf{x}\mathbf{v}_j^T](R[\mathbf{v}_j\mathbf{v}_j^T])^{1/2\dagger}\|_F^2 \\ &\quad + \sum_{j=1}^p \|A_j - [(A_j)_{r_j} C_j^\dagger + K_j] C_j\|_F^2 \\ &= \|R^{1/2}[\mathbf{x}\mathbf{x}^T]\|_F^2 - \sum_{j=1}^p \|A_j\|_F^2 + \sum_{j=1}^p \|A_j - (A_j)_{r_j}\|_F^2 \\ &= \|R^{1/2}[\mathbf{x}\mathbf{x}^T]\|_F^2 - \sum_{j=1}^p \sum_{k=1}^{r_j} [\sigma_k(A_j)]^2 \quad (50) \end{aligned}$$

since  $(A_j)_{r_j} = (A_j)_{r_j} C_j^\dagger C_j$  by [18]. ■

*Remark 5:* In (45) the number  $p$  is the number of steps in the procedure described by (4)–(7). It follows from Theorem 4 that the error given by (45) decreases as  $p$  increases.

*Remark 6:* It follows from Theorems 3 and 4 that the components  $\tilde{G}_j$  and  $\hat{G}_j$  of the proposed filter are computed separately and require computation of  $m \times n$  and  $n \times n$  matrices for  $\tilde{G}_j$  and  $m \times r_j$  and  $r_j \times n$  matrices for  $\hat{G}_j$ . To the best of our knowledge, a filter that is able to process infinite signal sets is not previously known. Therefore we compare the computational loads needed for our proposed filters with those for known filters in the case of finite sets of signals only. In this case most known nonlinear filters [18] require computation of much larger matrices than those mentioned above. The procedures presented in Theorems 3 and 4 are advantageous from a numerical point of view because they require much less computational work.

### C. Choice of operators $\mathcal{P}_k$ for (1)–(7)

The purpose of using the operators  $\mathcal{P}_k$  in our filter has been discussed in Section I-A2. Here, we illustrate the choice of  $\mathcal{P}_k$  with some possible particular suggestions.

For instance,  $\mathcal{P}_k$  could be chosen using the Hadamard product as in [15], [16] or in the form

$$[\mathcal{P}_k(\mathbf{x}_1, \dots, \mathbf{x}_k)](\omega, \alpha) = \frac{1}{\gamma} \sum_{j=1}^k \gamma_j \mathbf{x}_j(\omega, \alpha)$$

where  $\gamma = \sum_{j=1}^k \gamma_j$  and  $\gamma_j \in \mathbb{R}$  is a constant.

Another possible choice for  $\mathcal{P}_k$  is a time shifting operator similar to that used in [12]. For the case considered in Example 4 of Section VI, we define

$$\begin{aligned} \mathcal{P}_k(\mathbf{x}_1, \dots, \mathbf{x}_k) &= \mathcal{P}_k[\mathbf{x}_k(\cdot, t_1), \dots, \mathbf{x}_k(\cdot, t_N)] \\ &= [\mathbf{x}_k(\cdot, t_1 - \Delta_1), \dots, \mathbf{x}_k(\cdot, t_N - \Delta_N)] \end{aligned}$$

with  $\Delta_k \in \mathbb{R}$  for all  $k = 1, \dots, p$ .

Alternatively  $\mathcal{P}_k$  could be chosen as a  $k$ -linear operator. In particular,  $\mathcal{P}_k$  could be a  $k$ -linear integral operator. We cite [2] in this regard.

D. Device for data compression and reconstruction

Let us denote  $D_j^{(1)} = U_{A_j r_j} \Sigma_{A_j r_j}$  and  $D_j^{(2)} = V_{A_j r_j}^T C_j^\dagger$ . See (42) and (43) in this regard. In (44) the matrix  $K_j$  is arbitrary. Let us assume that  $K_j = \mathbb{O}$ . Then the filter (1) based on solutions (35) and (44) to problems (4)–(5) and (6)–(7), respectively, is given by

$$F(y) = \sum_{j=1}^p D_j^{(1)} D_j^{(2)} v_j$$

where  $v_j = H_j x_j$ . Note that  $D_j^{(1)} \in \mathbb{R}^{m \times r_j}$  and  $D_j^{(2)} \in \mathbb{R}^{r_j \times n}$ . Thus  $D_j^{(2)}$  provides compression of a vector  $v_j$  with  $n$  components to a vector  $D_j^{(2)} v_j$  with  $r_j$  components while  $D_j^{(1)}$  provides decompression (reconstruction) from a vector with  $r_j$  components to a vector with  $m$  components. The compression ratio is given by

$$c = (r_1 + \dots + r_p) / \min\{m, n\}. \quad (51)$$

E. Advantages associated with filtering based on scheme (4)–(7)

1) *General advantages:* The proposed filter consists of the two parts. The first one is the filtering of a random signal  $\mathbf{y}$  based on the concatenation of solutions to  $p-1$  unconstrained error minimization problems given by (4)–(5). This filter is important in its own right since it can be considered as a stand alone filter aimed at optimal filtering of stochastic signals.

The second part is based on the solution of the constrained error minimization problem given by (6)–(7). As a result, while this part is still filtering the input, it also compresses and decompresses the input. The latter procedure is realized via the solution of the problem (6)–(7) and is described in Section III-B.

The overall advantages of our proposed approach are the ability to determine a *single* optimal filter to process an *infinite* set of signals and the progressive decrease in the filtering error with an increase in the number of steps (4)–(7).

The mathematical advantages include simplification of the numerical procedure for determining the filter components  $\hat{G}_1, \dots, \hat{G}_k$  in (4)–(6) which we address in Remark 6 above, the rigorous theoretical justification of the results given in Sections III-A and III-B and the effective procedure for signal compression and decompression presented in Section III-D.

In the following Section, we consider other specific advantages by comparison with known filters.

2) *Specific advantage: comparison with the known filters:*

As we have mentioned in Section I-A1, to the best of our knowledge, the previously known filters can process only finite sets of random signals. Our filter processes infinite sets of random signals. Therefore, a comparison between the known filters and the proposed one can only be done for the particular case when each set  $\mathbf{x}$  and  $\mathbf{y}$  (see Section II-A) consists of one signal only. In such a case, the results obtained above are presented in terms of the norm (54),  $\|\cdot\|_\Omega^2$  (see Section VI below). The norm (54) is a particular case of the norm (3),  $\|\cdot\|_{X,\Omega}^2$ .

The errors associated with the Karhunen-Loève filter (KLF) [18] and the Wiener filter [18] follow from (45) and (36),

respectively, if  $p = 1$ ,  $k = 1$  and the norm  $\|\cdot\|_{X,\Omega}^2$  is replaced with the norm  $\|\cdot\|_\Omega^2$ . The errors presented by (45) and (36) are less than those for the KLF and Wiener filter if  $p = 2, 3, \dots$  and  $k = 2, 3, \dots$ . Thus our filter provides improved accuracy when compared to the KLF and Wiener filter.

The compression ratio of the KLF is  $\eta / \min\{m, n\}$  where  $\eta$  is the number of components in the compressed signal. Thus, the compression ratio (51) of our filter is better than that of the KLF if  $r_1 + \dots + r_p < \eta$ .

A comparison with other known filters [18] can be done in a similar way and leads to similar conclusions.

IV. NUMERICAL EXAMPLES AND SIMULATIONS

The following elementary Example 1 illustrates the proposed filter performance in the case of infinite sets of signals.

*Example 1.* Let

$$\mathbf{x} = \{\mathbf{x}(\omega, t) = [\omega t^2, \omega^2 t]^T \mid \omega \in [0, 1], t \in [0, 1]\}$$

and

$$\mathbf{y} = \{\mathbf{y}(\omega, t) = [0.8\omega t^2, 1.2\omega^2 t]^T \mid \omega \in [0, 1], t \in [0, 1]\},$$

i.e.  $\mathbf{x}$  and  $\mathbf{y}$  are represented by infinite sets of signals. For such signals, the norm (3) is reduced to its particular case given by (56). Accordingly, matrices  $R[\mathbf{x}\mathbf{y}^T]$  and  $R[\mathbf{y}\mathbf{y}^T]$  given by (27)–(29) are determined in terms of the norm (56). For instance, entries of  $R[\mathbf{x}\mathbf{y}^T]$  are  $\langle \mathbf{x}^{(i)}, \mathbf{y}^{(j)} \rangle = \int_0^1 \int_0^1 \mathbf{x}^{(i)}(\omega, t) \mathbf{y}^{(j)}(\omega, t) d\omega dt$ . Therefore, we have

$$R[\mathbf{x}\mathbf{y}^T] = \begin{bmatrix} 0.053 & 0.075 \\ 0.050 & 0.080 \end{bmatrix}, \quad R[\mathbf{y}\mathbf{y}^T] = \begin{bmatrix} 0.011 & 0.060 \\ 0.060 & 0.096 \end{bmatrix}$$

and  $R[\mathbf{y}\mathbf{y}^T]^\dagger = \begin{bmatrix} -37.7358 & 23.5849 \\ 23.5849 & -4.3239 \end{bmatrix}$ .

Let us first consider a simplest case of our filter presented by (1) with  $p = 1$ ,  $\mathcal{H}_1$  and  $\mathcal{P}_1$  given by the identity, and  $G_1 = \tilde{G}_1$  where  $\tilde{G}_1$  is determined by (35) with  $M_1 = \mathbb{O}$ . Then for any  $\omega \in [0, 1]$  and  $t \in [0, 1]$ , the estimate of signals  $\mathbf{x}(\omega, t)$  by the considered particular case of our filter is given by

$$\tilde{\mathbf{x}}(\omega, t) = R[\mathbf{x}\mathbf{y}^T] R[\mathbf{y}\mathbf{y}^T]^\dagger [0.8\omega t^2, 1.2\omega^2 t]^T. \quad (52)$$

Note that (52) does not coincide with an estimate by the generalized Wiener filter which can process a fixed random signal-vector [18], but is not able to process infinite sets of signals. Differences between (52) and an estimate by the generalized Wiener filter [18] are matrices  $R[\mathbf{x}\mathbf{y}^T]$  and  $R[\mathbf{y}\mathbf{y}^T]^\dagger$  determined in terms of the norm (55) used in (52), and the term  $[0.8\omega t^2, 1.2\omega^2 t]^T$ .

Table 1 contains magnitudes of the error  $\Delta_1 = \|\mathbf{x}(\omega, t) - \tilde{\mathbf{x}}(\omega, t)\|_2^2$  with respect to some particular values of  $\omega$  and  $t$ .

Next, let us now consider other simplest case of the proposed filter when in (1), as before,  $p = 1$ ,  $\mathcal{H}_1$  and  $\mathcal{P}_1$  are the identity, but  $G_1 = \hat{G}_1$  where  $\hat{G}_1$  is determined by (44) with  $r_1 = 1$  and  $K_1 = \mathbb{O}$ . Then for any  $\omega \in [0, 1]$  and  $t \in [0, 1]$ , the estimate of signals  $\mathbf{x}(\omega, t)$  by this particular case of our filter is given by

$$\hat{\mathbf{x}}(\omega, t) = (A_1)_{r_1} C_1^\dagger [0.8\omega t^2, 1.2\omega^2 t]^T \quad (53)$$



where  $A_1 = R[\mathbf{xy}^T]R[\mathbf{yy}^T]^{1/2\dagger}$ ,  $C_1 = R[\mathbf{yy}^T]^{1/2}$  and  $(A_1)_{r_1}$  is determined by (43). Compression and decompression is realized by  $(A_1)_{r_1}C_1^\dagger$  as those described in Section III-D above. For the same values of  $\omega$  and  $t$  as those in Table 1, the error  $\Delta_2 = \|\mathbf{x}(\omega, t) - \hat{\mathbf{x}}(\omega, t)\|_2^2$  is worse than the error  $\Delta_1$  for approximately 20%. This is because  $\hat{\mathbf{x}}(\omega, t)$  is determined with the truncated SVD given by (43) while  $\tilde{\mathbf{x}}(\omega, t)$  is not.

The estimate  $\hat{\mathbf{x}}(\omega, t)$  does not coincide with an estimate by the KLF [18] by reasons which are similar to those described above for  $\tilde{\mathbf{x}}(\omega, t)$  regarding a comparison with the generalized Wiener filter.

Of course, in many instances, matrices  $R[\mathbf{xy}^T]$  and  $R[\mathbf{yy}^T]$  are unknown and should be estimated. Some estimation methods can be found in [18].

*Example 2.* Here, we consider a case when  $\mathbf{x}$  and  $\mathbf{y}$  are represented by finite signal sets and illustrate advantages of the proposed approach over the known methods. Such a case has shortly been discussed in Section I-A1.

Let  $\mathbf{x} = \{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(N)}\}$  and  $\mathbf{y} = \{\mathbf{y}_{(1)}, \mathbf{y}_{(2)}, \dots, \mathbf{y}_{(N)}\}$ , where  $N = 8$  and  $\mathbf{x}_{(j)}, \mathbf{y}_{(j)} \in L^2(\Omega, \mathbb{R}^n)$  with  $n = 256$  for each  $j = 1, \dots, 8$ . Random vectors  $\mathbf{y}_{(j)}$  and  $\mathbf{x}_{(j)}$  are interpreted as an input of the filter and its 'idealistic' output, respectively; i.e.  $\mathbf{x}_{(j)}$  is a signal that should be estimated by the filter. In this example,  $\mathbf{x}_{(j)}$  and  $\mathbf{y}_{(j)}$  are simulated as digital images presented by  $116 \times 256$  matrices  $X_{(j)}$  and  $Y_{(j)}$ , respectively. A column of the matrices  $X_{(j)}$  and  $Y_{(j)}$  represents a realization of signals  $\mathbf{x}_{(j)}$  and  $\mathbf{y}_{(j)}$ , respectively.

Each picture  $X_{(j)}$  in the sequence  $X_{(1)}, \dots, X_{(8)}$  has been taken at a certain time  $t_j$  with  $j = 1, \dots, 8$ . Images  $Y_{(1)}, \dots, Y_{(8)}$  have been simulated from  $X_{(1)}, \dots, X_{(8)}$  in the form  $Y_{(j)} = X_{(j)} \bullet \text{rand}_{(j)}$  for each  $j = 1, \dots, 8$ . Here,  $\bullet$  means the Hadamard product and  $\text{rand}_{(j)}$  is a  $116 \times 256$  matrix whose elements are uniformly distributed in the interval  $(0, 1)$ . Images  $X_{(1)}, \dots, X_{(8)}$  are presented in Fig. 1. Image  $Y_{(3)}$  is given in Fig. 2 (a) as an example of images  $Y_{(1)}, \dots, Y_{(8)}$ . Other images  $Y_{(j)}$  with  $j = 1, \dots, 8$ ,  $j \neq 3$  are similar.

For finite signal sets  $\mathbf{x}$  and  $\mathbf{y}$  considered in this example, the norm (55) is used. As a consequence, for  $j = 1, 2$ , matrix  $R[\mathbf{xv}_j^T]$  in (35) and (44) is presented by

$$R[\mathbf{xv}_j^T] = \{r_{ik}\}_{i,q=1}^{116,256}$$

$$\text{with } r_{iq} = \frac{1}{8} \sum_{k=1}^8 \int_{\Omega} \mathbf{x}^{(i)}(\omega, t_k) \mathbf{v}_j^{(q)}(\omega, t_k) d\mu(\omega).$$

Matrix  $R[\mathbf{v}_j \mathbf{v}_j^T]$  is determined similarly.

First, a simplest case of our filter defined from (4) by Theorem 3 with  $p = 1$  has been applied to the considered signal sets. Then  $R[\mathbf{xv}_1^T] = R[\mathbf{xy}^T]$  and  $R[\mathbf{v}_1 \mathbf{v}_1^T] = R[\mathbf{yy}^T]$ . In (1),  $G_1$  is determined as  $\tilde{G}_1$  by (35). An estimate of  $X_{(3)}$  by this filter is denoted by  $\tilde{X}_{1,(3)}$  and the filter itself is denoted by  $\tilde{\mathcal{F}}_1$ .

To illustrate the performance of the proposed filters associated with different steps of the scheme (5)-(7), their related versions have also been applied to the given signal sets.

Estimates of  $X_{(3)}$  by those filters are denoted as follows:

$\tilde{X}_{2,(3)}$  is the estimate by our filter  $\tilde{\mathcal{F}}_2$  defined by (5) and Theorem 3 with  $p = 3$ ,

$\hat{X}_{2,(3)}$  is the estimate by our filter  $\hat{\mathcal{F}}_2$  defined by (6)-(7) and Theorem 4 with  $p = 2$ ,  $r_1 = 10$  and  $r_2 = 15$ , with the compression ratio  $c = 25/116$ , and

$\bar{X}_{2,(3)}$  is the estimate by our filter  $\bar{\mathcal{F}}_2$  defined by (6)-(7) and Theorem 4 with  $p = 2$ ,  $r_1 = 4$  and  $r_2 = 5$ , with the compression ratio  $c = 9/116$ .

We point out again that by the proposed method, *the same* filter is applied to each pair of signals  $Y_{(j)}$  and  $X_{(j)}$  for  $j = 1, \dots, 8$ , i.e. its form presented by (1), (4)-(7) and Theorems 3, 4 is invariant with respect to different pairs  $(Y_{(j)}, X_{(j)})$ . In particular, matrices determined by (35) and (44) remain the same regardless to what pair  $(Y_{(j)}, X_{(j)})$  the filter is applied.

The known filters based on the Wiener filtering approach [18] should specifically be constructed for *each pair*  $(Y_{(j)}, X_{(j)})$ . As a result, such filters require much more computational loads. Therefore, in comparison of the proposed filters with the known ones, this important difference should be borne in mind. We denote by  $\tilde{X}_{2,(3)}$  and  $X_{KLF,(3)}$  the estimate of  $X_{(3)}$  by the Wiener filter and by the KLF of rank=25, respectively. The compression ratio of the KLF of rank=25 is  $c = 25/116$ .

In Fig. 2, estimates of signal  $X_{(3)}$  by different filters are presented. Numerical results related to estimation of a signal  $X_{(3)}$  are given in Table 2. Results associated with estimates of other signals  $X_{(j)}$  with  $j = 1, \dots, 8$ ,  $j \neq 3$  are similar.

In particular, while the error associated with the filter  $\tilde{\mathcal{F}}_1$  is  $\|X_{(3)} - \tilde{X}_{1,(3)}\|^2 = 0.9893e+07$ , the error  $\|X_{(3)} - \tilde{X}_{2,(3)}\|^2$  associated with the filter  $\tilde{\mathcal{F}}_2$  is significantly lesser. This is a numerical illustration to Remark ??.

Next, we observe, that for the same compression ratio,  $c = 25/116$ , the error  $\|X_{(3)} - \hat{X}_{2,(3)}\|^2$  associated with the filter  $\hat{\mathcal{F}}_2$ , is four times less than the error  $\|X_{(3)} - X_{KLF,(3)}\|^2$  associated with the KLF.

It is interesting to compare compression ratios of those filters in the case when associated error are similar. In particular, it follows from the third and fifth columns of Table 2, that while the errors associated with the filter  $\tilde{\mathcal{F}}_2$  and the KLF are almost the same, the compression ratio of the filter  $\tilde{\mathcal{F}}_2$  (i.e.  $c = 9/116$ ) is more than two times better than the KLF compression ratio (which is  $c = 25/116$ ).

## V. CONCLUSION

We have provided the theory for a new approach to filtering, compression and decompression for infinite sets of stochastic signals. Distinctive features of the proposed approach are as follows. While we consider processing infinite signal sets, the proposed filter is nevertheless fixed for all signal pairs. The filter is nonlinear and consists of  $p$  steps. Each step contains three components represented by three consecutive operations. The various operations are determined at each step by an iterative scheme designed to improve filter performance as the number of steps increases. Signal compression and decompression is provided by the final step of the scheme.

The final step is based on a new method [6] for the best rank-constrained matrix approximation. The error associated with our filter decreases when the number of steps in the iterative scheme increases.

#### VI. APPENDIX A: PARTICULAR CASES OF THE NORM (3)

Here, we consider some particular norms that lead to the norm (3) used in our statement of the problem in Sections II-C–II-D.

*Example 3.* Let  $\mathbf{x} \in L^2(\Omega, \mathbb{R}^m)$ . Then we set

$$\|\mathbf{x}\|_{\Omega}^2 = \int_{\Omega} \|\mathbf{x}(\omega)\|_2^2 d\mu(\omega). \quad (54)$$

Note, that most of results related to Wiener-like optimal filtering have been obtained using the norm (54). Some relevant references can be found in [18].

*Example 4.* Let  $\mathbf{x} = \{\mathbf{x}(\cdot, t_k) \in L^2(\Omega, \mathbb{R}^m) \mid t_k \in \mathbb{R} \forall k = 1, \dots, N\}$ . Thus,  $\mathbf{x} \in L^2(\Omega \times [t_1, \dots, t_N], \mathbb{R}^m)$ . In other words, in the space  $L^2(\Omega \times [t_1, \dots, t_N], \mathbb{R}^m)$ ,  $\mathbf{x}$  is a fixed signal, but in the space  $L^2(\Omega, \mathbb{R}^m)$ ,  $\mathbf{x}$  is represented by the set of signals  $\{\mathbf{x}(\cdot, t_k) \in L^2(\Omega, \mathbb{R}^m) \mid t_k \in \mathbb{R} \forall k = 1, \dots, N\}$ .

Let us put

$$\|\mathbf{x}\|_{[t_1, \dots, t_N], \Omega}^2 = \frac{1}{N} \sum_{k=1}^N \int_{\Omega} \|\mathbf{x}(\omega, t_k)\|_2^2 d\mu(\omega). \quad (55)$$

We note that  $\|\mathbf{x}\|_{[t_1, \dots, t_N], \Omega}^2$  can be represented as

$$\|\mathbf{x}\|_{[t_1, \dots, t_N], \Omega}^2 = \frac{1}{N} \sum_{k=1}^N \|\mathbf{x}(\cdot, t_k)\|_{\Omega}^2,$$

where

$$\|\mathbf{x}(\cdot, t_k)\|_{\Omega}^2 = \int_{\Omega} \|\mathbf{x}(\omega, t_k)\|_2^2 d\mu(\omega).$$

The norm (54) follows from (55) if the set  $\mathbf{x}$  consists of a single signal,  $\mathbf{x}(\cdot, t_k)$ , with  $t_k$  fixed.

*Example 5.* Let  $\mathbf{x} = \{\mathbf{x}(\cdot, t) \in L^2(\Omega, \mathbb{R}^m) \mid t \in [a, b] \subset \mathbb{R}\}$ . Thus  $\mathbf{x} \in L^2(\Omega \times [a, b], \mathbb{R}^m)$ . We set

$$\|\mathbf{x}\|_{[a, b], \Omega}^2 = \frac{1}{b-a} \int_a^b \int_{\Omega} \|\mathbf{x}(\omega, t)\|_2^2 d\mu(\omega) dt. \quad (56)$$

Similarly to Example 4,

$$\|\mathbf{x}\|_{[a, b], \Omega}^2 = \frac{1}{b-a} \int_a^b \|\mathbf{x}(\cdot, t)\|_{\Omega}^2 dt.$$

The norm (54) follows from (56) if the set  $\mathbf{x}$  consists of a single signal,  $\mathbf{x}(\cdot, t)$ , with  $t$  fixed.

*Example 6.* Let  $\mathbf{x} = \{\mathbf{x}(\cdot, \tau) \in L^2(\Omega, \mathbb{R}^m) \mid \tau \in C^q \subset \mathbb{R}^q\}$  where  $C^q$  is a  $q$ -dimensional cube in  $\mathbb{R}^q$ . We put

$$\|\mathbf{x}\|_{C^q, \Omega}^2 = \frac{1}{V} \int_{C^q} \int_{\Omega} \|\mathbf{x}(\omega, \tau)\|_2^2 d\mu(\omega) d\tau = \frac{1}{V} \int_{C^q} \|\mathbf{x}(\cdot, \tau)\|_{\Omega}^2 d\tau, \quad (57)$$

where  $V = \int_{C^q} d\tau$ . Then  $\|\mathbf{x}\|_{\Omega}^2$  follows from  $\|\mathbf{x}\|_{C^q, \Omega}^2$  if the set  $\mathbf{x}$  consists of a single signal,  $\mathbf{x}(\cdot, \tau)$ , with  $\tau$  fixed.

The norms given by (54)–(57) are generalized in the following way. Let  $\mathbf{x} = \{\mathbf{x}(\cdot, \alpha) \in L^2(\Omega, \mathbb{R}^m) \mid \alpha \in K\}$  where  $K$  is a measurable set [9] with a measure  $\lambda(\alpha)$ . Thus,  $\mathbf{x}$  is a single signal in the space  $L^2(\Omega \times K, \mathbb{R}^m)$  and is the infinite

set of signals  $\{\mathbf{x}(\cdot, \alpha) \in L^2(\Omega, \mathbb{R}^m) \mid \alpha \in K\}$  in the space  $L^2(\Omega, \mathbb{R}^m)$ .

Let us set the norm by

$$\|\mathbf{x}\|_{K, \Omega}^2 = \frac{1}{\lambda(K)} \int_K \int_{\Omega} \|\mathbf{x}(\omega, \alpha)\|_2^2 d\mu(\omega) d\lambda(\alpha), \quad (58)$$

where  $\lambda(K) = \int_K d\lambda(\mathbf{x})$ . Then the norms given by (54)–(57) follow from  $\|\mathbf{x}\|_{K, \Omega}^2$  for a suitable choice of  $K$  and  $\lambda(\alpha)$ . In particular,  $\|\mathbf{x}\|_{\Omega}^2$  follows from  $\|\mathbf{x}\|_{K, \Omega}^2$  if  $X$  consists of a single signal only,  $\mathbf{x}(\cdot, \alpha)$ , with  $\alpha$  fixed.

#### REFERENCES

- [1] T. L. Boullion and P. L. Odell, *Generalized Inverse Matrices*, John Wiley & Sons, Inc., New York, 1972.
- [2] N. Dunford and J. T. Schwartz, *Linear Operators, Part I, General Theory*, Wiley Classics Library, Wiley, New York, 1988.
- [3] C. Eckart and G. Young, The Approximation of One Matrix by Another of Lower Rank, *Psychometrika*, **1**, 211–218, 1936.
- [4] J.H. Ferziger, *Numerical Methods for Engineering Applications*, Wiley, New-York, 1999.
- [5] V. N. Fomin and M. V. Ruzhansky, Abstract optimal linear filtering, *SIAM J. Control Optim.*, **38**, pp. 1334–1352, 2000.
- [6] S. Friedland and A. P. Torokhti, Generalized rank-constrained matrix approximations, *SIAM J. Matrix Anal. Appl.*, **29**, issue 2, pp. 656–659, 2007.
- [7] G.H. Golub and C.F. Van Loan, *Matrix Computation*, Johns Hopkins Univ. Press, 3rd Ed., 1996.
- [8] S. Haykin, *Adaptive Filter Theory*, Prentice–Hall, Englewood Cliffs, N. J., 1991.
- [9] V. Hutson and J.S. Pym, *Applications of Functional Analysis and Operator Theory*, Academic Press, London, 1980.
- [10] I.T. Jolliffe, *“Principal Component Analysis,”* Springer Verlag, New York, 1986.
- [11] A.N. Kolmogorov and S.V. Fomin, *Elements of the Theory of Functions and Functional Analysis*, Dover, New York, 1999.
- [12] J. Manton and Y. Hua, Convolution reduced rank Wiener filtering, *Proc. of ICASSP’01*, **6**, pp. 4001–4004, 2001.
- [13] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*, J. Wiley & Sons, 2001.
- [14] L.L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, New York: Addison-Wesley Publishing Co., 1990.
- [15] A. Torokhti and P. Howlett, Optimal fixed rank transform of the second degree, *IEEE Trans. on Circuits and Systems. Part II, Analog & Digital Signal Processing*, **48**, 309–315, 2001.
- [16] A. Torokhti and P. Howlett, Method of recurrent best estimators of second degree for optimal filtering of random signals, *Signal Processing*, **83**, 5, 1013 – 1024, 2003.
- [17] A. Torokhti and P. Howlett, Best approximation of identity mapping: the case of variable memory, *J. Approx. Theory*, **143**, 1, 111–123, 2006.
- [18] A. Torokhti and P. Howlett, *Computational Methods for Modelling of Nonlinear Systems*, Elsevier, 2007.
- [19] N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications*, Academic Press, New York, 1949.

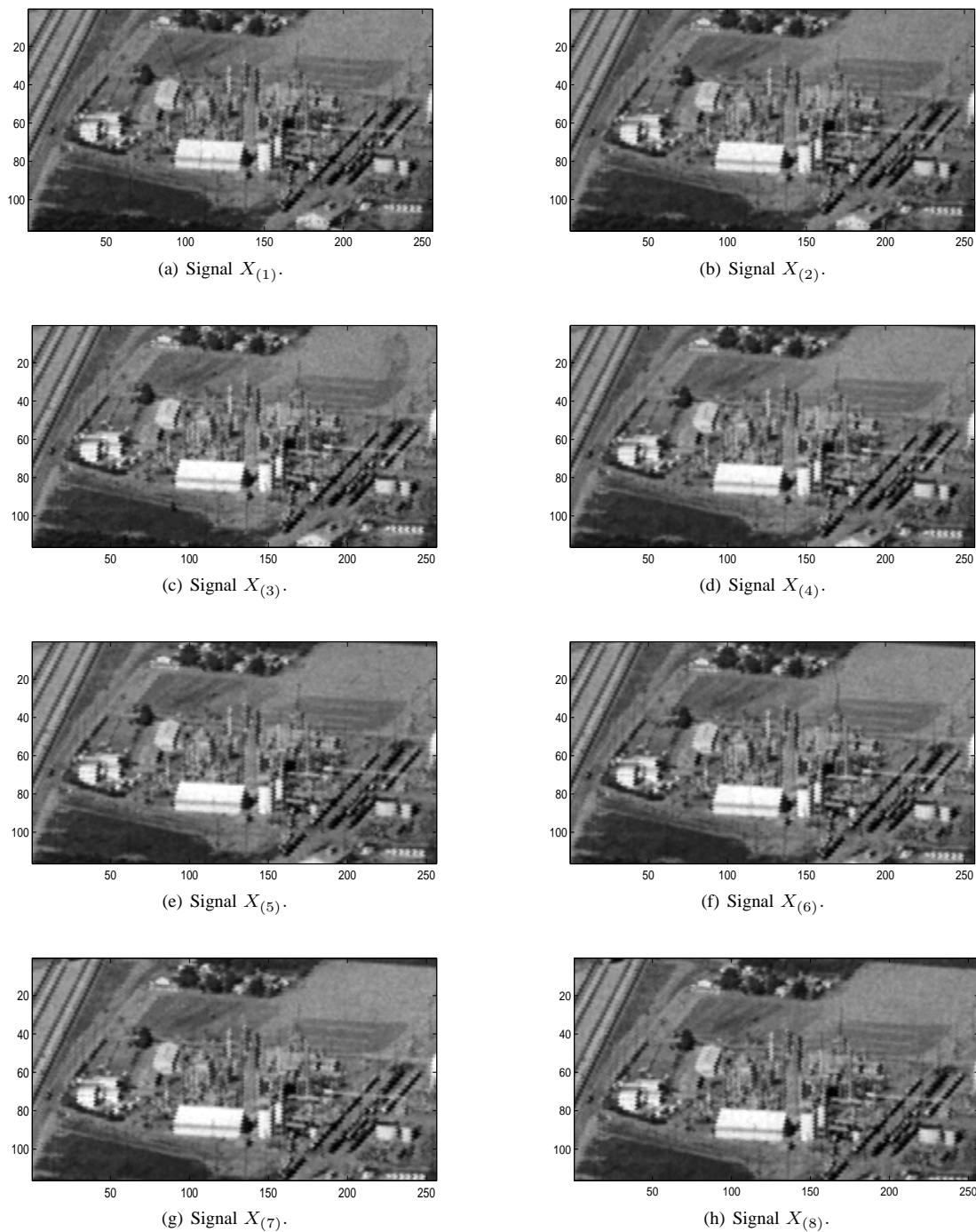


Fig. 1. Signals  $X_1, \dots, X_8$  to be estimated from observed data.

Table 1.

$\omega$	0.5	0.4	0.4	0.2	0.9
$t$	0.5	0.7	0.3	0.8	0.9
$\Delta_1$	$0.86 \times 10^{-4}$	$1.16 \times 10^{-2}$	$1.13 \times 10^{-4}$	$1.35 \times 10^{-2}$	$2.90 \times 10^{-3}$

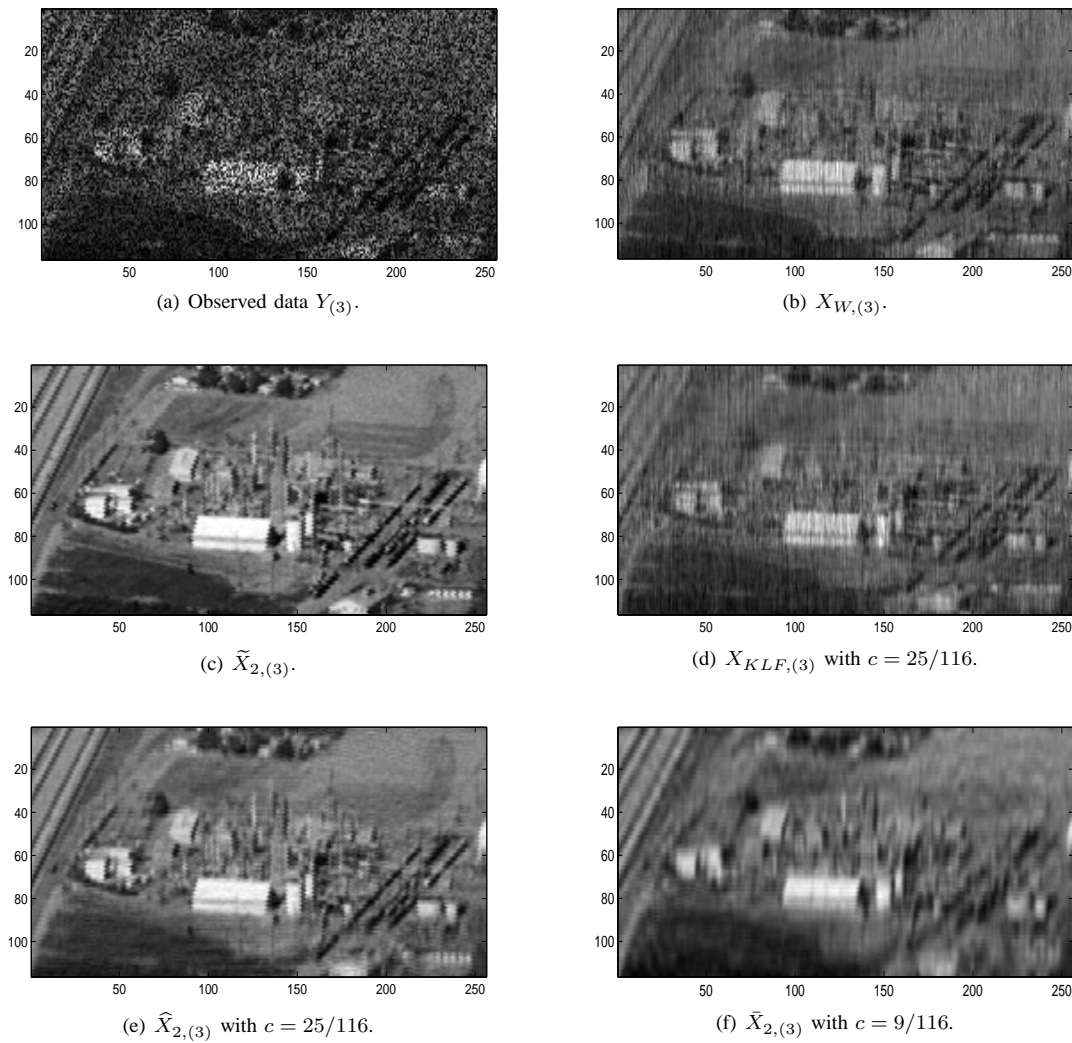


Fig. 2. Illustration to signal  $X_{(3)}$  estimation by different filters. Here,  $X_{W,(3)}$  is the estimate by the Wiener filter,  $\tilde{X}_{2,(3)}$  is the estimate that follows from (5) with  $p = 3$ ,  $X_{KLF,(3)}$  is the estimate by KLF with  $c = 25/116$ ,  $\hat{X}_{2,(3)}$  is the estimate that follows from (6)-(7) with  $c = 25/116$ , and  $\bar{X}_{2,(3)}$  is the estimate that follows from (6)-(7) with  $c = 9/116$ .

Table 2.

$\ X_{(3)} - X_{W,(3)}\ ^2$	$\ X_{(3)} - \tilde{X}_{2,(3)}\ ^2$	$\ X_{(3)} - X_{KLF,(3)}\ ^2$ $c = 25/116$	$\ X_{(3)} - \hat{X}_{2,(3)}\ ^2$ $c = 25/116$	$\ X_{(3)} - \bar{X}_{2,(3)}\ ^2$ $c = 9/116$
$0.7555e + 07$	$1.5089e - 07$	$8.8739e + 06$	$2.1730e + 06$	$8.2173e + 06$