14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

OPTIMAL LINEAR FILTERING WITH PIECEWISE-CONSTANT MEMORY

Anatoli Torokhti and Phil Howlett

University of South Australia, School of Mathematics and Statistics
1 Mawson Lakes Blvr., SA 5095, Adelaide, Australia
phone: +61 8 8302 3812, fax: +61 8 8302 5785, email: anatoli.torokhti@unisa.edu.au
web: http://people.unisa.edu.au/Anatoli.Torokhti

ABSTRACT

The paper concerns the optimal linear filtering of stochastic signals associated with the notion of piecewise constant memory. The filter should satisfy a specialized criterion formulated in terms of a so called lower stepped matrix A. To satisfy the special structure of the filter, we propose a new technique based on a block-partition of the lower stepped part of matrix A into lower triangular and rectangular blocks, L_{ij} and R_{ij} with $i = 1,\ldots,l$, $j = 1,\ldots,s_i$ where l and s_i are given. We show that the original error minimization problem in terms of the matrix A is reduced to l individual error minimization problems in terms of blocks L_{ij} and R_{ij}. The solution to each problem is provided and a representation of the associated error is given.

1. INTRODUCTION

While the general theory of optimal filtering is well elaborated (see, e.g., [1]), the theory of optimal *constrained* filtering is still not so well developed, although this is an area of intensive recent research (see, e.g., [2]). Despite increasing demands from applications, this subject is hardly tractable because of intrinsic difficulties in computing techniques, when the filter should have a specific structure implied by the underlying problem.

This paper concerns the theory of optimal linear filtering subject to a specialized criterion associated with the notion of piece-wise constant memory. The problem stems from an observation considered in Section 1.2. A formulation of the problem is given in Section 3. The solution is provided in Section 5.

1.1 Preliminary notation

LetΩ be the set of outcomes in a probability space (Ω,Σ,μ) for which Σ is a σ–field of measurable subsets of Ω and $\mu : \Sigma \to [0,1]$ is an associated probability measure with $\mu(\Omega) = 1$. The random variables $\mathbf{x}_k : \Omega \to \mathbb{R}$ and $\mathbf{y}_k : \Omega \to \mathbb{R}$ are measurable functions on Ω for each $\omega \in \Omega$ and $k = 1,2,\ldots,n$. If $\mathbf{x}_k$ and $\mathbf{y}_k$ are square integrable for each $k = 1,2,\ldots,n$ then the square integrable random vectors $\mathbf{x} \in L^2(\Omega,\mathbb{R}^n)$ and $\mathbf{y} \in L^2(\Omega,\mathbb{R}^n)$ are denoted by $\mathbf{x} = [\mathbf{x}_1,\ldots,\mathbf{x}_n]^T$ and $\mathbf{y} = [\mathbf{y}_1,\ldots,\mathbf{y}_n]^T$. We write

$$x_k = \mathbf{x}_k(\omega), \quad y_k = \mathbf{y}_k(\omega), \quad x = \mathbf{x}(\omega), \quad y = \mathbf{y}(\omega) \tag{1}$$

$$x = [x_1,\ldots,x_n]^T \quad \text{and} \quad y = [y_1,\ldots,y_n]^T. \tag{2}$$

Let $A \in \mathbb{R}^{n\times n}$ and let $\mathscr{A} : L^2(\Omega,\mathbb{R}^n) \to L^2(\Omega,\mathbb{R}^n)$ be a linear filter defined by the formula

$$[\mathscr{A}(\mathbf{y})](\omega) = A[\mathbf{y}(\omega)] \quad \forall \quad \mathbf{y} \in L^2(\Omega,\mathbb{R}^n) \text{ and } \omega \in \Omega \tag{3}$$

so that

$$\tilde{\mathbf{x}} = \mathscr{A}(\mathbf{y}) \quad \text{where } \tilde{\mathbf{x}} = [\tilde{\mathbf{x}}_1,\ldots,\tilde{\mathbf{x}}_n]^T.$$

Next, let us partition $\tilde{\mathbf{x}}$ in such a way that

$$\tilde{\mathbf{x}} = [\tilde{\mathbf{u}}_1^T, \tilde{\mathbf{u}}_2^T, \ldots, \tilde{\mathbf{u}}_l^T]^T, \tag{4}$$

where $\tilde{\mathbf{u}}_i = [\tilde{\mathbf{x}}_{p_1+\ldots+p_{i-1}+1},\ldots,\tilde{\mathbf{x}}_{p_1+\ldots+p_i}]^T$, $i = 1,\ldots,l$, $p_0 = 0$, $\tilde{\mathbf{u}}_i \in L^2(\Omega,\mathbb{R}^{p_i})$, and $p_1+\ldots+p_l = n$.

1.2 The underlying problem

We interpret random vectors $\mathbf{y}$ and $\mathbf{x}$ as observable data and reference vector, respectively. It is assumed that $\mathbf{y}$ contains $\mathbf{x}$ and is contaminated with a random noise, and it is required to find A so that $\mathscr{A}(\mathbf{y})$ estimates $\mathbf{x}$ in the best possible in terms of minimizing the mean square error. Moreover, to determine a best $\tilde{\mathbf{u}}_i$ in (4), the filter $\mathscr{A}$ may transform no more than $m(i)$ components $\mathbf{y}_{s_i},\ldots,\mathbf{y}_{p_1+\ldots+p_i}$ of $\mathbf{y}$, where

$$m_i = (p_1+\ldots+p_i) - s_i + 1, \quad q_i = 1,2,\ldots,(p_1+\ldots+p_i),$$
$$s_i = q_i, q_i+1,\ldots,(p_1+\ldots+p_i) \quad \text{and} \quad i = 1,\ldots,l.$$

Such an filter $\mathscr{A}$ is called the filter with piecewise-constant memory $\{m_1,\ldots,m_l\}$.

The above constraint implies that the filter $\mathscr{A}$ and consequently the matrix A, must have a compatible structure. Essential conditions are that the components $\tilde{\mathbf{x}}_{p_1+\ldots+p_i}$ and $\mathbf{y}_{p_1+\ldots+p_i}$ have the same subscript and that s_i in (5) is different for each i, i.e., for each $\tilde{\mathbf{u}}_i$ in (4). This respectively means that all entries above the diagonal of the matrix A are zeros and second, that for each i, there can be a zero-rectangular block in A from the left hand side of the diagonal.

An example of such a matrix A is given in Fig. 1 for $l = 10$ where the shaded part designates non-zero entries and non-shaded parts denote zero entries of A (and where p_1+p_2 denotes a (p_1+p_2)-th row, etc.). For lack of a better name, we will refer to A similar to that in Fig. 1 as the lower stepped matrix. We say that non-zero entries of the matrix A form a lower stepped part of A.

Such an unusual structure of the filter $\mathscr{A}$ makes the problem of finding the best $\mathscr{A}$ quite specific. This subject has a long history [3], but to the best of our knowledge, even for a much simpler structure of the filter $\mathscr{A}$ when $\mathscr{A}$ is defined by a lower triangular matrix, the problem of determining the best $\mathscr{A}$ has only been solved under the hard assumption of positive definiteness of an associated covariance matrix (see [3, 4, 5]). We avoid such an assumption and solve the problem in the general case of the lower stepped matrix (Theorem 1). The proposed technique is substantially different from those considered in [3, 4, 5].

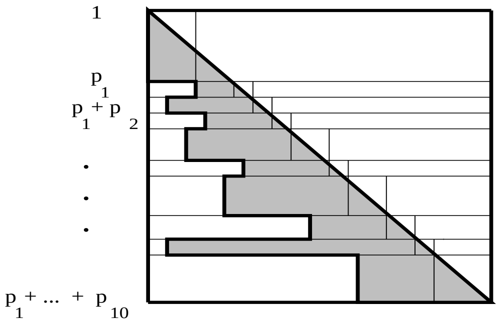

Figure 1: A lower stepped matrix and its partition.

2. LINEAR CAUSAL FILTER WITH PIECEWISE-CONSTANT MEMORY

To define a linear causal filters with piece-wise constant memory, we first need to formally define a lower stepped matrix. It is done below with a special partition of A in such a way that its lower stepped part consists from rectangular and lower triangular blocks as it is illustrated in Fig. 1. To realize such a representation, we need to choose a non-uniform partition of A in a form similar to that in Fig. 1.

The block-matrix representation for $\mathscr{A}$ is as follows. Let

$$A = \{A_{ij} \mid A_{ij} \in \mathbb{R}^{p_i \times q_{ij}}, i = 1, \ldots, l, j = 1, \ldots, 4\}, \quad (5)$$

where $p_1 + \ldots + p_l = n$ and $q_{i1} + \ldots + q_{i4} = n$.

Let $\emptyset$, $\mathbb{O}_{ij} \in \mathbb{R}^{p_i \times q_{ij}}$, $L_{ij} \in \mathbb{R}^{p_i \times q_{ij}}$ and $R_{ij} \in \mathbb{R}^{p_i \times q_{ij}}$ be the empty block, zero block, lower triangular block and rectangular block, respectively.

We write $A = \begin{bmatrix} A_1 \\ \vdots \\ A_l \end{bmatrix}$, where $A_i = [A_{i1}, \ldots, A_{i4}]$ for each $i = 1, \ldots, l$. Here, A_i is called the block-row.

Now, let

$$A_1 = [\emptyset, \emptyset, L_{13}, \mathbb{O}_{14}], \quad A_i = [\mathbb{O}_{i1}, R_{i2}, L_{i3}, \mathbb{O}_{i4}]$$

and

$$A_{l1} = [\mathbb{O}_{l1}, R_{l2}, L_{l3}, \emptyset],$$

where $i = 2, \ldots, l-1$.

For $i = 1, \ldots, l-1$, we also set

$$m_1 = q_{13}, \quad q_{i3} = p_i, \quad m_{i+1} = q_{i+1,2} + p_{i+1} \quad (6)$$
$$\text{and} \quad q_{i+1,1} + q_{i+1,2} = q_{i,1} + m_i, \quad (7)$$

where $q_{11} = 0$. Then the matrix A is represented as follows:

$$A = \begin{bmatrix} L_{13} & & & \mathbb{O}_{14} & \\ \mathbb{O}_{21} & R_{22} & L_{23} & & \mathbb{O}_{24} \\ \vdots & \ddots & \ddots & & \vdots \\ \mathbb{O}_{l-1,1} & & R_{l-1,2} & L_{l-1,3} & \mathbb{O}_{l-1,4} \\ & \mathbb{O}_{l1} & & R_{l2} & L_{l3} \end{bmatrix} \quad (8)$$

Definition 1 *The matrix A given by (8) is called a lower stepped matrix. The set of lower stepped matrices is denoted by $\mathbb{L}^n_m$.*

Definition 2 *The linear filter $\mathscr{A} : L^2(\Omega, \mathbb{R}^n) \to L^2(\Omega, \mathbb{R}^n)$ is called a causal filter with piece-wise constant memory $\{m_1, \ldots, m_l\}$ where*

$$m_i = \begin{cases} q_{13} & \text{if} \quad i = 1, \\ q_{i2} + q_{i3} & \text{if} \quad i = 2, \ldots, l, \end{cases} \quad (9)$$

if $\mathscr{A}$ is defined by the lower stepped matrix $A \in \mathbb{R}^{n \times n}$ given by (8). The set of such filters is denoted by $\mathbb{A}^n_m$.

3. STATEMENT OF THE PROBLEM

For any $\mathbf{x}, \mathbf{y} \in L^2(\Omega, \mathbb{R}^n)$ and $\mathscr{A} \in \mathbb{A}^n_m$, let

$$J(A) = E\left[\|\mathbf{x} - \mathscr{A}(\mathbf{y})\|^2\right], \quad (10)$$

where

$$E\left[\|\mathbf{x} - \mathscr{A}(\mathbf{y})\|^2\right] = \int_\Omega \|\mathbf{x}(\omega) - [\mathscr{A}(\mathbf{y})](\omega)\|^2_E d\mu(\omega)$$

with $\|\cdot\|_E$ the Euclidean norm.

The problem is to find a filter $\mathscr{A}^0 \in \mathbb{A}^n_m$ such that

$$J(A^0) = \min_{A \in \mathbb{L}^n_m} J(A). \quad (11)$$

Here, $[\mathscr{A}^0(\mathbf{y})](\omega) = A^0[\mathbf{y}(\omega)]$ and $A \in \mathbb{L}^n_m$.

It is assumed that $\mathbf{x}$ is unknown and no relationship between $\mathbf{x}$ and $\mathbf{y}$ is known except covariance matrices or their estimates formed from subvectors of $\mathbf{y}$ and $\mathbf{x}$. We note that similar assumptions are conventional for the known methods [1]-[7] concerning filtering of stochastic signals. The methods of a covariance matrix estimation can be found in [6].

4. AUXILIARY RESULTS

The solution of the problem (11) given below, consists of the following steps. First, vector $\mathbf{y}$ is partitioned in subvectors $\mathbf{v}_{13}, \mathbf{v}_{22}, \mathbf{v}_{23}, \ldots, \mathbf{v}_{l2}, \mathbf{v}_{l3}$ in a way which is compatible with the partition of matrix A in (8). Then the original problem can be represented as l independent problems (26)–(27). Second, to solve the problems (26)–(27), orthogonalization of subvectors $\mathbf{v}_{13}, \mathbf{v}_{22}, \mathbf{v}_{23}, \ldots, \mathbf{v}_{l2}, \mathbf{v}_{l3}$ is used. Finally, in Theorem 1, the solution of the original problem is derived in terms of matrices formed from orthogonalized subvectors.

We begin with partitions of $\mathbf{x}$ and $\mathbf{y}$.

4.1 Compatible representation of $\mathscr{A}(\mathbf{y})$

Partitions of $\mathbf{x}$ and $\mathbf{y}$ which are compatible with the partition of matrix A above are as follows.

We write

$$x = [u_1^T, u_2^T, \ldots, u_l^T]^T \quad \text{and} \quad \mathbf{x} = [\mathbf{u}_1^T, \mathbf{u}_2^T, \ldots, \mathbf{u}_l^T]^T \quad (12)$$

where $u_1 \in \mathbb{R}^{p_1}$, $u_2 \in \mathbb{R}^{p_2}$, …, $u_l \in \mathbb{R}^{p_l}$ are such that

$$u_1 = [x_1, \ldots, x_{p_1}]^T, \quad u_2 = [x_{p_1+1}, \ldots, x_{p_1+p_2}]^T, \quad (13)$$
$$u_l = [x_{p_1+\ldots+p_{l-1}+1}, \ldots, x_{p_1+\ldots+p_l}]^T, \quad (14)$$

and $\mathbf{u}_1 \in L^2(\Omega, \mathbb{R}^{p_1})$, $\mathbf{u}_2 \in L^2(\Omega, \mathbb{R}^{p_2})$, …, $\mathbf{u}_l \in L^2(\Omega, \mathbb{R}^{p_l})$ are defined via $u_1, u_2, \ldots, u_l$ similarly to (1).

Next, let $v_{11} = \emptyset$, $v_{12} = \emptyset$, $v_{13} = [y_1, \ldots, y_{q_{13}}]^T$ and $v_{14} = \emptyset$.

For $i = 2, \ldots, l-1$, we set

$$v_{i1} = [y_1, \ldots, y_{q_{i1}}]^T, \quad v_{i2} = [y_{q_{i1}+1}, \ldots, y_{q_{i1}+q_{i2}}]^T,$$
$$v_{i3} = [y_{q_{i1}+q_{i2}+1}, \ldots, y_{q_{i1}+q_{i2}+q_{i3}}]^T, \; v_{i4} = [y_{q_{i1}+q_{i2}+q_{i3}+1}, \ldots, y_n]^T.$$

If $i = l$, then

$$v_{l1} = [y_1, \ldots, y_{q_{l1}}]^T, \quad v_{l2} = [y_{q_{l1}+1}, \ldots, y_{q_{l1}+q_{l2}}]^T,$$
$$v_{l3} = [y_{q_{l1}+q_{l2}+1}, \ldots, y_n]^T, \quad v_{l4} = \emptyset.$$

Therefore

$$Ay = \begin{bmatrix} L_{13}v_{13} \\ R_{22}v_{22} + L_{23}v_{23} \\ \vdots \\ R_{l2}v_{l2} + L_{l3}v_{l3} \end{bmatrix}. \tag{15}$$

We define $\mathscr{L}_{ij}$ and $\mathscr{R}_{ij}$ via L_{ij} and R_{ij} respectively, in the manner of $\mathscr{A}$ defined via A by (3). The vector $\mathbf{v}_{ij} \in L^2(\Omega, \mathbb{R}^{q_{ij}})$ are defined similarly to those in (1).

Now, we can write $J(A)$ given by (10), in the form

$$J(A) = J_1(L_{13}) + \sum_{i=2}^{l} J_i(R_{i2}, L_{i3}) \tag{16}$$

where

$$J_1(L_{13}) = E\left[\|\mathbf{u}_1 - \mathscr{L}_{13}(\mathbf{v}_{13})\|^2\right]$$

and

$$J_i(R_{i2}, L_{i3}) = E\left[\|\mathbf{u}_i - [\mathscr{R}_{i2}(\mathbf{v}_{i2}) + \mathscr{L}_{i3}(\mathbf{v}_{i3})]\|^2\right]. \tag{17}$$

We note that matrix A can be represented so that

$$Ay = BPy,$$

where

$$B \in \mathbb{R}^{n\times q} \quad \text{and} \quad P \in \mathbb{R}^{q\times n}$$

with

$$q = q_{13} + \sum_{i=1}^{l}(q_{i2} + q_{i3})$$

are such that

$$B = \begin{bmatrix} L_{13} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \ldots & \mathbb{O} & \mathbb{O} \\ \mathbb{O} & R_{22} & L_{23} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \ldots & \mathbb{O} & \mathbb{O} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & & \vdots & \mathbb{O} \\ \mathbb{O} & \ldots & \ldots & \ldots & \mathbb{O} & R_{l-1,2} & L_{l-1,3} & \mathbb{O} & \mathbb{O} \\ \mathbb{O} & \ldots & \ldots & \ldots & \mathbb{O} & \mathbb{O} & \mathbb{O} & R_{l2} & L_{l3} \end{bmatrix} \tag{18}$$

and $Py = \begin{bmatrix} v_1 \\ \vdots \\ v_l \end{bmatrix}$. Here, $\mathbb{O}$ is the zero block, $v_1 = v_{13}$ and $v_i = \begin{bmatrix} v_{i2} \\ v_{i3} \end{bmatrix}$ for $i = 2, \ldots, l-1$. The size of each zero block is such that BPy is represented in the form (15). The matrix B consists of $l \times (2l-1)$ blocks. The vector $v = Py$ consists of $2l-1$ subvectors v_{13}, v_{22}, v_{23}, $\ldots$, v_{l2}, v_{l3}.

The filter $\mathscr{A}$ can be written as

$$\mathscr{A}(\mathbf{y}) = \mathscr{B}\mathscr{P}(\mathbf{y})$$

where

$$[\mathscr{B}(\mathbf{v})](\omega) = B[(\mathbf{v})(\omega)], \quad \mathbf{v} = \mathscr{P}(\mathbf{y})$$

and

$$[\mathscr{P}(\mathbf{y})](\omega) = P[(\mathbf{y})(\omega)].$$

4.2 Orthogonality of random vectors

For any $\mathbf{x}, \mathbf{y} \in L^2(\Omega, \mathbb{R}^n)$, we denote

$$E_{xy} = E[\mathbf{x}\mathbf{y}^T] = \left\{E[\mathbf{x}_i \mathbf{y}_j]\right\}_{i,j=1}^{n}$$

where $E[\mathbf{x}_i\mathbf{y}_j] \stackrel{\text{def}}{=} \int_\Omega \mathbf{x}_i(\omega)\mathbf{y}_j(\omega)d\mu(\omega)$. The pseudo-inverse matrix for any matrix M is denoted by $M^\dagger$.

Definition 3 [6, 7] *Let $\mathbf{w}_{ij} \in L^2(\Omega, \mathbb{R}^{q_{ij}})$ for each $i = 1, \ldots, l$ and $j = 1, \ldots, 4$. The random vectors $\mathbf{w}_{11}, \ldots, \mathbf{w}_{l4}$ are called pairwise orthogonal if*

$$E_{w_{ir}w_{is}} = \mathbb{O}_{ii} \quad \text{for} \quad r \neq s,$$

where $\mathbb{O}_{ii}$ is $p_i \times p_i$ zero matrix. The pairwise orthogonal random vectors $\mathbf{w}_{11}, \ldots, \mathbf{w}_{l4}$ are said to be pairwise orthonormal if it is also true that

$$E_{w_{is}w_{is}} = I \quad \text{for} \quad s = 1, \ldots, 4.$$

Lemma 1 [6, 7] *Let $\mathbf{v}_{ij} \in L^2(\Omega, \mathbb{R}^{q_{ij}})$ for each $i = 1, \ldots, l$ and $j = 1, \ldots, 4$, and let*

$$\mathbf{w}_{i1} = \mathbf{v}_{i1} \quad \text{and} \quad \mathbf{w}_{is} = \mathbf{v}_{is} - \sum_{\ell=1}^{s-1} \mathscr{Z}_{is\ell}(\mathbf{w}_{i\ell}) \quad \text{for } s = 2,3,4 \tag{19}$$

where $\mathscr{Z}_{is\ell} : L^2(\Omega, \mathbb{R}^{q_{i\ell}}) \to L^2(\Omega, \mathbb{R}^{q_{is}})$ is defined in the manner of (3) by the matrix $Z_{is\ell} \in \mathbb{R}^{q_{is}\times q_{i\ell}}$ given by

$$Z_{is\ell} = E_{w_{is}w_{i\ell}}E^\dagger_{w_{i\ell}w_{i\ell}} + M_{is\ell}(I - E_{w_{i\ell}w_{i\ell}}E^\dagger_{w_{i\ell}w_{i\ell}}) \tag{20}$$

where $M_{k\ell} \in \mathbb{R}^{q_{is}\times q_{i\ell}}$ is arbitrary. Then $\mathbf{w}_{i1}, \ldots, \mathbf{w}_{i4}$ are pairwise orthogonal random vectors.

In (16), the terms $J_1(L_{13})$ and $J_i(R_{i2}, L_{i3})$ is defined by the operators $\mathscr{L}_{13}$, $\mathscr{R}_{i2}$ and $\mathscr{L}_{i3}$ and their action on the random block-vectors $\mathbf{v}_{13}$, $\mathbf{v}_{i2}$ and $\mathbf{v}_{i3}$ respectively. The corresponding mutually orthogonal random vectors are

$$\mathbf{w}_{13} = \mathbf{v}_{13}, \quad \mathbf{w}_{i2} = \mathbf{v}_{i2} \quad \text{and} \quad \mathbf{w}_{i3} = \mathbf{v}_{i3} - \mathscr{Z}_i(\mathbf{v}_{i2}) \tag{21}$$

where $i = 2, \ldots, l$ and the operator $\mathscr{Z}_i : L^2(\Omega, \mathbb{R}^{q_{i2}}) \to L^2(\Omega, \mathbb{R}^{q_{i3}})$ is defined by the matrix

$$Z_i = E_{v_{i3}v_{i2}}E^\dagger_{v_{i2}v_{i2}} + M_i(I - E_{v_{i2}v_{i2}}E^\dagger_{v_{i2}v_{i2}}) \tag{22}$$

with $M_i \in \mathbb{R}^{q_{i3}\times q_{i2}}$ arbitrary.

We write

$$\mathbf{w}(\omega) = [\mathbf{w}_{13}(\omega)^T, \quad \mathbf{w}_{22}(\omega)^T, \quad \mathbf{w}_{23}(\omega)^T,$$
$$\ldots, \mathbf{w}_{l2}(\omega)^T, \quad \mathbf{w}_{l3}(\omega)^T]^T,$$

and

$$Z = \begin{bmatrix} I_{13} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & I_{22} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & -Z_2 & I_{23} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & \mathbb{O} & \mathbb{O} & I_{32} & \mathbb{O} & \mathbb{O} & \dots & \mathbb{O} \\ \mathbb{O} & \mathbb{O} & \mathbb{O} & -Z_3 & I_{33} & \mathbb{O} & \dots & \mathbb{O} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & & \vdots \\ \mathbb{O} & \dots & \dots & \dots & \mathbb{O} & \mathbb{O} & I_{l2} & \mathbb{O} \\ \mathbb{O} & \dots & \dots & \dots & \mathbb{O} & \mathbb{O} & -Z_l & I_{l3} \end{bmatrix}$$

where I_{ij} is $q_{ij} \times q_{ij}$ identity matrix for $i = 1,\dots,l$ and $j = 2,3$, and Z_i is defined by (22) for $i = 2,\dots,l$.

The matrix Z consists of $(2l-1)\times(2l-1)$ blocks.

Then (21) can be written in the matrix form as

$$\mathbf{w}(\omega) = Z\mathbf{v}(\omega)$$

with $\mathbf{v}$ given above. Matrix Z implies the operator $\mathscr{Z} : L^2(\Omega,\mathbb{R}^n) \to L^2(\Omega,\mathbb{R})$ defined in the manner of (3).

Since Z is invertible, we can represent $\mathscr{A}$ as follows:

$$\mathscr{A}(\mathbf{y}) = \mathscr{K}[\mathscr{Z}(\mathscr{P}(\mathbf{y}))] \quad \text{where} \quad \mathscr{K} = \mathscr{B}\mathscr{Z}^{-1}. \tag{23}$$

A matrix representation of $\mathscr{K}$ is

$$K = \begin{bmatrix} L_{13} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \dots & \mathbb{O} & \mathbb{O} \\ \mathbb{O} & T_2 & L_{23} & \mathbb{O} & \mathbb{O} & \mathbb{O} & \dots & \mathbb{O} & \mathbb{O} \\ \mathbb{O} & \mathbb{O} & \mathbb{O} & T_3 & L_{33} & \mathbb{O} & \dots & \mathbb{O} & \mathbb{O} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & & \vdots & \vdots \\ \mathbb{O} & \dots & \dots & \dots & \mathbb{O} & T_{l-1} & L_{l-1,3} & \mathbb{O} & \mathbb{O} \\ \mathbb{O} & \dots & \dots & \dots & \mathbb{O} & \mathbb{O} & \mathbb{O} & T_l & L_{l3} \end{bmatrix}$$

where

$$T_i = R_{i2} + L_{i3}Z_i \tag{24}$$

for $i = 2,\dots,l$. We note that K consists of $l\times(2l-1)$ blocks.

As a result, in (17),

$$\begin{aligned} R_{i2}\mathbf{v}_{i2}(\omega) + L_{i3}\mathbf{v}_{i3}(\omega) &= R_{i2}\mathbf{w}_{i2}(\omega) + L_{i3}[\mathbf{w}_{i3}(\omega) + Z_i\mathbf{w}_{i2}(\omega)] \\ &= T_i\mathbf{w}_{i2}(\omega) + L_{i3}\mathbf{w}_{i3}(\omega) \end{aligned}$$

and hence

$$J(A) = J_1(L_{13}) + \sum_{i=2}^{l} \mathscr{J}_i(T_i, L_{i3}), \tag{25}$$

where

$$\mathscr{J}_i(T_i, L_{i3}) = E[\|\mathbf{u}_i - [\mathscr{T}_i\mathbf{w}_{i2}(\omega) + \mathscr{L}_{i3}\mathbf{w}_{i3}]\|^2]$$

with $\mathscr{T}_i$ defined by

$$[\mathscr{T}_i\mathbf{w}_{i2}](\omega) = T_i[\mathbf{w}_{i2}(\omega)]$$

for all $i = 2,\dots,l$.

5. MAIN RESULTS

Lemma 2 *For $A \in \mathbb{L}^n_m$, the following is true:*

$$\begin{aligned} \min_{A\in\mathbb{L}^n_m} J(A) &= \min_{L_{13}} J_1(L_{13}) + \sum_{i=2}^{l} \min_{T_i, L_{i3}} \mathscr{J}_i(T_i, L_{i3}) \quad (26) \\ &= \min_{L_{13}} J_1(L_{13}) + \sum_{i=2}^{l} \min_{R_{i2}, L_{i3}} J_i(R_{i2}, L_{i3}). \quad (27) \end{aligned}$$

Now, we are in the position to prove the main result given in Theorem 1 below. To this end, we use the following notation.

For $i = 1,\dots,l$, let λ_i be the rank of the matrix $E_{w_{i3}w_{i3}} \in \mathbb{R}^{p_i\times p_i}$ and let[1]

$$E^{1/2}_{w_{i3}w_{i3}} = Q_iU_i$$

be the QR-decomposition for $E^{1/2}_{w_{i3}w_{i3}}$ where $Q_i \in \mathbb{R}^{p_i\times\lambda_i}$ and $Q_i^TQ_i = I$ and $U_i \in \mathbb{R}^{\lambda_i\times p_i}$ is upper trapezoidal with rank λ_i. We write $G_i = U_i^T$ and use the notation

$$G_i = [g_{i1},\dots,g_{i\lambda_i}] \in \mathbb{R}^{p_i\times\lambda_i}$$

where $g_{ij} \in \mathbb{R}^{p_i}$ denotes the j-th column of G_i. We also write

$$G_{i,s} = [g_{i1},\dots,g_{is}] \in \mathbb{R}^{p_i\times s}$$

for $s \le \lambda_i$ to denote the matrix consisting of the first s columns of the matrix G_i.

The s-th row of the unit matrix $I \in \mathbb{R}^{p_i\times p_i}$ is denoted by $e_s^T \in \mathbb{R}^{1\times p_i}$.

For a square matrix $M = \{m_{ij}\}^n_{i,j=1}$, we also write

$$M = M_\triangledown + M_\triangle$$

where

$$M_\triangledown = \{m_{ij} \mid m_{ij} = 0 \quad \text{if} \quad i < j\}$$

and

$$M_\triangle = \{m_{ij} \mid m_{ij} = 0 \quad \text{if} \quad i \ge j\},$$

i.e. $M_\triangledown$ is lower triangular and $M_\triangle$ is strictly upper triangular.

Theorem 1 *The solution to the problem (11) is given by the operator $\mathscr{A}^0 \in \mathbb{A}^n_m$ defined by the lower stepped matrix $A^0 \in \mathbb{L}^n_m$ where*

$$L^0_{i3} = \begin{bmatrix} \ell^0_{i,1} \\ \vdots \\ \ell^0_{i,p_i} \end{bmatrix} \quad \text{and} \quad R^0_{i2} = T^0_{i2} - L^0_{i3}Z_i \quad \text{for} \quad i = 1,\dots,l. \tag{28}$$

In (28), for each $i = 1,2,\dots,l$ and $s = 1,2,\dots,p_i$, the s-th row $\ell^0_{i,s}$ is defined by

$$\ell^0_{i,s} = e_s^T E_{u_iw_{i3}} E^\dagger_{w_{i3}w_{i3}} G_{i,s}G^\dagger_{i,s} + b_i^T(I - G_{i,s}G^\dagger_{i,s}) \tag{29}$$

where $b_i^T \in \mathbb{R}^{1\times p_i}$ is arbitrary; the matrix T^0_{i2} is such that

$$T^0_{i2} = E_{u_iw_{i2}}E^\dagger_{w_{i2}w_{i2}} + F_i(I - E_{w_{i2}w_{i2}}E^\dagger_{w_{i2}w_{i2}}) \tag{30}$$

with $F_i \in \mathbb{R}^{p_i\times q_{i2}}$ arbitrary and I the $q_{i2}\times q_{i2}$ identity matrix.

The error associated with the operator $\mathscr{A}^0$ is given by

$$\begin{aligned} E[\|\mathbf{x} - \mathscr{A}^0(\mathbf{y})\|^2] = \sum_{i=1}^{l} \Bigg[&\sum_{s=1}^{\lambda_i}\sum_{j=s+1}^{p_i} E\left[|e_s^T E_{u_iw_{i3}} E^\dagger_{w_{i3}w_{ki3}} g_{i,j}|^2\right] \\ &+ \|E^{1/2}_{u_iu_i}\|^2_F - \|E_{u_iw_{i2}}E^{\dagger 1/2}_{w_{i2}w_{i2}}\|^2 - \|E_{u_iw_{i3}}E^{\dagger 1/2}_{w_{i3}w_{i3}}\|^2_F \Bigg]. \end{aligned} \tag{31}$$

[1] We recall that by (6), $q_{i3} = p_i$.

Remark 1 *The matrix $G_i \in \mathbb{R}^{p_i \times r}$ has rank λ_i and hence has λ_i independent columns. It follows that $G_{i,s} \in \mathbb{R}^{p_i \times s}$ also has independent columns and therefore also has rank s. Thus $G_{i,s}^T G_{i,s} \in \mathbb{R}^{\lambda_i \times \lambda_i}$ is non-singular and so $G_{i,s}^\dagger = (G_{i,s}^T G_{i,s})^{-1} G_{i,s}^T$. Hence*

$$\begin{aligned} \ell_{i,s}^0 &= e_s^T E_{u_i w_{i3}} E_{w_{i3} w_{i3}}^\dagger G_{i,s} (G_{i,s}^T G_{i,s})^{-1} G_{i,s}^T \\ &+ b_i^T [I - G_{i,s} (G_{i,s}^T G_{i,s})^{-1} G_{i,s}^T] \end{aligned}$$

for all $i = 1, 2, \ldots, l$.

We note that the results by Bode and Shannon [3], Fomin and Ruzhansky [4], Ruzhansky and Fomin [5], and Wiener [1, 2, 6] are particular cases of Theorem 1 above.

5.1 Simulations

To illustrate the proposed method, we consider the best approximator $\mathscr{A}^0 \in \mathbb{A}_m^n$ with $n = 51$ and memory $m = \{m_1, \ldots, m_5\}$, where $m_1 = 20$, $m_2 = 25$, $m_3 = 15$, $m_4 = 35$ and $m_5 = 25$.

Then the blocks of the matrix A^0 are

$$L_{13}^0 \in \mathbb{R}^{20 \times 20}, \quad R_{22}^0 \in \mathbb{R}^{10 \times 15}, \quad L_{23}^0 \in \mathbb{R}^{10 \times 10}, \tag{32}$$

$$R_{32}^0 \in \mathbb{R}^{5 \times 10}, \quad L_{33}^0 \in \mathbb{R}^{5 \times 5}, \quad R_{42}^0 \in \mathbb{R}^{10 \times 25}, \quad L_{43}^0 \in \mathbb{R}^{10 \times 10}. \tag{33}$$

$$R_{52}^0 \in \mathbb{R}^{5 \times 20} \quad \text{and} \quad L_{53}^0 \in \mathbb{R}^{5 \times 5}. \tag{34}$$

We apply $\mathscr{A}^0 \in \mathbb{A}_m^{51}$ to the random vector $\mathbf{y}$ under conditions as follows. In accordance with the assumption made above, we suppose that a reference random vector $\mathbf{x} \in L^2(\Omega, \mathbb{R}^{51})$ is unknown and that noisy observed data $\mathbf{y} \in L^2(\Omega, \mathbb{R}^{51})$ is given by q realizations of $\mathbf{y}$ in the form of a matrix $Y \in \mathbb{R}^{n \times q}$ with $q = 101$. Matrices $E_{u_1 v_{13}}$, $E_{v_{13} v_{13}}$ and matrices $E_{u_i v_{i2}}$, $E_{u_i v_{i3}}$, $E_{v_{i2} v_{i2}}$ and $E_{v_{i3} v_{i3}}$ for $i = 2, \ldots, 5$, or their estimates are assumed to be known.

In practice, these matrices or their estimates are given numerically, not analytically. Similarly to our methods presented in [6, 7], the proposed method works, of course, under this condition. In this example, we model the matrices used in the simulations with analytical expressions in the following way. First, we set $X \in \mathbb{R}^{n \times q}$ and $Y \in \mathbb{R}^{n \times q}$ by

$$X = [\cos(\alpha) + \cos(0.3\alpha)]^T [\cos(0.5\beta) + \sin(5\beta)]$$

and

$$Y = [\cos(\alpha) \bullet r_1 + \cos(0.3\alpha)]^T [\cos(0.5\beta) + \sin(5\beta) \bullet r_2],$$

where

$$\alpha = [\alpha_0, \alpha_1, \ldots, \alpha_{n-1}], \quad \alpha_{k+1} = \alpha_k + 0.4, \quad k = 0, 1, \ldots, n-1,$$

$$\alpha_0 = 0, \quad \beta_0 = 0,$$

$$\beta = [\beta_0, \beta_1, \ldots, \beta_{q-1}], \quad \beta_{j+1} = \beta_j + 0.4, \quad j = 0, 1, \ldots, q-1,$$

$$\cos(\alpha) = [\cos(\alpha_0), \ldots, \cos(\alpha_n)],$$

$$\sin(\beta) = [\sin(\beta_0), \ldots, \sin(\beta_{q-1})],$$

the symbol $\bullet$ means the Hadamard product, r_1 is a $1 \times n$ normally distributed random vector and r_2 is a $1 \times q$ uniformly distributed random vector. Here, r_1 and r_2 simulate noise.[2]

[2] The matrix X can be interpreted as a sample of $\mathbf{x}$. By the assumptions of the proposed method, it is not necessary to know X. We use matrix X for illustration purposes only.

Each column of Y is a particular realization of $\mathbf{y}$.

By the proposed procedure , we partition each column of X and Y in subvectors

$$u_1, \ldots, u_5 \quad \text{and} \quad v_{13},\, v_{22},\, v_{23},\, \ldots,\, v_{52},\, v_{53},$$

respectively.

Furthermore, v_{13}, v_{22}, v_{23}, v_{32}, v_{33} and v_{34} have been orthogonalized to w_{11}, w_{22}, w_{23}, w_{32}, w_{33} and w_{34}. Matrices (32)–(34) have then been evaluated by the procedure presented in Theorem 1 from $u_1, \ldots, u_3$, and w_{11}, w_{22}, w_{23}, w_{32}, w_{33} and w_{34}.

As a result, the estimate $\hat{\mathbf{x}}^0$ has been evaluated in the form $\hat{x}^0$ such that

$$\hat{x}^0 = \begin{bmatrix} L_{13}^0 w_{13} \\ R_{22}^0 w_{22} + L_{23}^0 w_{23} \\ \vdots \\ R_{52}^0 w_{52} + L_{53}^0 w_{53} \end{bmatrix}.$$

On Fig. 1, the plots of columns 51 and 52 of the matrix Y are presented. They are typical representatives of the noisy data under consideration. On Fig. 2, the plots of columns 51 and 52 of the matrix X (solid line) and their estimates (dashed line with circles) by our filter are given.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice–Hall, Englewood Cliffs, N. J., 1991.

[2] V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing,* J. Wiley & Sons, 2001.

[3] H. W. Bode and C. E. Shannon, A Simplified Derivation of Linear Least Square Smoothing and Prediction Theory, *Proc. IRE,* 38, pp. 417–425, 1950.

[4] V. N. Fomin and M. V. Ruzhansky, Abstract optimal linear filtering, *SIAM J. Control Optim.,* 38, pp. 1334–1352, 2000.

[5] M. Ruzhanski and V. Fomin, Optimal Filter Construction for a General Quadratic Cost Functional, *Bulletin of St. Petersburg University. Mathematics,* 28, pp. 50–55, 1995.

[6] A. Torokhti, P. Howlett, *Computing Methods for Modelling of Nonlinear Systems,* Elsevier, 302 p. (in press).

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