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

OPTIMAL ESTIMATION OF A RANDOM SIGNAL FROM PARTIALLY MISSED DATA

Anatoli Torokhti, Phil Howlett and Charles Pearce

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

We provide a new technique for random signal estimation under the constraints that the data is corrupted by random noise and moreover, some data may be missed. We utilize nonlinear filters defined by multi-linear operators of degree r, the choice of which allows a trade–off between the accuracy of the optimal filter and the complexity of the corresponding calculations. A rigorous error analysis is presented.

1. INTRODUCTION

The tasks considered in this paper are motivated by restrictions arising in real applications. They mainly concern with a practical way for gathering data. In many practical cases, to estimate a component $\mathbf{x}_k$ of the reference signal $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_m)^T$, a filter $\mathscr{A}$ uses (or "remembers") no more than the $p_k = 1, \ldots, v_k$ most recent components $\mathbf{y}_{s_k}, \ldots, \mathbf{y}_{v_k}$ from the measurement data $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_m)^T$, where s_k and v_k are respectively defined by

$$s_k = v_k - p_k + 1 \quad \text{and} \quad v_k = 1, \ldots, k. \tag{1}$$

We say that such a filter $\mathscr{A}$ has arbitrarily variable *incomplete* memory $p = \{p_1, \ldots, p_m\}$.

Such a restriction makes the problem of finding the best $\mathscr{A}$ quite specific. This is, perhaps, a reason that despite a long history of the subject [3], even for a simplest structure of the filter $\mathscr{A}$ when $\mathscr{A}$ is defined by a 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]) and for the case of *complete* memory only. We avoid such bottlenecks and solve the problem in the general case of the polynomial filter with the arbitrarily variable *incomplete* memory p. The proposed technique is substantially different from those considered, for the case of linear filters, in [3, 4, 5] and for the case of non-linear regression, in [6].

A distinguishing feature of our solution is that the filter should be non-linear and causal with incomplete memory. The simplest particular case of our desired filter is an optimal *linear* filter defined by a lower p-band matrix. There is no published solution even for this simplest case but Fomin and Ruzhansky [4, 5] have recently proposed an optimal *causal* linear filter where no restriction is imposed on the memory. In effect their filter has *unlimited* memory.

The *novelty* of our approach derives from our formulation of a new class of filters, the new reduction technique to determine an optimal representative and a rigorous error analysis. The main results are Proposition 1 and Theorems 1 and 2.

1.1 Computational impact

Many known results assume the relevant covariance matrices are invertible. Such an assumption is quite restrictive from a computational point of view and is often associated with numerical instability. Our formula for the optimal filters can always be computed. In addition the reduction procedure (Section 4) means that our optimal filters are defined by matrices of reduced size. Consequently the computational load should compare favorably with known methods. On the other hand we use non-linear filters to provide improved accuracy and it is natural to expect additional computation in problems where increased accuracy is desired.

2. NONLINEAR CAUSAL FILTERS WITH INCOMPLETE MEMORY

The vector $\mathbf{y}$ is interpreted as observable data containing information about the reference signal $\mathbf{x}$, but may be contaminated by noise. No specific relationship between the reference signal and noise is assumed. Using an heuristic idea of causality, we expect that the present value of the estimate is not affected by future values of the data. Since the filters under consideration must have incomplete memory $p = \{p_1, \ldots, p_m\}$ the estimate of x_k must be obtained from the data components $y_{s_k}, \ldots, y_{v_k}$. For each $k = 1, 2, \ldots, m$ and random m–vector $\mathbf{y}$, we define τ_k by

$$\tau_k(\mathbf{y}) = (y_{s_k}, y_{s_k+1}, \ldots, y_{v_k}).$$

Definition 1 *An m–vector operator $\mathscr{A}$ acting on $\mathbf{y}$ is called a* causal *filter with* incomplete memory $p = \{p_1, \ldots, p_m\}$ *if $\mathscr{A}(\mathbf{y})$ takes the form*

$$\mathscr{A}(\mathbf{y}) = \begin{bmatrix} \mathscr{A}_1(\tau_1(\mathbf{y})) \\ \vdots \\ \mathscr{A}_m(\tau_m(\mathbf{y})) \end{bmatrix}. \tag{2}$$

For a brevity, we often omit the term "causal" for such a filter.

For an appropriate choices of $\mathscr{A}$, the vector $\mathscr{A}(\mathbf{y})$ provides an estimate of $\mathbf{x}$.

A simple example in which $\mathscr{A}(\mathbf{y})$ is a first–order polynomial is

$$\mathscr{A}(\mathbf{y}) = \mathbf{a} + B_1\mathbf{y}$$

with $\mathbf{a}$ an m–vector and B_1 an $m \times m$ matrix. By the definition, the filter $\mathscr{A}$ is causal with incomplete memory $p = \{p_1, \ldots, p_m\}$ if the matrix $B_1 = \{b_{kj}\}$ is such that

$$b_{kj} = 0 \quad \text{for } j = \{1, \ldots, s_k - 1\} \bigcup \{v_k + 1, \ldots, m\}$$

We call B_1 a *lower p-band matrix*. The set of lower p-band matrices in $\mathbb{R}^{m\times m}$ is denoted by $\mathscr{R}^{m\times m}_{p_1,\ldots,p_m}$.

For instance, if

$$m=4,\quad p_1=1,\quad p_2=2,\quad p_3=2,\quad p_4=3,$$
$$v_1=1,\quad v_2=2,\quad v_3=3,\quad v_4=3,$$

then $B_1 \in \mathscr{R}^{4\times 4}_{1223}$ is given by

$$B_1 = \begin{bmatrix} \bullet & 0 & 0 & 0 \\ \bullet & \bullet & 0 & 0 \\ 0 & \bullet & \bullet & 0 \\ \bullet & \bullet & \bullet & 0 \end{bmatrix}$$

where $\bullet$ denotes an entry which is not necessarily zero.

More generally, we employ as in [7] an operator $\mathscr{P}_r$ of degree $r\in\mathbb{N}$ given by

$$\mathscr{P}_r(\mathbf{y}) = \mathbf{a} + \sum_{q=1}^{r} \mathscr{B}_q(\mathbf{y}^q).$$

Here $\mathbf{y}^q = (\mathbf{y},\ldots,\mathbf{y})$ (q terms), $\mathbf{a}$ is an m–vector and $\mathscr{B}_q$ is q–linear, that is, $\mathscr{B}_q(\mathbf{z}_1,...,\mathbf{z}_q)$ is linear in each of $\mathbf{z}_1$, ... , $\mathbf{z}_q$. We wish to employ $\mathscr{P}_r\mathbf{y}$ as a filter for $\mathbf{x}$. Motivation for such usage follows from known results [7, 9] where strong estimating properties of $\mathscr{P}_r$ have been demonstrated.

For $j=1,\ldots,m$, denote by $\mathbf{e}_j$ the m–vector with unity in place j and zeros elsewhere. Then by q–linearity

$$\begin{aligned}\mathscr{B}_q[\mathbf{y}^q] &= \mathscr{B}_q\left[\sum_{j_1=1}^{m} y_{j_1}\mathbf{e}_{j_1},\ldots,\sum_{j_q=1}^{m} y_{j_q}\mathbf{e}_{j_q}\right] \\ &= \sum_{j_1=1}^{m}\cdots\sum_{j_q=1}^{m} y_{j_1}\cdots y_{j_q}\mathscr{B}_q[\mathbf{e}_{j_1},\ldots,\mathbf{e}_{j_q}] \\ &= \sum_{j_1=1}^{m}\cdots\sum_{j_q=1}^{m} y_{j_1}\cdots y_{j_q} B_{q,j_1,\cdots,j_q},\end{aligned}$$

say, for $q\geq 1$. Here each $B_{q,j_1\cdots j_q}$ is an m–vector. Thus

$$\mathscr{P}_r(\mathbf{y}) = \mathbf{a} + \sum_{q=1}^{r}\sum_{j_1=1}^{m}\cdots\sum_{j_q=1}^{m} y_{j_1}\cdots y_{j_q} B_{q,j_1,\cdots,j_q}. \quad (3)$$

We denote by $\mathscr{P}_{r,k}(\mathbf{y})$ the k-th element of $\mathscr{P}_r(\mathbf{y})$ ($k=1,2,\ldots,m$), with corresponding definitions for $B_{q,j_1,\cdots,j_q,(k)}$ and a_k. Then

$$\mathscr{P}_{r,k}(\mathbf{y}) = a_k + \sum_{q=1}^{r}\sum_{j_1=1}^{m}\cdots\sum_{j_q=1}^{m} y_{j_1}\ldots y_{j_q} B_{q,j_1,\cdots,j_q,(k)}. \quad (4)$$

The products $y_{j_1}\ldots y_{j_q}$ in (3) and (4) are not all distinct. Suppose we collect together terms with the common factors $y_1^{i_1}\cdots y_m^{i_m}$ into a class $\mathscr{S}(\mathbf{i})$ for each nonzero vector $\mathbf{i}=(i_1,\ldots,i_m)$ of nonnegative integers. We may then express (4) as

$$\mathscr{P}_{r,k}(\mathbf{y}) = a_k + \sum_{\mathbf{i}\,\in\,\mathscr{S}_{m,r}} y_1^{i_1}\cdots y_m^{i_m} \quad B_{\mathbf{i},(k)}, \quad (5)$$

where

$$\mathscr{S}_{m,r} = \bigcup_{i_1+\ldots+i_m\leq r} \mathscr{S}(\mathbf{i})$$

and

$$B_{\mathbf{i}} = \sum_{\mathbf{i}}{}' B_{q,j_1,\cdots,j_q}.$$

Here the summation $\sum_{\mathbf{i}}{}'$ is over all $(j_1,\ldots,j_q)$ for which

$$i_n = \sum_{\ell=1}^{q}\delta(n,j_\ell) \quad (1\leq n\leq m),$$

where δ is the Kronecker delta.

In fact with a filter of incomplete memory $p=\{p_1,\ldots,p_m\}$, the number of terms is further reduced. Let us temporary denote $s=s_k$ and $v=v_k$. We set

$$\mathscr{T}_r(\mathbf{y}) = \begin{bmatrix}\mathscr{T}_{r,1}(\tau_1(\mathbf{y})) \\ \mathscr{T}_{r,2}(\tau_2(\mathbf{y})) \\ \vdots \\ \mathscr{T}_{r,m}(\tau_m(\mathbf{y}))\end{bmatrix} = \begin{bmatrix}\mathscr{T}_{r,1}(y_{v_1}) \\ \mathscr{T}_{r,2}(y_{s_2},y_{v_2}) \\ \vdots \\ \mathscr{T}_{r,m}(y_{s_m},\ldots,y_{v_m})\end{bmatrix}, \quad (6)$$

where for each $k=1,\ldots,m$ we have

$$\mathscr{T}_{r,k}(\tau_k(\mathbf{y})) = a_k + \sum_{\mathbf{i}^k\,\in\,\mathscr{S}_{m,r}} y_s^{i_s}\ldots y_v^{i_v} \quad \beta_{\mathbf{i}^k}(\tau_k(\mathbf{y})), \quad (7)$$

where $\mathbf{i}^k=(0,\ldots,i_s,\ldots,i_k,0,\ldots,0)$ and $\beta_{\mathbf{i}^k}$ is an appropriate restriction of $\mathscr{B}_{\mathbf{i},(k)}$. Thus $\mathscr{T}_{r,k}$ is constructed from $\mathscr{P}_{r,k}$ when the general terms $y_1^{i_1}\cdots y_m^{i_m}\ \mathscr{B}_{\mathbf{i},(k)}(\mathbf{y})$ in (5) are restricted to terms of the form $y_s^{i_s}\ldots y_k^{i_k}\ \beta_{\mathbf{i}^k}(\tau_k(\mathbf{y}))$. Note that it is possible to have different degree operators for each component. In such cases we simply replace r by r_k in (7).

Although $\mathscr{T}_{r,k}(y_s,\ldots,y_v)$ is multilinear in the original variables $y_1,\ldots,y_m$, the dependence on the product terms $y_s^{i_s}\ldots y_v^{i_v}$ is linear (ones again, we write $s=s_k$ and $v=v_k$ here). We denote by $N_k=N_k(r)$ the number of such terms, the terms by $h_{jk}=h_{jk}(r)$ and the corresponding coefficients by $\eta_{jk}=\eta_{jk}(r)$ ($j=1,2,\ldots,N_k$) in any convenient ordering. Thus we write

$$\mathscr{T}_{r,k}(y_{s_k},\ldots,y_{v_k}) = a_k + \sum_{j=1}^{N_k}\eta_{jk}h_{jk} = a_k + {\eta_k}^T\mathbf{h}_k, \quad (8)$$

where ${\eta_k}^T=(\eta_{1k},\ldots,\eta_{N_k k})$ and ${\mathbf{h}_k}^T=(h_{1k},\ldots,h_{N_k k})$ for $k=1,2,\ldots,m$.

3. FORMULATION OF THE PROBLEM

Let

$$J(\mathscr{T}_r) = E[\|\mathbf{x}-\mathscr{T}_r(\mathbf{y})\|^2],$$

where $\mathscr{T}_r$ is defined by (6) and (8). The problem is to find $\mathscr{T}_r^0$ such that

$$J(\mathscr{T}_r^0) = \min_{\mathscr{T}_r} J(\mathscr{T}_r). \quad (9)$$

An optimal filter $\mathscr{T}_r^0$ in the class of causal r-th degree filters with incomplete memory p takes the general form

$$\mathscr{T}_r^0(\mathbf{y}) = \begin{bmatrix} \mathscr{T}_{r,1}^0(y_{s_k},\ldots,y_{v_k}) \\ \vdots \\ \mathscr{T}_{r,m}^0(y_{s_k},\ldots,y_{v_k}) \end{bmatrix}, \quad (10)$$

where the component $\mathscr{T}_{r,k}^0$ is given by

$$\mathscr{T}_{r,k}^0(y_{s_k},\ldots,y_{v_k}) = a_k^0 + \eta_k^{0\,T}\mathbf{h}_k \quad (11)$$

for each $k = 1,\ldots,m$. Finding an optimal representative $\mathscr{T}_r^0$ is therefore a matter of finding optimal values a_k^0 and $\eta_k^{0\,T}$ for the constants a_k and vectors η_k^T.

4. REDUCTION OF ORIGINAL PROBLEM TO SEVERAL INDEPENDENT PROBLEMS

The special structure of the operators makes direct solution of (9) a difficult problem. Suffice it to say that a solution is known only for the special case where $\mathscr{T}_r$ is linear and has complete memory [4, 5]. Moreover the solution in [4, 5] has been obtained with a quite restrictive assumption that the covariance matrix $E[\mathbf{y}\mathbf{y}^T]$ is nonsingular. Indeed we observe that direct determination of $\mathscr{T}_r^0$ from (9) is not straightforward because of difficulties imposed by the embedded lower p-band structure of the matrices. To avoid these difficulties we show that the problem (9) can be reduced to m independent problems. Define

$$J_k(\mathscr{T}_{r,k}) = E[|x_k - \mathscr{T}_{r,k}(y_{s_k},\ldots,y_{v_k})|^2] \quad (12)$$

for $k = 1,\ldots,m$, where $\mathscr{T}_{r,k}$ is defined by (8). We have the following result.

Proposition 1 *We have*

$$\min_{\mathscr{T}_r} J(\mathscr{T}_r) = \sum_{k=1}^{m} \min_{\mathscr{T}_{r,k}} J_k(\mathscr{T}_{r,k}). \quad (13)$$

Expression (13) allows us to reformulate (9) in an equivalent form: for each $k = 1,\ldots,m$, find $\mathscr{T}_{r,k}^0$ such that

$$J_k(\mathscr{T}_{r,k}^0) = \min_{\mathscr{T}_{r,k}} J_k(\mathscr{T}_{r,k}). \quad (14)$$

Any optimal operator $\mathscr{T}_{r,k}^0$ is the k-th component of an optimal estimator $\mathscr{T}_r^0$. Hence an optimal estimator $\mathscr{T}_r^0$ can be constructed from any solutions $\mathscr{T}_{r,1}^0,\ldots,\mathscr{T}_{r,m}^0$ to the m independent problems (14).

5. DETERMINATION OF THE OPTIMAL CAUSAL FILTER WITH INCOMPLETE MEMORY

To describe the solution to (9) we introduce some additional notation. If A is a nonnegative square matrix then we write $A^{1/2}$ for a nonnegative square matrix satisfying $A^{1/2}A^{1/2} = A$. We also denote by $A^\dagger$ the generalised inverse matrix for A. Both matrices are normally defined via the singular–value decomposition. For any two random vectors $\mathbf{w}$, $\mathbf{q}$ we write $\hat{\mathbf{w}} = \mathbf{w} - E[\mathbf{w}]$ and $\hat{\mathbf{q}} = \mathbf{q} - E[\mathbf{q}]$ and define

$$\mathbb{E}_{wq} = E[\hat{\mathbf{w}}\hat{\mathbf{q}}^T] = E[\mathbf{w}\mathbf{q}^T] - E[\mathbf{w}]E[\mathbf{q}^T].$$

Then we have

$$\mathbb{E}_{wq}\mathbb{E}_{qq}^\dagger\mathbb{E}_{qq} = \mathbb{E}_{wq} \quad (15)$$

(see [7]).

It is convenient to use a special notation in two particular cases. We introduce $\mathbb{H}_k := \mathbb{H}_k(r)$ and $\mathbb{Q}_k := \mathbb{Q}_k(r)$ by the formulæ

$$\mathbb{H}_k = E[\mathbf{h}_k\mathbf{h}_k^T] - E[\mathbf{h}_k]E[\mathbf{h}_k^T] \text{ and } \mathbb{Q}_k = E[x_k\mathbf{h}_k^T] - E[x_k]E[\mathbf{h}_k^T]. \quad (16)$$

We have from (15) that

$$\mathbb{Q}_k\mathbb{H}_k^\dagger\mathbb{H}_k = \mathbb{Q}_k. \quad (17)$$

Theorem 1 below solves problem (14) for $k = 1,2,\ldots,m$.

Theorem 1 *For each $k = 1,2,\ldots,m$, an optimal causal r-th degree filter $\mathscr{T}_{r,k}^0(y_{s_k},\ldots,y_{v_k})$ with incomplete memory p is defined by*

$$\eta_k^{0\,T} = \mathbb{Q}_k\mathbb{H}_k^\dagger + M_k[I_k - \mathbb{H}_k\mathbb{H}_k^\dagger], \quad (18)$$

where the N_k–vector M_k is arbitrary, I_k is the $N_k \times N_k$ identity matrix and

$$a_k^0 = E[x_k] - \eta_k^{0\,T}E[\mathbf{h}_k]. \quad (19)$$

Remark 1 *The solution structure given by Theorem 1 may formally be recognized in terms of the known statistical least–squares prediction problem (see, for example, [6]). By the known technique, we seek a least–squares prediction of a random variable X in the form*

$$\hat{X} = a_0 + \sum_{i=1}^{n} a_i W_i.$$

Let

$$\gamma = \begin{bmatrix} cov\,(X,W_1) \\ \vdots \\ cov\,(X,W_n) \end{bmatrix}, \qquad \mu_W = \begin{bmatrix} E(W_1) \\ \vdots \\ E(W_n) \end{bmatrix},$$

Γ be the $n \times n$ matrix with (i,j)–th entry $cov\,(W_i,W_j)$, and suppose that these quantities, $E(X)$ and $var\,(X)$ are all known. If $a = (a_1,\ldots,a_n)^T$, the solution of this problem is known to be

$$a_0 = E(X) - a^T\mu_W,$$

where $\mathbf{a}$ is any solution to

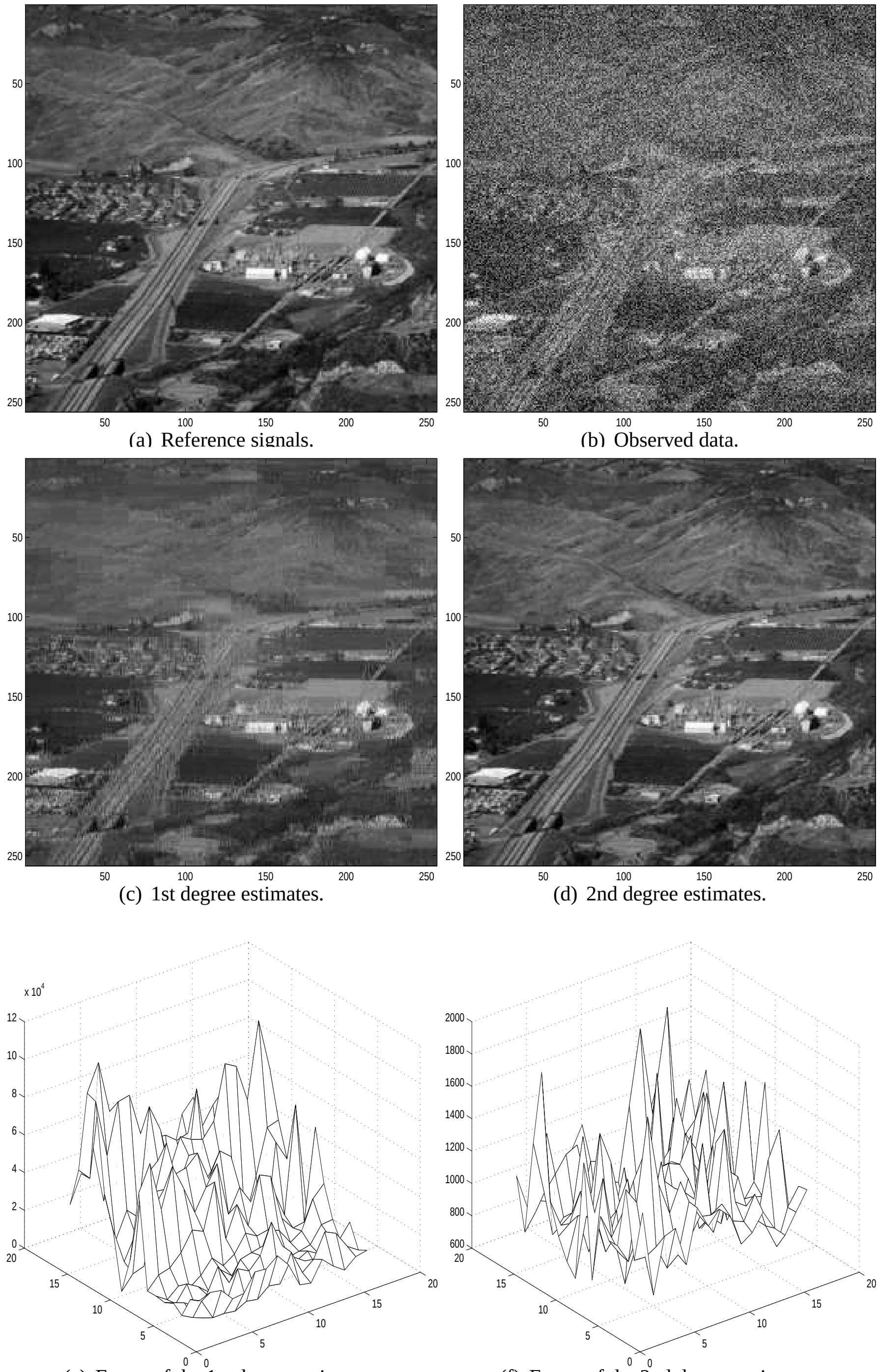
$$\Gamma a = \gamma.$$

Further, the associated mean–square error is

$$var\,(X) - a^T\gamma.$$

The differences of our approach (and as a result, of Theorem 1) from the known methods are implied by the specific structure of the filter used. In turn, such a specific structure is implied by the incomplete memory condition (Definition 1) and by the reduction to m independent problems (Proposition 1). Moreover, nonlinear terms involving the data values y_i are used for the estimation. The latter implies the significant improvement in the accuracy of estimation. In this regard, see also references [2, 7].

Remark 2 *The covariances used in (18) and (19) and in similar relationships below, are assumed to be known. This assumption is a well-known and significant limitation in problems dealing with estimation of random signals. The covariances can be estimated in various ways and particular techniques are given in many papers (e.g., see [10]).*



(a) Reference signals. (b) Observed data.

(c) 1st degree estimates. (d) 2nd degree estimates.

(e) Errors of the 1st degree estimates. (f) Errors of the 2nd degree estimates.

Figure 1: The performance of the proposed method.

Theorem 2 *The mean–square error* $E[\|\mathbf{x}-\mathscr{T}_r^0(\mathbf{y})\|^2]$ *for any optimal filter* $\mathscr{T}_r^0$ *defined by (11), (18) and (19) is*

$$E[\|\mathbf{x}-\mathscr{T}_r^0(\mathbf{y})\|^2] = \sum_{k=1}^{m} J_k(\mathscr{T}_{r,k}^0) = \sum_{k=1}^{m} \mathbb{E}_{x_k x_k} - \mathbb{Q}_k \mathbb{H}_k^{\dagger} \mathbb{Q}_k^T. \tag{20}$$

6. SIMULATIONS

To illustrate the performance of the method, we apply the proposed filters to the problem of extracting information about images of the surface of the earth obtained by air observations. The 256×256 reference matrix $\mathscr{X}$ to be estimated is a numerical representation of the image of a chemical plant. The data can be found in
http://sipi.usc.edu/services/database/Database.html .
We consider two different cases. In the first the data is disturbed by additive noise and in the second by multiplicative noise. In each case the raw data set is represented by a 256×256 matrix $\mathscr{Y}$. In the first case we set $\mathscr{Y}=\mathscr{Y}^{(1)}$ and in the second $\mathscr{Y}=\mathscr{Y}^{(2)}$ where

$$\mathscr{Y}^{(1)} = \mathscr{X} + 150\, R_1 \quad \text{and} \quad \mathscr{Y}^{(2)} = \mathscr{X} * R_2,$$

and where R_1 and R_2 are 256×256 matrices of randomly generated numbers from independent uniform distributions on the interval $(0,1)$. The symbol "$*$" denotes the Hadamard product of two square matrices of the same order, given by $(A*B)_{i,j} = A_{i,j}B_{i,j}$.

Because the procedure is formally the same in each case we will give a generic description with $\mathscr{X}$ denoting the reference matrix that we wish to estimate and $\mathscr{Y}$ denoting the observed data. In each case we begin the analysis by partitioning $\mathscr{X}$ and $\mathscr{Y}$ into smaller blocks and we consider two different schemes. In the first instance we use 64 separate blocks with sub-matrices

$$\{X_{i,j}\}_{i,j=1,\ldots,32} \quad \text{and} \quad \{Y_{i,j}\}_{i,j=1,\ldots,32}$$

and in the second a more refined partition

$$\{X_{i,j}\}_{i,j=1,\ldots,16} \quad \text{and} \quad \{Y_{i,j}\}_{i,j=1,\ldots,16}$$

with 256 separate blocks. Since the procedure is essentially the same whichever scheme is used our subsequent description will not distinguish between the two.

To apply the estimation procedure to each fixed block (i,j) we set $X = X_{i,j}$ and $Y = Y_{i,j}$. The ℓ-th columns $x^{(\ell)} = x_{i,j}^{(\ell)}$ and $y^{(\ell)} = y_{i,j}^{(\ell)}$ of X and Y respectively are regarded as the ℓ-th realisations of the random signals $\mathbf{x}$ and $\mathbf{y}$. To model the incomplete memory requirement, we assume that at each time k we can observe at most the seven most recent rows of data. Thus our estimate for the k-th element x_k can use only the observed elements $y_s,\ldots,y_k$ with $p=7$. We have applied standard MATLAB routines to compute these estimates using our proposed optimal causal filters of degrees one and two. For each $k=1,\ldots,m$ and each $r=1,2$ the optimal filters are denoted by $T_{r,k}^0$ and are given by (11),(18) and (19) with $M_k = 0$.

The covariances have been estimated from the samples as maximum likelihood estimates. We have used this method for illustrative purposes only. The results of the simulations are presented in Figure 1 and Table 1. For each case in Table 1 we write

$$\Delta_{r,ij} = \|X_{ij} - T_r^0(Y_{ij})\|^2$$

for $r=1,2$. The results are consistent with the theoretical analysis. Table 1 shows that the error associated with the second degree filter T_2^0 is less than that for the first degree filter T_1^0.

Table 1. Maximum errors for the proposed filters

	16×16 submatrices		32×32 submatrices	
	Errors by T_1^0 and T_2^0		Errors by T_1^0 and T_2^0	
Case	$\max_{ij}\Delta_{1,ij}$	$\max_{ij}\Delta_{2,ij}$	$\max_{ij}\Delta_{1,ij}$	$\max_{ij}\Delta_{2,ij}$
1	1.16×10^5	0.02×10^5	5.32×10^5	0.71×10^5
2	2.85×10^5	0.54×10^5	1.05×10^6	0.29×10^6

The proposed method has also been tested with other simulations including EEG data. Those tests were also consistent with the theoretical results obtained above.

It is inappropriate to compare causal filters with incomplete memory proposed above to filters that are not restricted in this way. One would naturally expect unrestricted filters to exhibit superior performance but there are many realistic applications where such filters cannot be used.

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] P.J. Brockwell, R.A. Davis, *Introduction to Time Series and Forecasting,* Springer, New York., 2002.

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

[8] 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.

[9] P. G. Howlett, A. P. Torokhti, & Pearce, C. E. M.: A Philosophy for the Modelling of Realistic Non-linear Systems, *Proc. of Amer. Math. Soc.,* **131**, 2, pp. 353–363, 2003.

[10] L. I. Perlovsky and T. L. Marzetta, Estimating a Covariance Matrix from Incomplete Realizations of a Random Vector, *IEEE Trans. on Signal Processing,* 40, pp. 2097-2100, 1992.