

Article

Not peer-reviewed version

On Constructive Approximation of Nonlinear Operators

[Anatoli Torokhti](#) * and [Peter Pudney](#)

Posted Date: 10 December 2025

doi: 10.20944/preprints202512.0842.v1

Keywords: approximation of nonlinear mappings; error minimization; optimization; interpolation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

On Constructive Approximation of Nonlinear Operators

Anatoli Torokhti *  and Peter Pudney

STEM discipline, University of South Australia, Mawson Lakes, SA, Australia

* Correspondence: anatoli.torokhti@gmail.com

Abstract

Suppose K_Y and K_X are the image and the preimage of a nonlinear operator $\mathcal{F} : K_Y \rightarrow K_X$. It is supposed that the cardinality of each K_Y and K_X is N and N is large. We provide an approximation to the map \mathcal{F} that requires a prior information only on a few elements p from K_Y , where $p \ll N$, but still effectively represents $\mathcal{F}(K_Y)$. It is achieved under quite non-restrictive assumptions. The device behind the proposed method is based on a special extension of the piecewise linear interpolation technique to the case of sets of stochastic elements. The proposed technique provides a single operator that transforms any element from the arbitrarily large set K_Y . The operator is determined in terms of pseudo-inverse matrices so that it always exists.

Keywords: approximation of nonlinear mappings; error minimization; optimization; interpolation

1. Introduction

A purpose of the proposed methodology is to provide an effective way to transform large data sets. The methodology is motivated by the problems arising in signal processing where the nonlinear operator $\mathcal{F} : K_Y \rightarrow K_X$ is interpreted as a nonlinear system (or a nonlinear filter) transforming a set of stochastic signals K_Y to the set of stochastic signals K_X . Therefore, below we refer to this terminology.

The device behind the proposed method is based on a special extension of the piecewise linear interpolation technique to the case of stochastic signal sets. The device is not straightforward and requires the careful substantiation presented in Sections 2.3, 3.4, 4.2 and 4.4 below.

1.1. Motivations

The problem under consideration is motivated by the following observations.

1.1.1. Transformation of Large Sets of Signals

Suppose we need to transform a set of signals K_Y to another set of signals K_X . The signals are represented by finite stochastic vectors¹. A major associated difficulty and inconvenience which is common to many known filtering methodologies (see, for example, [1]–[9], [11,13,22,23,25]) is that they require a prior information on each reference signal to be estimated². In particular, the filters in [22,23,25] are based on the use of either the reference signal $\mathbf{x} \in K_X$ itself, as in [22,23], or its estimate, as in [25]. The Wiener filtering approach (see, for example, [1]–[13,23,25]) assumes that covariance matrices formed from a reference signal, $\mathbf{x} \in K_X$, and an observed signal, $\mathbf{y} \in K_Y$, are known or can be estimated. The latter can be done, for instance, from samples of \mathbf{x} and \mathbf{y} . In particular, this means that the reference signal \mathbf{x} can be measured.

In the case of processing large signal sets, such restrictions become much more inconvenient.

¹ We say a stochastic vector \mathbf{x} is finite if its realization has a finite number of scalar components.

² To the best of our knowledge, the exception is the methodology in [10,12] where the filtering techniques exploit information on reference signals in the form of the vector obtained from averaging over reference signal sets.

The major motivating question for this work is as follows. Let $\mathcal{F} : K_Y \rightarrow K_X$ denote a filter that estimates a large set of reference signals, K_X , from a large set of observed signals, K_Y . Each set contains N signals. Is it possible to construct a filter \mathcal{F} that requires a prior information only on few signals, $p \ll N$, from K_X but performs better than the known filters based on a prior information on every reference signal from K_X ? We denote such a filter by $\mathcal{F}^{(p-1)}$.

It is shown in Sections 2.3 and 4.4 that the positive answer is achievable under quite unrestrictive assumptions. The required features of filter $\mathcal{F}^{(p-1)}$ are satisfied by its special structure described in Sections 2.3, 3.1 and 3.4. The related conditions are also considered in those Sections.

1.1.2. Filtering Based on Idea of Piecewise Function Interpolation

The specific structure of the proposed filter follows from the extension of piecewise function interpolation [14]. This is because the technique of piecewise function interpolation [14] has significant advantages over the methods of linear and polynomial approximation used in known filtering techniques (such as, for example, those in [5,9]).

The structure of the proposed filter is presented in Sections 2.3, 3.1 and 4.2 below.

1.1.3. Exploiting Pseudo-Inverse Matrices in the Filter Model

Most of the known filtering techniques, for example, those ones in [1]–[3], [6]–[8], [11,23,25], are based on exploiting inverse matrices in their mathematical models. In the cases of grossly corrupted signals or erroneous measurements those inverse matrices may not exist and, thus, those filters cannot be applied. The examples in Section 5 illustrate this case.

The filter proposed here avoids this drawback since its model is based on exploiting pseudo-inverse matrices. As a result, the proposed filter always exist. That is, it processes any kind of noisy signals. An extension of the filtering techniques to the case of implementation of the pseudo-inverse matrices is done on the basis of theory presented in [5].

1.1.4. Computational Work

Let m and n be the number of components of $\mathbf{x} \in K_X$ and of $\mathbf{y} \in K_Y$, respectively, where K_X and K_Y each contains N signals. The known filtering techniques (e.g. see [1]–[8,11], [23,25]), applied to \mathbf{x} and \mathbf{y} , require computation of a product of an $m \times n$ matrix and an $n \times n$ matrix, as well as computation of an $n \times n$ inverse or pseudo-inverse matrix for each pair of signals $\mathbf{x} \in K_X$ and $\mathbf{y} \in K_Y$. This requires $O(2mn^2)$ and $O(26n^3)$ flops, respectively [26]. Thus, for the processing of all signals in K_X and K_Y , the filters in [1]–[8,11], [23,25] require $O(2mn^2N) + O(26n^3N)$ operations.

Alternatively, K_X and K_Y can be represented by vectors, χ and γ , each with mN and nN components, respectively. In such a case, the techniques in [1]–[8,11], [23,25] can be applied to χ and γ as opposed to each signals in K_X and K_Y . computational requirement is then $O(2mn^2N^2)$ and $O(26n^3N^3)$ operations, respectively [26].

In both cases, but especially when N is large, computational work associated with the approaches [1]–[8,11], [23,25] becomes unreasonable hard.

For the filter $\mathcal{F}^{(p-1)}$ to be introduced below, the associated computational work is substantially less. This is because $\mathcal{F}^{(p-1)}$ requires computation of only p pseudo-inverse matrices associated with p selected signals in K_X , where p is much less than the number of signals in K_X . Therefore, for processing of the signal sets, K_X and K_Y , $\mathcal{F}^{(p-1)}$ requires only $O(2mn^2p) + O(26n^3p)$ flops where $p \ll N$. This comparison is illustrated in Section 5.

1.2. Relevant works

Some particular filtering techniques relevant to the method proposed below are as follows.

1.2.1. Generic Optimal Linear (GOL) Filter [5]

The generic optimal linear (GOL) filter in [5] is a generalization of the Wiener filter to the case when covariance matrix is not invertible and observable signal is arbitrarily noisy (i.e. when, in

particular, noise is not necessarily additive and Gaussian). The GOL filter has been developed for processing an individual stochastic signal. Some ideas from [5] are used in the proof of Theorem 1 below.

1.2.2. Simplicial Canonical Piecewise Linear Filter [23]

A complex Wiener adaptive filter was developed in [23] from the two-dimensional complex-valued simplicial canonical piecewise linear filter [24]. The filter in [23] was developed for the processing of an individual stochastic signal and can be exploited when the reference signal is known and a 'covariance-like' matrix is invertible. The latter precludes an application to the signal types considered, for example, in Section 5: the matrices used in [23] are not invertible for the signals as those in Section 5. Similarly, the filters studied in [8,11] were developed for the processing of a single signal when the covariance matrices are invertible.

For the filter proposed here, these restrictions are removed.

1.2.3. Adaptive Piecewise Linear Filter [22]

A piecewise linear filter in [22] was proposed for a fixed image denoising (given by a matrix), corrupted by an additive Gaussian noise. That is, the method involved a non stochastic reference signal and required its knowledge. No theoretical justification for the filter was given in [22].

1.2.4. Averaging Polynomial Filter [10,12]

The averaging polynomial filter proposed in [10,12] was developed for the purpose of processing infinite signal sets. The filter was based on an argument involving the 'averaging' over sets of signals under consideration. This device allows one to determine a single filter for the processing of infinite signal sets. At the same time, it leads to an increase in the associated error when signals differ considerably from each other. This effect is illustrated in Section 5 below.

1.2.5. Other Relevant Filters

The technique developed in [13] is an extension of the GOL filter to the constraint problem with respect to the filter rank. It concerns data compression.

The methods in [6,7,15,16] have been developed for deterministic signals. Motivated by the results achieved in [15,16], adaptive filters were elaborated in [17]. A theoretical basis for the device proposed in [15,16] is provided in [18].

We note that the idea of piecewise linear filtering has been used in the literature in several very different conceptual frameworks, despite exploiting some very similar terms (as in [15]–[24]). At the same time, a common feature of those techniques is that they were developed for the processing of a single signal, not of large signal sets as in this paper. In particular, piecewise linear filters in [19] have been obtained by arranging linear filters and thresholds in a tree structure. Piecewise linear filters discussed in [20] were developed using so-called threshold decomposition, which is a segmentation operator exploited to split a signal into a set of multilevel components. Filter design methods for piecewise linear systems proposed in [21] were based on a piecewise Lyapunov function.

1.3. Difficulties Associated with the Known Filtering Techniques

Basic difficulties associated with applying the known filtering techniques to the case under consideration (i.e. to processing of large signal sets, K_x and K_y) are that:

- (i) they require an information on each reference signal (in the form of a sample, for example),
- (ii) matrices used in the known filters can be not invertible (as in the simulations considered below in Section 5) and then the filter does not exist, and
- (iii) the associated computation work may require a very long time. For example, in simulations (Section 5), MATLAB was out of memory for computing the GOL filter [5] when each of sets K_x and K_y was represented by a long vector (this option has been discussed in Section 1.1.4 above).

1.4. Differences from the Known Filtering Techniques

The differences from the known filtering techniques discussed above are as follows.

(i) We consider a single filter that processes arbitrarily large input-output sets of stochastic signal-vectors. The known filters [1]–[9], [11,13,15]–[25] have been developed for the processing of an individual signal-vector only. In the case of their application to arbitrarily large signal sets, they imply difficulties described in Sections 1.1 and 1.3 above.

(ii) As a result, our piecewise linear filter model (Section 3), the statement of the problem (Section 3.3 below) and consequently, the device of its solution (Section 4 below) are different from those considered in [15]–[24]. In this regard, see also Section 1.2.5.

(iii) The above naturally leads to a new structure of the filter (presented in Sections 3.4 and 4.2 below) which is very different from the known ones.

1.5. Contribution

In general, for the processing of large data sets, the proposed filter allows us to achieve better results in comparison with the known techniques in [1]–[25]. In particular, it allows us to

- (i) achieve a desired accuracy in signal estimation³,
- (ii) exploit a prior information only on few reference signals, p , from the set K_X that contains $N \gg p$ signals or even infinite number of signals,
- (iii) find a single filter to process any signal from the arbitrarily large signal set,
- (vi) determine the filter in terms of pseudo-inverse matrices so that the filter always exists, and
- (v) decrease the computational load compared to the related known techniques.

2. Some Preliminaries

2.1. Notation

The signal sets we consider are, in fact, special representations of time series.

Let (Ω, Σ, μ) be a probability space⁴, and K_X and K_Y be arbitrarily large sets of signals such that

$$K_X = \{\mathbf{x}(t, \cdot) \in L^2(\Omega, \mathbb{R}^m) \mid t \in T\} \quad \text{and} \quad K_Y = \{\mathbf{y}(t, \cdot) \in L^2(\Omega, \mathbb{R}^n) \mid t \in T\}$$

where $T := [a, b] \subseteq \mathbb{R}$. We interpret $\mathbf{x}(t, \cdot)$ as a reference signal and $\mathbf{y}(t, \cdot)$ as an observable signal, an input to the filter \mathcal{F} studied below⁵. The variable $t \in T \subseteq \mathbb{R}$ represents time.⁶ Then, for example, the stochastic signal $\mathbf{x}(t, \cdot)$ can be interpreted as an arbitrary stationary time series.

Let $\{t_k\}_1^p \subset T$ be a sequence of fixed time-points such that

$$a = t_1 < \dots < t_p = b. \quad (1)$$

Because of the partition (1), the sets K_Y and K_X are divided in ‘smaller’ subsets $K_{X,1}, \dots, K_{X,p-1}$ and $K_{Y,1}, \dots, K_{Y,p-1}$, respectively, so that, for each $j = 1, \dots, p$,

$$K_{X,j} = \{\mathbf{x}(t, \cdot) \mid t_j \leq t \leq t_{j+1}\} \quad \text{and} \quad K_{Y,j} = \{\mathbf{y}(t, \cdot) \mid t_j \leq t \leq t_{j+1}\}. \quad (2)$$

³ This means that any desired accuracy is achieved theoretically, as is shown in Section 4.4 below. In practice, of course, the accuracy is increased to a prescribed reasonable level.

⁴ As usually, $\Omega = \{\omega\}$ is the set of outcomes, Σ a σ -field of measurable subsets in Ω and $\mu : \Sigma \rightarrow [0, 1]$ an associated probability measure on Σ . In particular, $\mu(\Omega) = 1$.

⁵ In an intuitive way \mathbf{y} can be regarded as a noise-corrupted version of \mathbf{x} . For example, \mathbf{y} can be interpreted as $\mathbf{y} = \mathbf{x} + \mathbf{n}$ where \mathbf{n} is white noise. In this paper, we do not restrict ourselves to this simplest version of \mathbf{y} and assume that the dependence of \mathbf{y} on \mathbf{x} and \mathbf{n} is arbitrary.

⁶ More generally, T can be considered as a set of parameter vectors $\alpha = (\alpha^{(1)}, \dots, \alpha^{(q)})^T \in C^q \subseteq \mathbb{R}^q$, where C^q is a q -dimensional cube, i.e., $\mathbf{y} = \mathbf{y}(\alpha, \cdot)$ and $\mathbf{x} = \mathbf{x}(\alpha, \cdot)$. One coordinate, say $\alpha^{(1)}$ of α , could be interpreted as time.

Therefore, K_Y and K_X can now be represented as

$$K_X = \bigcup_{j=1}^{p-1} K_{X,j} \quad \text{and} \quad K_Y = \bigcup_{j=1}^{p-1} K_{Y,j}. \quad (3)$$

2.2. Brief Description of the Problem

Given two arbitrarily large sets of stochastic signals, K_Y and K_X , find a single filter $\mathcal{F} : K_Y \rightarrow K_X$ that estimates the signal $\mathbf{x} \in K_X$ with a controlled, associated error. Note that in our formulation the set K_Y can be finite or infinite.

2.3. Brief description of the method

The solution of the above problem is based on the representation of the proposed filter in the form of a sum with $p - 1$ terms $\mathcal{F}_1, \dots, \mathcal{F}_{p-1}$ where each term, \mathcal{F}_j , is interpreted as a particular sub-filter (see (4) and (5) below). Such a filter is denoted by $\mathcal{F}^{(p-1)} : K_Y \rightarrow K_X$.

The sub-filter \mathcal{F}_j transforms signals that belong to 'piece' $K_{Y,j}$ of set K_Y to signals in 'piece' $K_{X,j}$ of K_X , i.e. $\mathcal{F}_j : K_{Y,j} \rightarrow K_{X,j}$. Each sub-filter \mathcal{F}_j depends on two parameters, α_j and \mathcal{B}_j .

The prime idea is to determine \mathcal{F}_j (i.e. α_j and \mathcal{B}_j) separately, for each $j = 1, \dots, p - 1$. The required α_j and \mathcal{B}_j follow from the solutions of the equation (11) and an associated minimization problem (11) (see Sections 3.4 and 4.2 below). This procedure adjusts \mathcal{F}_j so that the error associated with the estimation of $\mathbf{x}(t, \cdot) \in K_{X,j}$ is minimal.

A motivation for such a structure of the filter $\mathcal{F}^{(p-1)}$ is as follows. The method of determining α_j and \mathcal{B}_j provides an estimate $\mathcal{F}_j[\mathbf{y}(t, \cdot)]$ that interpolates $\mathbf{x}(t, \cdot) \in K_{X,j}$ at $t = t_j$ and $t = t_{j+1}$. In other words, the filter is flexible to variations in the sets of observed and reference signals K_Y and K_X , respectively. Due to this way of determining \mathcal{F}_j , it is natural to expect that the processing of a 'smaller' signal set, $K_{Y,j}$, may lead to a smaller associated error than that for the processing of the whole set K_Y by a filter which is not specifically adjusted to each particular piece $K_{Y,j}$.

As a result, $\mathcal{F}^{(p-1)}[\mathbf{y}(t, \cdot)]$ represents a special piecewise interpolation procedure and, thus, should be attributed with the associated advantages such as, for example, the high accuracy of estimation.

In Section 4.4, this observation is confirmed. In Sections 4.5 and 5, it is also shown that the proposed technique allows us to avoid the difficulties discussed in Section 1.3 above.

3. Description of the Problem

3.1. Piecewise Linear Filter Model

Let $\mathcal{F}^{(p-1)} : K_Y \rightarrow K_X$ be a filter such that, for each $t \in T$,

$$\mathcal{F}^{(p-1)}[\mathbf{y}(t, \cdot)] = \sum_{j=1}^{p-1} \delta_j \mathcal{F}_j[\mathbf{y}(t, \cdot)], \quad (4)$$

where

$$\mathcal{F}_j[\mathbf{y}(t, \cdot)] = \alpha_j + \mathcal{B}_j[\mathbf{y}(t, \cdot)] \quad \text{and} \quad \delta_j = \begin{cases} 1, & \text{if } t_j \leq t \leq t_{j+1}, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Here, \mathcal{F}_j is a sub-filter defined for $t_j \leq t \leq t_{j+1}$. In (5), $\alpha_j = [\alpha_j^{(1)}, \dots, \alpha_j^{(m)}]^T \in \mathbb{R}^m$ and $\mathcal{B}_j : L^2(\Omega, \mathbb{R}^n) \rightarrow L^2(\Omega, \mathbb{R}^m)$ is a linear operator given by a matrix $B_j \in \mathbb{R}^{m \times n}$, so that

$$[\mathcal{B}_j(\mathbf{y})](t, \omega) = B_j[\mathbf{y}(t, \omega)].$$

Thus, \mathcal{F}_j is defined by an operator $F_j : \mathbb{R}^n \rightarrow \mathbb{R}^m$ such that

$$F_j[\mathbf{y}(t, \omega)] = \alpha_j + B_j[\mathbf{y}(t, \omega)]. \quad (6)$$

Filter $\mathcal{F}^{(p-1)}$ defined by (4)-(6) is called the piecewise filter⁷.

3.2. Assumptions

In the known approaches related to filtering of stochastic signals (e.g. see [1]–[13,23,25]), it is assumed that covariance matrices formed from the reference signal and observed signal are known or can be estimated.

The assumption used here is similar. The covariance matrices that are assumed to be known or can be estimated, are formed from selected signal pairs $\{\mathbf{x}(t_j, \cdot), \mathbf{y}(t_j, \cdot)\}$ with $j = 1, \dots, p$ and p to be a small number⁸, $p \ll N$, where N is the number of signals in K_X or K_Y .

3.3. The Problem

In (4)-(6), parameters of the filter $\mathcal{F}^{(p-1)}$, i.e. vector α_j and matrix B_j , for $j = 1, \dots, p-1$, are unknown. Therefore, under the assumptions described in Section 3.2, the problem is to determine α_j and B_j , for $j = 1, \dots, p-1$. The related problem is to estimate an error associated with the filter $\mathcal{F}^{(p-1)}$.

Solutions to the both problems are given in Sections 4.2 and 4.4, respectively. In particular, in the following Section 3.4, interpolation conditions (8) and (11) are introduced that lead to a determination of α_j and B_j .

3.4. Interpolation Conditions

Let us denote

$$\|\mathbf{x}(t_j, \cdot)\|_{\Omega}^2 = \int_{\Omega} \|\mathbf{x}(t_j, \omega)\|_2^2 d\mu(\omega) \quad (7)$$

where $\|\mathbf{x}(t_j, \omega)\|_2$ is the Euclidean norm of $\mathbf{x}(t_j, \omega) \in \mathbb{R}^m$.

For $t = t_1$, let $\hat{\mathbf{x}}(t_1, \cdot)$ be an estimate of $\mathbf{x}(t_1, \cdot)$ determined by known methods [1]–[13,23,25]. This is the initial condition of the proposed technique.

For $j = 1, \dots, p-1$, each sub-filter F_j in (5)-(6) is defined so that α_j and B_j satisfy the conditions as follows.

Sub-filter \mathcal{F}_1 : For $j = 1$, α_1 and B_1 solve

$$\hat{\mathbf{x}}(t_1, \cdot) = \alpha_1 + B_1[\mathbf{y}(t_1, \cdot)] \quad \text{and} \quad \min_{B_1} \|\alpha_1 + B_1[\mathbf{y}(t_2, \cdot)] - \hat{\mathbf{x}}(t_2, \cdot)\|_{\Omega}^2, \quad (8)$$

respectively. Then an estimate of $\mathbf{x}(t, \cdot)$, $\hat{\mathbf{x}}(t, \cdot)$, for $t \in [t_1, t_2]$, is determined as

$$\hat{\mathbf{x}}(t, \cdot) = \mathcal{F}_1[\mathbf{y}(t, \cdot)] = \alpha_1 + B_1[\mathbf{y}(t, \cdot)] = \hat{\mathbf{x}}(t_1, \cdot) + B_1[\mathbf{y}(t, \cdot) - \mathbf{y}(t_1, \cdot)] \quad (9)$$

where α_1 and B_1 satisfy (8). In particular, $\alpha_1 = \hat{\mathbf{x}}(t_1, \cdot) - B_1[\mathbf{y}(t_1, \cdot)]$ and

$$\hat{\mathbf{x}}(t_2, \cdot) = \mathcal{F}_1[\mathbf{y}(t_2, \cdot)].$$

Extending this procedure up to $j = k-1$, where $k = 3, \dots, p$, we set the following. Let $\hat{\mathbf{x}}(t_{k-1}, \cdot)$ be an estimate of $\mathbf{x}(t_{k-1}, \cdot)$ defined by the preceding steps as

$$\hat{\mathbf{x}}(t_{k-1}, \cdot) = \mathcal{F}_{k-2}[\mathbf{y}(t_{k-1}, \cdot)]. \quad (10)$$

Then sub-filter \mathcal{F}_{k-1} is defined as follows.

⁷ Hereinafter, we will use a non-curly symbol to denote an operator and associated matrix (e.g., the operator $\mathcal{F}_j : L^2(\Omega, \mathbb{R}^n) \rightarrow L^2(\Omega, \mathbb{R}^m)$ and the associated matrix $F_j \in \mathbb{R}^{m \times n}$ are denoted by F_j).

⁸ It is worthwhile to note that it is not assumed that the covariance matrices are known for each signal pair from $K_X \times K_Y$, $\{\mathbf{x}(t, \cdot), \mathbf{y}(t, \cdot)\}$ with $t \in [a, b]$.

Sub-filter \mathcal{F}_{k-1} : For $j = k - 1, \alpha_{k-1}$ and \mathcal{B}_{k-1} solve

$$\hat{\mathbf{x}}(t_{k-1}, \cdot) = \alpha_{k-1} + \mathcal{B}_{k-1}[\mathbf{y}(t_{k-1}, \cdot)] \quad \text{and} \quad \min_{\mathcal{B}_{k-1}} \|\mathbf{x}(t_k, \cdot) - \alpha_{k-1} - \mathcal{B}_{k-1}[\mathbf{y}(t_k, \cdot)]\|_{\Omega}^2, \quad (11)$$

respectively. Then an estimate of $\mathbf{x}(t, \cdot)$, $\hat{\mathbf{x}}(t, \cdot)$, for $t \in [t_{k-1}, t_k]$, is determined as

$$\hat{\mathbf{x}}(t, \cdot) = \mathcal{F}_{k-1}[\mathbf{y}(t, \cdot)] = \alpha_{k-1} + \mathcal{B}_{k-1}[\mathbf{y}(t, \cdot)] = \hat{\mathbf{x}}(t_{k-1}, \cdot) + \mathcal{B}_1[\mathbf{y}(t, \cdot) - \mathbf{y}(t_{k-1}, \cdot)]. \quad (12)$$

The conditions (8) and (11) are motivated by the device of piecewise function interpolation and associated advantages [14].

Filter $\mathcal{F}^{(p-1)}$ of the form (4)-(5) with α_j and \mathcal{B}_j satisfying (8) and (11) is called the *piecewise linear interpolation filter*. The pair of signals $\{\mathbf{x}(t_k, \cdot), \mathbf{y}(t_k, \cdot)\}$ associated with time t_k defined by (1) is called the *interpolation pair*.

4. Main Results

4.1. General Device

In accordance with the scheme presented in Sections 3.1 and 3.4 above, an estimate of the reference signal $\mathbf{x}(t, \cdot)$, for any $t \in T = [a, b]$, by the piecewise linear interpolation filter $\mathcal{F}^{(p-1)}$, is given by

$$\hat{\mathbf{x}}(t, \cdot) = \mathcal{F}^{(p-1)}[\mathbf{y}(t, \cdot)] = \sum_{j=1}^{p-1} \delta_j \mathcal{F}_j[\mathbf{y}(t, \cdot)], \quad (13)$$

where, for each $j = 1, \dots, p - 1$, the sub-filter \mathcal{F}_j is given by (5), and is defined from the interpolation conditions (8) and (11).

Below, we show how to determine \mathcal{F}_j to satisfy the conditions (8) and (11).

4.2. Determination of Piecewise Linear Interpolation Filter

Let us denote

$$\mathbf{z}(t_j, t_{j+1}, \cdot) = \mathbf{x}(t_{j+1}, \cdot) - \hat{\mathbf{x}}(t_j, \cdot) \quad \text{and} \quad \mathbf{w}(t_j, t_{j+1}, \cdot) = \mathbf{y}(t_{j+1}, \cdot) - \mathbf{y}(t_j, \cdot). \quad (14)$$

We need to represent $\mathbf{z}(t_j, t_{j+1}, \cdot)$ and $\mathbf{w}(t_j, t_{j+1}, \cdot)$ in terms of their components as follows:

$$\begin{aligned} \mathbf{z}(t_j, t_{j+1}, \cdot) &= [\mathbf{z}^{(1)}(t_j, t_{j+1}, \cdot), \dots, \mathbf{z}^{(m)}(t_j, t_{j+1}, \cdot)]^T \\ \text{and} \quad \mathbf{w}(t_j, t_{j+1}, \cdot) &= [\mathbf{w}^{(1)}(t_j, t_{j+1}, \cdot), \dots, \mathbf{w}^{(n)}(t_j, t_{j+1}, \cdot)]^T, \end{aligned}$$

where $\mathbf{z}^{(j)}(t_j, t_{j+1}, \cdot) \in L^2(\Omega, \mathbb{R})$ and $\mathbf{w}^{(i)}(t_j, t_{j+1}, \cdot) \in L^2(\Omega, \mathbb{R})$ are stochastic variables, for all $j = 1, \dots, m$.

Then we can introduce the covariance matrix

$$E_{z_j w_j} = \left\{ \left\langle \mathbf{z}^{(i)}(t_j, t_{j+1}, \cdot), \mathbf{w}^{(k)}(t_j, t_{j+1}, \cdot) \right\rangle \right\}_{i,k=1}^{m,n}, \quad (15)$$

where $\left\langle \mathbf{z}^{(i)}(t_j, t_{j+1}, \cdot), \mathbf{w}^{(k)}(t_j, t_{j+1}, \cdot) \right\rangle = \int_{\Omega} \mathbf{z}^{(i)}(t_j, t_{j+1}, \omega) \mathbf{w}^{(k)}(t_j, t_{j+1}, \omega) d\mu(\omega)$.

Below, M^\dagger is the Moor-Penrose generalized inverse of a matrix M .

Now, we are in a position to establish the main results.

Theorem 1. *Let*

$$K_X = \{\mathbf{x}(t, \cdot) \in L^2(\Omega, \mathbb{R}^m) \mid t \in T = [a, b]\} \quad \text{and} \quad K_Y = \{\mathbf{y}(t, \cdot) \in L^2(\Omega, \mathbb{R}^n) \mid t \in T = [a, b]\}$$

be sets of reference signals and observed signals, respectively. Let $t_j \in [a, b]$, for $j = 1, \dots, p$, be such that

$$a = t_1 < \dots < t_p = b.$$

For $t = t_1$, let $\hat{\mathbf{x}}(t_1, \cdot)$ be a known estimate of $\mathbf{x}(t_1, \cdot)$ ⁹. Then, for any $t \in [a, b]$, the proposed piecewise linear interpolation filter $\mathcal{F}^{(p-1)} : L^2(\Omega, \mathbb{R}^n) \rightarrow L^2(\Omega, \mathbb{R}^m)$ transforming any signal $\mathbf{y}(t, \cdot) \in L^2(\Omega, \mathbb{R}^m)$ to an estimate of $\mathbf{x}(t, \cdot)$, $\hat{\mathbf{x}}(t, \cdot)$, is given by

$$\hat{\mathbf{x}}(t, \cdot) = \mathcal{F}^{(p-1)}[\mathbf{y}(t, \cdot)] = \sum_{j=1}^{p-1} \delta_j \mathcal{F}_j[\mathbf{y}(t, \cdot)] \quad (16)$$

where

$$\mathcal{F}_j[\mathbf{y}(t, \cdot)] = \hat{\mathbf{x}}(t_j, \cdot) + \mathcal{B}_j[\mathbf{y}(t, \cdot) - \mathbf{y}(t_j, \cdot)], \quad (17)$$

$$\hat{\mathbf{x}}(t_j, \cdot) = \mathcal{F}_{j-1}[\mathbf{y}(t_j, \cdot)] \quad (\text{for } j = 2, \dots, p-1), \quad (18)$$

$$\mathcal{B}_j = E_{z_j w_j} E_{w_j w_j}^\dagger + M_{B_j} [I_n - E_{w_j w_j} E_{w_j w_j}^\dagger], \quad (19)$$

and where I_n is the $n \times n$ identity matrix and M_{B_j} is an $m \times n$ arbitrary matrix.

Proof: The proof of Theorem 1 is given in Section A. □

It is worthwhile to observe that, due to an arbitrary matrix M_{B_j} in (19), the filter $\mathcal{F}^{(p-1)}$ is not unique. In particular, M_{B_j} can be chosen as the zero matrix \mathbb{O} similarly to the generic optimal linear [5] (which is also not unique by the same reason).

4.3. Numerical Realization of Filter $\mathcal{F}^{(p-1)}$ and Associated Algorithm

4.3.1. Numerical Realization

In practice, the set $T = [a, b]$ (see Section 2.1) is represented by a finite set $\{\tau_k\}_{k=1}^N$, i.e. $[a, b] = [\tau_1, \tau_2, \dots, \tau_N]$ where $a \leq \tau_1 < \tau_2 < \dots < \tau_N \leq b$.

For $k = 1, \dots, N$, the estimate of $\mathbf{x}(\tau_k, \cdot)$, $\hat{\mathbf{x}}(\tau_k, \cdot)$, and observed signal $\mathbf{y}(\tau_k, \cdot)$ are represented by $m \times q$ and $n \times q$ matrices

$$\hat{X}^{(k)} = [\hat{\mathbf{x}}(\tau_k, \omega_1), \dots, \hat{\mathbf{x}}(\tau_k, \omega_q)] \quad \text{and} \quad Y^{(k)} = [\mathbf{y}(\tau_k, \omega_1), \dots, \mathbf{y}(\tau_k, \omega_q)]. \quad (20)$$

The sequence of fixed time-points $\{t_k\}_1^p \subset [a, b]$ introduced in (1) is such that

$$\tau_1 = t_1 < \dots < t_p = \tau_N, \quad (21)$$

where $t_1 = \tau_{n_0}$, $t_2 = \tau_{n_0+n_1}$, \dots , $t_p = \tau_{n_0+n_1+\dots+n_{p-1}}$, and where $n_0 = 1$ and n_1, \dots, n_{p-1} are positive integers such that $N = n_0 + n_1 + \dots + n_{p-1}$.

For $j = 1, \dots, p$, vectors $\hat{\mathbf{x}}(t_j, \cdot)$ and $\mathbf{y}(t_j, \cdot)$ associated with t_j in (21) are represented, respectively, by

$$\hat{X}_j^{(k)} = [\hat{\mathbf{x}}(t_j, \omega_1), \dots, \hat{\mathbf{x}}(t_j, \omega_N)] \quad \text{and} \quad Y_j = [\mathbf{y}(t_j, \omega_1), \dots, \mathbf{y}(t_j, \omega_N)].$$

⁹ As it has been mentioned in Section 3.4, $\hat{\mathbf{x}}(t_1, \cdot)$ can be determined by the known methods.

4.3.2. Algorithm

As it has been mentioned in Section 3.4, it is supposed that, for $t = t_1$, an estimate of X_1 , \hat{X}_1 , is known and can be determined by the known methods. This is the initial condition of the proposed technique.

On the basis of the results obtained in Sections 3.4 and 4.2, the performance algorithm of the proposed filter consists of the following steps. For $j = 1 \dots, p$, we write $N_j = n_0 + n_1 + \dots + n_{j-1}$.

Initial parameters: $Y^{(1)}, \dots, Y^{(N)}$, $\{t_j\}_{j=1}^p$ (see (21)), $\{E_{z_j w_j}\}_{j=1}^p$, $\{E_{w_j w_j}\}_{j=1}^p$ (see (14) and (15)), \hat{X}_1 , $n_0 = 1$, $N_0 = 0$ and $M_{B_j} = \mathbb{O}$, for $j = 1, \dots, p-1$.

(Possible ways to get estimates of $E_{z_j w_j}$ and $E_{w_j w_j}$ are discussed below in Section 4.5.)

Final parameters: $\hat{X}^{(2)}, \hat{X}^{(3)}, \dots, \hat{X}^{(N)}$.

Algorithm:

- for $j = 1$ to p do
 - begin

$$B_j = E_{z_j w_j} E_{w_j w_j}^\dagger;$$

- for $k = N_{j-1} + 1$ to N_j do
 - begin

$$\hat{X}^{(k)} = \hat{X}_j + B_j(Y^{(k)} - Y_j);$$

$$\hat{X}_{j+1} = \hat{X}^{(N_j)};$$

- end
- end

4.4. Error Analysis

It is natural to expect that the error associated with the piecewise interpolating filter $\mathcal{F}^{(p-1)}$ decreases when $\max_{j=1, \dots, p-1} \Delta t_j$ decreases. Below, in Theorem 3, we justify that this observation is true. To this end, first, in the following Theorem 2, we establish an estimate of the error associated with the filter F .

Let us introduce the norm by

$$\|\mathbf{x}(t, \cdot)\|_{T, \Omega}^2 = \frac{1}{b-a} \int_T \|\mathbf{x}(t, \cdot)\|_{\Omega}^2 dt. \quad (22)$$

We also denote $\|\mathbf{x}(t, \omega)\|_{T, \Omega}^2 = \|\mathbf{x}(t, \cdot)\|_{T, \Omega}^2$.

Let us suppose that $\mathbf{x}(\cdot, \omega)$ and $\mathbf{y}(\cdot, \omega)$ are Lipschitz continuous signals, i.e. that there exist real non-negative constants λ_j and γ_j , with $j = 1, \dots, p$, such that, for $t \in [t_j, t_{j+1}]$,

$$\|\mathbf{x}(t, \omega) - \mathbf{x}(t_j, \omega)\|_{T, \Omega}^2 \leq \lambda_j \Delta t_j \quad \text{and} \quad \|\mathbf{y}(t, \omega) - \mathbf{y}(t_{j+1}, \omega)\|_{T, \Omega}^2 \leq \gamma_j \Delta t_j \quad (23)$$

where $\Delta t_j = |t_{j+1} - t_j|$.

Theorem 2. Under the conditions (23) the error associated with the piecewise interpolation filter, $\|\mathbf{x}(t, \omega) - F^{(p-1)}[\mathbf{y}(t, \omega)]\|_{T, \Omega}^2$, is estimated as follows:

$$\begin{aligned} & \|\mathbf{x}(t, \omega) - F^{(p-1)}[\mathbf{y}(t, \omega)]\|_{T, \Omega}^2 \\ & \leq \max_{j=1, \dots, p-1} \left[(\lambda_j + \gamma_j \|B_j\|^2) \Delta t_j + \|E_{z_j z_j}^{1/2}\|^2 - \|E_{z_j w_j} (E_{w_j w_j}^{1/2})^\dagger\|^2 \right]. \end{aligned} \quad (24)$$

Proof: The proof of Theorem 2 is given in Section A. □

Further, to show that the error of the reference signal estimate tends to the zero, we need to assume that, for $t \in [t_1, t_2]$, the known estimate $\hat{\mathbf{x}}(t_1, \omega)$ differs from $\mathbf{x}(t, \omega)$ for the value of the order Δt_1 , i.e. that, for some constant $c_1 \geq 0$,

$$\|\mathbf{x}(t, \omega) - \hat{\mathbf{x}}(t_1, \omega)\|_{\Omega}^2 \leq c_1 \Delta t_1, \quad \text{for } t \in [t_1, t_2]. \quad (25)$$

Theorem 3. Let the conditions (23) and (25) be true. Then the error associated with the piecewise interpolating filter F , $\|\mathbf{x}(t, \omega) - F^{(p-1)}[\mathbf{y}(t, \omega)]\|_{T, \Omega}^2$, decreases in the following sense:

$$\|\mathbf{x}(t, \omega) - F^{(p-1)}[\mathbf{y}(t, \omega)]\|_{T, \Omega}^2 \rightarrow 0 \quad \text{as} \quad \max_{j=1, \dots, p-1} \Delta t_j \rightarrow 0 \quad \text{and} \quad p \rightarrow \infty. \quad (26)$$

Proof: The proof of Theorem 3 is given in Section A. □

Remark 1. We would like to emphasize that the statement of Theorem 3 is fulfilled only under assumptions (23) and (25). At the same time, the assumptions (23) and (25) are not restrictive from a practical point of view. The condition (23) is true for Lipschitz continuous signals \mathbf{x} and \mathbf{y} , i.e. for very wide class of signals. The condition (25) is achieved by a choosing an appropriate known method (e.g. see [1]–[13,23,25]) to find the estimate $\hat{\mathbf{x}}(t_1, \omega)$ used in the proposed filter $\mathcal{F}^{(p-1)}$ (see (8) and Theorem 1).

4.5. Some Remarks Related to the Assumptions of the Method

As it has been mentioned in Section 3.2, for $j = 1, \dots, p$, matrices $E_{z_j w_j}$ and $E_{w_j w_j}$ in (19) are assumed to be known or can be estimated. Here, p is a chosen number of selected interpolation signal pairs (see Section 3.4). We note that normally p is much smaller than the number of input-output signals $\mathbf{x}(t, \cdot)$ and $\mathbf{y}(t, \cdot)$. Therefore, to estimate any signal $\mathbf{x}(t, \cdot)$ from an arbitrarily large set K_x , only a small number, p , of matrices $E_{z_j w_j}$ and $E_{w_j w_j}$ should be estimated (or be known). This issue has also been discussed in Sections 1.1.1 and 1.1.4.

By the proposed method, $\mathbf{x}(t, \cdot)$ is estimated for $t \in [t_j, t_{j+1}]$. While $E_{w_j w_j}$ in (19) can be directly estimated from observed signals $\mathbf{y}(t_{j+1}, \cdot)$ and $\mathbf{y}(t_j, \cdot)$, an estimate of matrix $E_{z_j w_j}$ depends on the reference signal $\mathbf{x}(t_{j+1}, \cdot)$ (see (14) and (15)) which is unknown (because the estimate is considered for $t \in [t_j, t_{j+1}]$).

Some possible approaches to an estimation of matrix $E_{z_j w_j}$ could be as follows.

1. In the general case, when $\mathbf{x}(t, \cdot)$ and $\mathbf{y}(t, \cdot)$ are arbitrary signals as discussed in Section 2.1 above, matrix $E_{z_j w_j}$ can be estimated as proposed, for example, in [27], from samples of z_j and w_j .
2. In the case of incomplete observations, the method proposed in [28,29] can be used.
3. Let $E_{z_j w_j}$ be a matrix obtained from matrix $E_{z_j w_j}$ where the term $\mathbf{x}(t_{j+1}, \cdot)$ is replaced by $\hat{\mathbf{x}}(t, \cdot)$ with $t \in [t_{j-1}, t_j]$. Since $\hat{\mathbf{x}}(t, \cdot)$ with $t \in [t_{j-1}, t_j]$ is known, matrix $E_{z_j w_j}$ can be considered as an estimate of $E_{z_j w_j}$.
4. In the important case of an *additive* noise, $E_{z_j w_j}$ can be represented in the explicit form. Indeed, if

$$\mathbf{y}(t, \cdot) = \mathbf{x}(t, \cdot) + \zeta(t, \cdot)$$

where $\zeta(t, \cdot) \in L^2(\Omega, \mathbb{R}^m)$ is a random noise, then $\mathbf{z}(t_j, t_{j+1}, \cdot) = \mathbf{y}(t_{j+1}, \cdot) - \zeta(t_{j+1}, \cdot) - \hat{\mathbf{x}}(t_j, \cdot)$ and matrix $E_{z_j w_j}$ can be represented as follows:

$$E_{z_j w_j} = E_{(y_{j+1} - \zeta_{j+1})(y_{j+1} - y_j)} - E_{\hat{\mathbf{x}}_j(y_{j+1} - y_j)} \quad (27)$$

We note that the RHS of (27) depends only on observed signals $\mathbf{y}(t_j, \cdot)$, $\mathbf{y}(t_{j+1}, \cdot)$, estimated signal $\hat{\mathbf{x}}(t_j, \cdot)$, and noise $\zeta(t_{j+1}, \cdot)$, not on the reference signal $\mathbf{x}(t_{j+1}, \cdot)$. In particular, in (27), the term $E_{\zeta_{j+1}(y_{j+1} - y_j)}$ can be estimated as $\pm(E[\zeta_{j+1}^2])^{1/2}(E[(y_{j+1} - y_j)^2])^{1/2}$ where $E[\zeta_{j+1}^2] = \int_{\Omega} [\zeta(t_{j+1}, \omega)]^2 d\mu(\omega)$. It is motivated by the Holder's inequality for integrals. The second term in (27), $E_{\hat{\mathbf{x}}_j(y_{j+1} - y_j)}$, can be estimated from the samples of $\hat{\mathbf{x}}(t_{j+1}, \cdot)$ and $\mathbf{y}(t_{j+1}, \cdot) - \mathbf{y}(t_j, \cdot)$.

We also note that the first term in the RHS of (27), $E_{(y_{j+1}-\xi_{j+1})(y_{j+1}-y_j)}$, is similar to the related covariance matrix in the Wiener filtering approach [5].

5. Other known ways to estimate $E_{\xi_{j+1}(y_{j+1}-y_j)}$ can be found in [5], Section 5.3.

In general, an estimation of covariance matrices is a special research topic which is not a subject of this paper. The relevant references can be found, for example, in [5,29].

5. Simulations

5.1. General Consideration

In these simulations, in accordance with Section 4.3.1, signal sets K_X and K_Y (see Section 2.1) are given by

$$K_X = \{\mathbf{x}(\tau_1, \cdot), \mathbf{x}(\tau_2, \cdot), \dots, \mathbf{x}(\tau_N, \cdot)\} \quad \text{and} \quad K_Y = \{\mathbf{y}(\tau_1, \cdot), \mathbf{y}(\tau_2, \cdot), \dots, \mathbf{y}(\tau_N, \cdot)\},$$

where, for $k = 1, \dots, N$, $\mathbf{x}(\tau_k, \cdot) \in L^2(\Omega, \mathbb{R}^m)$ and $\mathbf{y}(\tau_k, \cdot) \in L^2(\Omega, \mathbb{R}^n)$. In many practical problems (arising, for example, in a DNA analysis the number N is quite large, for instance, $N = \mathcal{O}(10^4)$).

We set $N = 141$ and $m = n = 116$. Thus, in these simulations, the interval $T = [a, b]$ (see Sections 2.1 and 4.3.1) is modelled as 141 points τ_k with $k = 1, \dots, 141$ so that $[a, b] = [\tau_1, \tau_2, \dots, \tau_{141}]$.

The sequence of fixed time-points $\{t_k\}_1^p \subset T$ in (1) is now such that

$$\tau_1 = t_1 < \dots < t_p = \tau_{141}. \quad (28)$$

Below, in Examples 1-12, four particular choices of the specific interpolation signal pairs $\{\mathbf{x}(t_j, \cdot), \mathbf{y}(t_j, \cdot)\}_1^p$ (introduced in Section 3.4) are considered, for $p = 5, 8, 15$ and 28. Points t_1, \dots, t_p are as follows.

For $p = 5, 8, 15$, if $j = 1, \dots, p$, then $t_j = t_j^{(p)} = \tau_1 + (j-1)\Delta_p$, respectively, where $\Delta_5 = 35$, $\Delta_8 = 20$ and $\Delta_{15} = 10$.

For $p = 28$, if $j = 1, \dots, p-1$, then $t_j = t_j^{(p)} = \tau_1 + (j-1)\Delta_{28}$, and if $j = p$, then $t_{28} = t_{28}^{(28)} = t_{27} + 6 = 141$, where $\Delta_{28} = 5$.

Signals $\mathbf{x}(\tau_k, \cdot)$ and $\mathbf{y}(\tau_k, \cdot)$ have been simulated as digital images represented by 116×256 matrices

$$X^{(k)} = [\mathbf{x}(\tau_k, \omega_1), \dots, \mathbf{x}(\tau_k, \omega_{256})] \quad \text{and} \quad Y^{(k)} = [\mathbf{y}(\tau_k, \omega_1), \dots, \mathbf{y}(\tau_k, \omega_{256})], \quad (29)$$

respectively, for $k = 1, \dots, 141$, so that $X^{(k)}$ represents an image that should be estimated from an observed image $Y^{(k)}$. A column of matrices $X^{(k)}$ and $Y^{(k)}$, $\mathbf{x}(\tau_k, \omega_i) \in \mathbb{R}^{116}$ and $\mathbf{y}(\tau_k, \omega_i) \in \mathbb{R}^{116}$, for $i = 1, \dots, 256$, represents a realization of signals $\mathbf{x}(\tau_k, \cdot)$ and $\mathbf{y}(\tau_k, \cdot)$, respectively.

Note that $X^{(1)}, \dots, X^{(141)}$ did not used in the piecewise linear filter $F^{(p-1)}$ below since they are not supposed to be known. They are represented here for illustration purposes only. In particular, $X^{(1)}, \dots, X^{(141)}$ are used to compare their estimates by different filters.

Observed noisy signals $Y^{(1)}, \dots, Y^{(141)}$ have been simulated in different forms presented by (40), (49), (50) and (51) in the Examples 1-12 below. We note that the considered observed signals are grossly corrupted.

To estimate the signals $X^{(1)}, \dots, X^{(141)}$ from the observed signals $Y^{(1)}, \dots, Y^{(141)}$, the proposed piecewise linear filter $F^{(p-1)}$, the generic optimal linear (GOL) filters [5] and the averaging polynomial filter [12] have been used.

The filters proposed in [12,13,22,23] have not been applied here by the reasons discussed in Section 1. In particular, the filter in [23] cannot be applied to signals represented by $Y^{(1)}, \dots, Y^{(141)}$ in the form (40), (49), (50) and (51) below because the associated inverse matrices used in [23] do not exist.

For signals under consideration (given by matrices $X^{(k)}$ and $Y^{(k)}$ with $k = 1, \dots, 141$), the filter $F^{(p-1)}$, the generic optimal linear (GOL) filters [5] and the averaging polynomial filter [10,12] are represented as follows.

(i) *Piecewise linear filter* $F^{(p-1)}$. For $j = 1, \dots, p$, $\{X_j, Y_j\}$ designates an interpolation pair defined similarly to that in Section 3.4. Each X_j and Y_j is associated with t_j in (28) so that

$$X_j = [\mathbf{x}(t_j, \omega_1), \dots, \mathbf{x}(t_j, \omega_{256})] \quad \text{and} \quad Y_j = [\mathbf{y}(t_j, \omega_1), \dots, \mathbf{y}(t_j, \omega_{256})].$$

The estimate $\hat{X}^{(k)}$ of $X^{(k)}$ by the filter $F^{(p-1)}$ is given by

$$\hat{X}^{(k)} = F^{(p-1)}[Y^{(k)}], \quad (30)$$

where, by (16)-(19) in Section 4.2,

$$F^{(p-1)}[Y^{(k)}] = \sum_{j=1}^{p-1} \delta_j F_j[Y^{(k)}], \quad \delta_j = \begin{cases} 1, & \text{if } t_j \leq \tau_k \leq t_{j+1}, \\ 0, & \text{otherwise,} \end{cases} \quad (31)$$

$$F_j^{(p-1)}[Y^{(k)}] = \hat{X}_j + B_j[Y^{(k)} - Y_j], \quad (32)$$

$$\hat{X}_j = F_{j-1}[Y_j], \quad \hat{X}_1 \text{ is given,} \quad (33)$$

$$B_j = E_{Z_j W_j} (E_{W_j W_j})^\dagger, \quad (34)$$

and where $E_{Z_j W_j}$ and $E_{W_j W_j}$ are estimates of matrices $E_{Z_j W_j}$ and $E_{W_j W_j}$ in (19), respectively. In particular, $E_{W_j W_j}$ can be represented in the form

$$E_{W_j W_j} = W_j W_j^T, \quad \text{where } W_j = Y_{j+1} - Y_j. \quad (35)$$

Further, matrix $E_{Z_j W_j}$ depends on $Z_j = X_{j+1} - \hat{X}_j$ where X_{j+1} is unknown. Therefore a determination of $E_{Z_j W_j}$ is reduced, in fact, to finding an estimate of X_{j+1} . Since it is customary to find $E_{Z_j W_j}$ in terms of signal samples [5], $E_{Z_j W_j}$ has been presented as

$$E_{Z_j W_j} = \tilde{Z}_j W_j^T, \quad \text{where } \tilde{Z}_j = \tilde{X}_{j+1} - \hat{X}_j \quad (36)$$

and \tilde{X}_{j+1} has been constructed from a sample of X_{j+1} as follows. The sample of X_{j+1} is a 116×128 matrix presented by odd columns of X_{j+1} . Then an estimate of X_{j+1} is chosen as a 116×256 matrix \tilde{X}_{j+1} where each odd column is a related odd column of X_{j+1} , and each even column is an average of two adjacent columns. The last column in \tilde{X}_{j+1} is the same as its preceding column.

This way of estimating $E_{Z_j W_j}$ was chosen for illustration purposes only. Other related methods have been considered in Section 4.5.

The errors associated with the filter $F^{(p-1)}$ are given by

$$\varepsilon_{k,F}^{(p-1)} = \left\| X^{(k)} - F^{(p-1)}[Y^{(k)}] \right\|_F^2, \quad \text{for } k = 1, \dots, 141. \quad (37)$$

(ii) *Generic optimal linear (GOL) filters* [5]. To each signal $Y^{(k)}$, an individual GOL filter W_k has also been applied, so that W_k estimates $X^{(k)}$ from $Y^{(k)}$ in the form

$$W_k Y^{(k)} = E_{X^{(k)} Y^{(k)}} E_{Y^{(k)} Y^{(k)}}^\dagger Y^{(k)},$$

for each $k = 1, \dots, 141$. Thus, the GOL filter W_k requires an estimate of 141 matrices $E_{X^{(k)} Y^{(k)}}$, for each $k = 1, \dots, 141$.

Similarly to matrix $E_{Z_j W_j}$ in the filter $F^{(p-1)}$ above, the matrix $E_{X^{(k)} Y^{(k)}}$ has been estimated from samples of each $X^{(k)}$, $\tilde{X}^{(k)}$, for each $k = 1, \dots, 141$.

One of the advantages of the proposed filter $F^{(p-1)}$ is that $F^{(p-1)}$ requires a smaller number, p , of samples of X_j , \tilde{X}_j , to be known (where $j = 1, \dots, p$).

The errors associated with filters W_k are given by

$$\epsilon_{k,w} = \|X^{(k)} - W_k Y^{(k)}\|_F^2. \quad (38)$$

(iii) *Averaging polynomial filters* [10,12]. By the methodology in [10], the averaging polynomial filter W is based on the use of the estimates of the covariance matrices, E_{XY} and E_{YY} , in the form

$$E_{XY} = \frac{1}{141} \sum_{k=1}^{141} \tilde{X}^{(k)} (Y^{(k)})^T \quad \text{and} \quad E_{YY} = \frac{1}{141} \sum_{k=1}^{141} Y^{(k)} (Y^{(k)})^T.$$

Then, for each, $k = 1, \dots, 141$, the estimate of $X^{(k)}$ is given by

$$WY^{(k)} = E_{XY} E_{YY}^\dagger Y^{(k)}.$$

The errors associated with the filter W are given by

$$\epsilon_{k_W} = \|X^{(k)} - WY^{(k)}\|_F^2, \quad \text{for } k = 1, \dots, 141. \quad (39)$$

5.2. Simulations with Signals Modelled from Images 'Plant': Application of Piecewise Interpolation Filter and GOL Filters

Here, results of simulations for reference signals represented by matrices $X^{(1)}, \dots, X^{(141)}$ (see (29) above) formed from images 'plant'¹⁰ are considered. Typical selected images $X^{(k)}$ are shown in Figure 9.

Observed noisy images $Y^{(1)}, \dots, Y^{(141)}$ have been simulated in the form

$$Y^{(k)} = X^{(k)} \bullet \text{randn}_{(k)} \bullet \text{rand}_{(k)}, \quad (40)$$

for each $k = 1, \dots, 141$. Here, \bullet means the Hadamard product, and $\text{randn}_{(k)}$ and $\text{rand}_{(k)}$ are 116×256 matrices with random entries. The entries of $\text{randn}_{(k)}$ are normally distributed with mean zero, variance one and standard deviation one. The entries of $\text{rand}_{(k)}$ are uniformly distributed in the interval $(0, 1)$. A typical example of such images is given in Figure 10 (a).

To demonstrate the effectiveness of the proposed filter $F^{(p-1)}$, sub-filters $F_j^{(p-1)}$ and associated interpolation signal pairs $\{X_j, Y_j\}_{j=1}^p$ have been chosen in four different ways as follows.

Example 1. First, for $p = 5$, the interpolation signal pairs are

$$\{X_1, Y_1\} = \{X^{(1)}, Y^{(1)}\}, \quad \{X_2, Y_2\} = \{X^{(35)}, Y^{(35)}\}, \quad \{X_3, Y_3\} = \{X^{(70)}, Y^{(70)}\}, \quad (41)$$

$$\{X_4, Y_4\} = \{X^{(105)}, Y^{(105)}\}, \quad \{X_5, Y_5\} = \{X^{(141)}, Y^{(141)}\}. \quad (42)$$

The error values $\{\epsilon_{k,F}^{(4)}\}_1^{141}$ associated with filter $F^{(4)}$ are evaluated by (37). The graph of $\{\epsilon_{k,F}^{(4)}\}_1^{141}$ is presented in Figure 11 (a).

Example 2. For $p = 8$, the interpolation signal pairs are

$$\{X_1, Y_1\} = \{X^{(1)}, Y^{(1)}\}, \quad \{X_j, Y_j\} = \{X^{(20(j-1))}, Y^{(20(j-1))}\}, \quad \text{for } j = 2, \dots, 7; \quad (43)$$

$$\text{and } \{X_8, Y_8\} = \{X^{(141)}, Y^{(141)}\}. \quad (44)$$

The error magnitudes $\{\epsilon_{k,F}^{(7)}\}_1^{141}$ associated with the piecewise interpolation filter $F^{(7)}$ constructed by (31)-(36) with the interpolation signal pairs given by (43)-(44) are diagrammatically shown in Figure 11 (b).

It follows from Figure 11 (b) that the errors associated with filter $F^{(7)}$ is less than those of filter $F^{(4)}$. This is a confirmation of Theorem 3.

¹⁰ The database is available in <http://sipi.usc.edu/services/database.html>.

Example 3. Further, for $p = 15$, the interpolation pairs are

$$\{X_1, Y_1\} = \{X^{(1)}, Y^{(1)}\}, \quad \{X_j, Y_j\} = \{X^{(10(j-1))}, Y^{(10(j-1))}\} \quad \text{for } j = 2, \dots, 14; \quad (45)$$

$$\text{and } \{X_{15}, Y_{15}\} = \{X^{(141)}, Y^{(141)}\}. \quad (46)$$

In Figure 11 (c), the errors $\{\varepsilon_{k,F}^{(15)}\}_1^{141}$ associated with the piecewise interpolation filter $F^{(15)}$ are presented. The Figure 11 (c) demonstrates a further confirmation of Theorem 3: the errors associated with the piecewise interpolation filter diminishes as p increases.

Example 4. Finally, the number of interpolation signal pairs $\{X_j, Y_j\}_{j=1}^p$ is $p = 29$ so that

$$\{X_1, Y_1\} = \{X^{(1)}, Y^{(1)}\}, \quad \{X_j, Y_j\} = \{X^{(5(j-1))}, Y^{(5(j-1))}\} \quad \text{for } j = 2, \dots, 28; \quad (47)$$

$$\text{and } \{X_{29}, Y_{29}\} = \{X^{(141)}, Y^{(141)}\}. \quad (48)$$

In this case, when p is grater than in the previous Examples 1-3, the errors $\{\varepsilon_{k,F}^{(29)}\}_1^{141}$ associated with the piecewise interpolation filter $F^{(29)}$ are smaller than those associated with filters $F^{(4)}$, $F^{(8)}$ and $F^{(15)}$ - see Figure 11 (d).

The diagrams of errors associated with the GOL filters [5] are also presented in Figure 11. It follows from Figure 11 that proposed filters $F^{(4)}$, $F^{(8)}$, $F^{(15)}$ and $F^{(29)}$ provide the better accuracy then that of the GOL filters.

At the same time, the filter $F^{(p-1)}$ is easier to implement since it requires less initial information compared to GOL filters, as it has been discussed in Sections 1.1.1 and 1.1.4.

Results of the application of the averaging polynomial filter [10,12] are discussed in Section 5.4 below.

5.3. Simulations with Signals Modelled from Images 'Boat': Application of Piecewise Interpolation Filter and GOL Filters

In this section, results of the simulations for a different type of signals than those considered in Section 5.2 above are presented. Here, the reference signals $X^{(1)}, \dots, X^{(141)}$ are formed from images 'boat'¹¹.

Observed noisy signals $Y^{(1)}, \dots, Y^{(141)}$ have been simulated in the form

$$Y^{(k)} = X^{(k)} \bullet \text{randn}_{(k)}, \quad (49)$$

for each $k = 1, \dots, 141$. The noise term is different from that in (40).

Typical selected images $X^{(k)}$ and $Y^{(k)}$ are shown in Figures 12 and 13, respectively.

As in Section 5.2, the piecewise interpolation filter $F^{(p-1)}$ is constructed by (31)-(36). In Examples 5-8 below, the number $p - 1$ of sub-filters $F_j^{(p-1)}$ and associated interpolation signal pairs $\{X_j, Y_j\}_{j=1}^p$ have been chosen in four different ways.

Example 5. First, similar to Example 1, the number of interpolation signal pairs $\{X_j, Y_j\}_{j=1}^p$ has been chosen as $p = 5$, and X_j and Y_j have been presented as in (41)-(42).

The error values $\{\varepsilon_{k,F}^{(4)}\}_1^{141}$ associated with the piecewise interpolation filter $F^{(4)}$ applied to these data are presented in Figure 14 (a).

Example 6. For the grater number of interpolation signal pairs than that in Example 5, $p = 8$, and for X_j and Y_j ($j = 1, \dots, 8$) chosen as in (43)-(44), the error magnitudes $\{\varepsilon_{k,F}^{(7)}\}_1^{141}$ associated with the piecewise interpolation filter $F^{(7)}$ are diagrammatically shown in Figure 14 (b). A comparison between Figures 14 (a) and (b) demonstrates that the increase in p implies the decrease in the errors associated with the filter $F^{(p-1)}$.

¹¹ The database is available in <http://sipi.usc.edu/services/database.html>.

Example 7. For $p = 15$, and for X_j and Y_j ($j = 1, \dots, 15$) chosen as in (45)-(46), the errors $\{\varepsilon_{k,F}^{(14)}\}_1^{141}$ associated with the piecewise interpolation filter $F^{(14)}$ are further less than those for filters $F^{(4)}$ and $F^{(7)}$. See Figure 14 (c) in this regard.

Example 8. The further increase in p to $p = 29$, confirms this tendency. The piecewise interpolation filter $F^{(28)}$ with X_j and Y_j ($j = 1, \dots, 29$) chosen similar to (47)-(48) produces the associated errors $\{\varepsilon_{k,F}^{(28)}\}_1^{141}$ represented in Figure 14 (d). They are, clearly, less than the errors associated with filters $F^{(4)}$, $F^{(7)}$ and $F^{(15)}$.

The errors associated with the GOL filters are also presented in Figures 14 (a)-(d). The figures clearly demonstrate the advantage of the piecewise interpolation filter $F^{(p-1)}$.

Results of the application of the averaging polynomial filter [12] are discussed in Section 5.4 below.

5.4. Results of Simulations for Averaging Polynomial Filter [10,12]

To further illustrate the effectiveness of the proposed piecewise interpolation filter, in this Section, results of simulations for the averaging polynomial filter [10,12] are presented. The filter has been applied to two different types of data considered in Sections 5.2 and 5.3.

Example 9. The filter [10,12] applied to signals considered in Section 5.2 gives the associated errors $\{\varepsilon_{k_W}\}_{k=1}^{141}$ (see (39)) represented in Figure 15 (a). For a comparison, the errors associated with the piecewise interpolation filter $F^{(28)}$ and the GOL filters [5] are also given in Figure 15 (a).

A typical example of the estimated signal by the averaging polynomial filter [10,12] is presented in Figure 10 (d) above.

Example 10. The averaging polynomial filter [12] applied to signals considered in Section 5.3 produces the associated errors $\{\varepsilon_{k_W}\}_{k=1}^{141}$ shown in Figure 15 (b). The errors associated with the piecewise interpolation filter $F^{(28)}$ are much smaller and they are not discerned in Figure 15 (b).

Together with Figures 10, 11, 13 and 14, Figures 15 (a) and (b) illustrate the advantage of the piecewise interpolation filter.

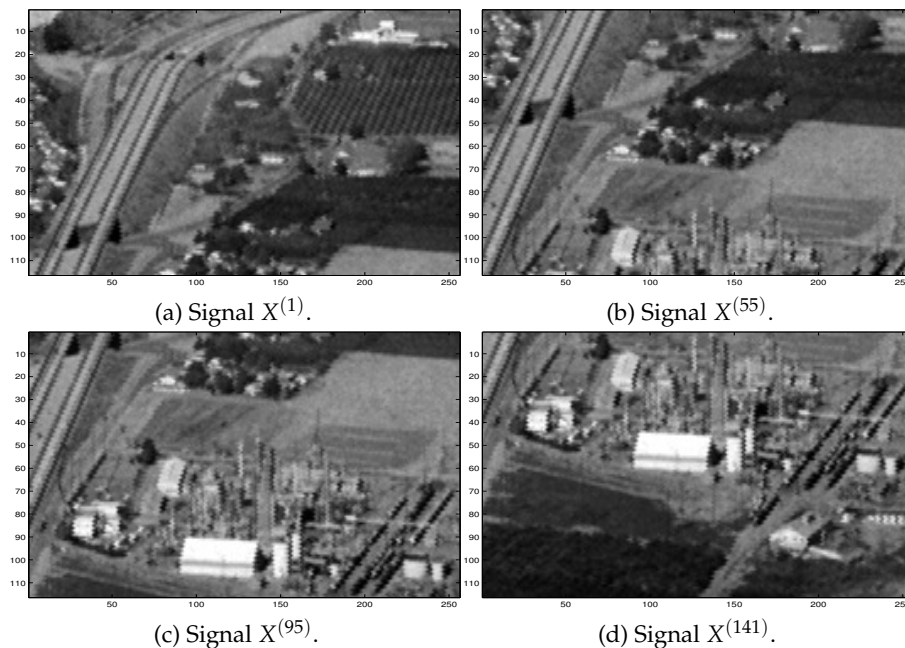


Figure 1. Examples of selected signals to be estimated from observed data.

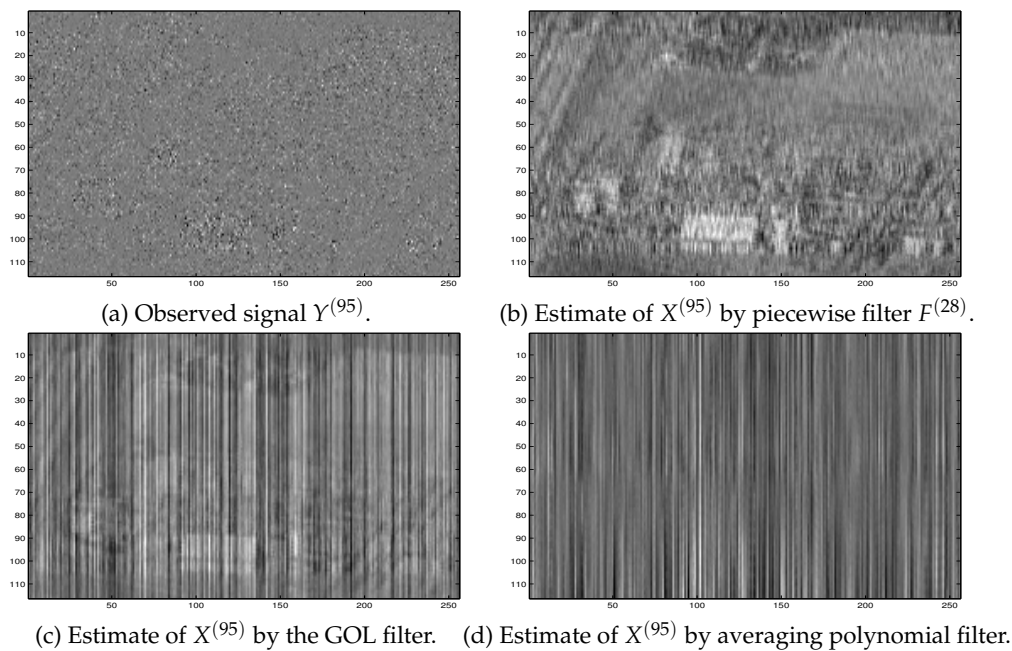


Figure 2. Examples of the observed signal and the estimates obtained by different filters.

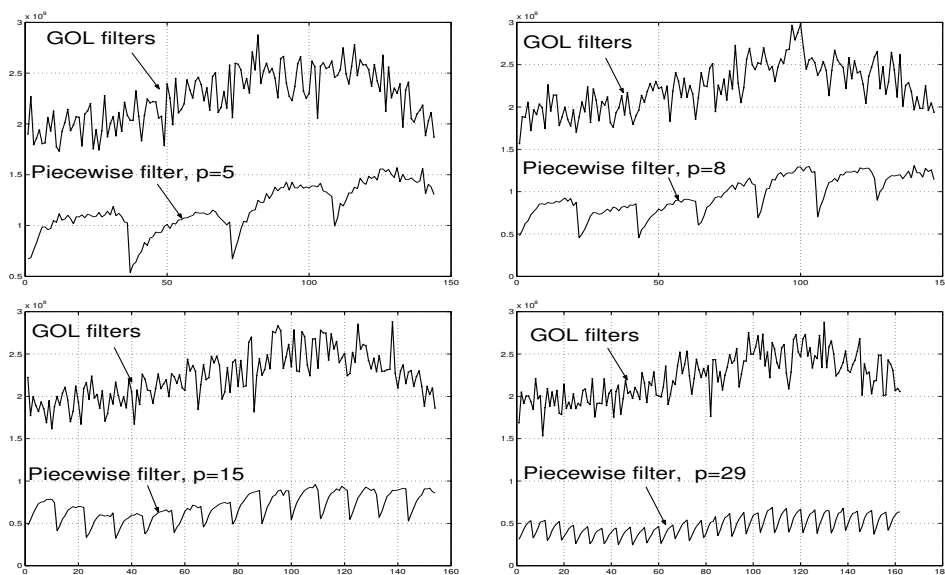


Figure 3. Illustration of the errors associated with the piecewise interpolation filters $F^{(p-1)}$ and the GOL filters [5] applied to signals described in Examples 1–4.

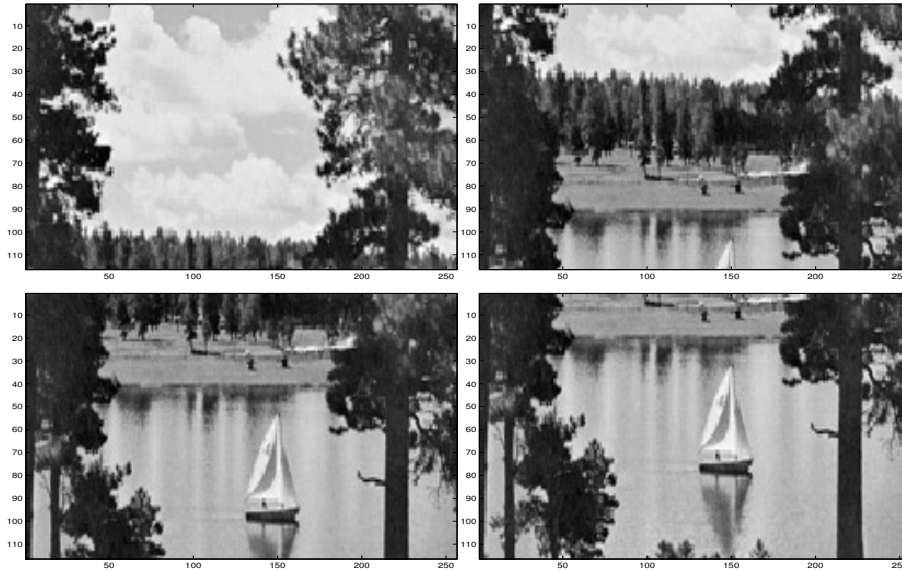


Figure 4. Examples of selected signals to be estimated from observed data considered in Example 5-9.

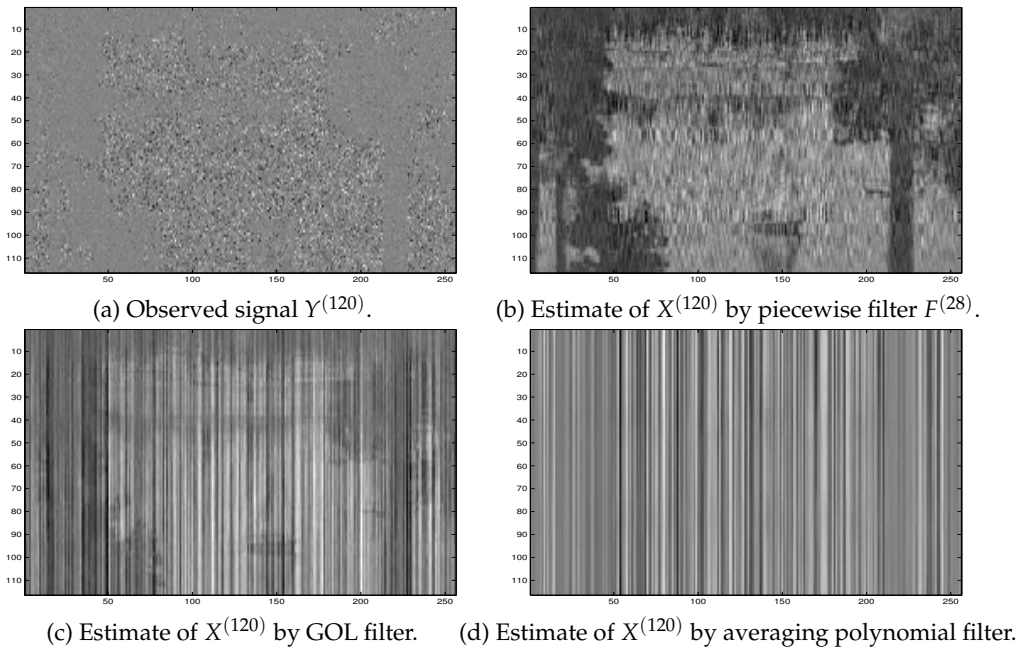


Figure 5. Examples of the observed signal and the estimates obtained by different filters.

5.5. Further Simulations with Different Type of Noise

In Examples 11 and 12 below, a different type of noise is considered. Unlike the multiplicative noise in (40) and (49), here, the noise is additive.

Example 11. First, the piecewise interpolation filter $F^{(28)}$, the GOL filters [5] and the averaging polynomial filter [12] have been applied to the observed signals given by

$$Y^{(k)} = X^{(k)} + 900 \times \text{randn}_{(k)}, \quad \text{for } k = 1, \dots, 141. \quad (50)$$

where $X^{(k)}$ is as in Section 5.2, i.e. $X^{(k)}$ is formed from the images 'plant'. In Figure 8 (a), the diagrams of the errors associated with filter $F^{(28)}$ and the GOL filters [5] are given. The errors associated with the averaging polynomial filter [12], $\{\epsilon_{k_W}\}_{k=1}^{141}$, are much greater (of order $\mathcal{O}(10^9)$) and they are not presented in Figure 8 (a).

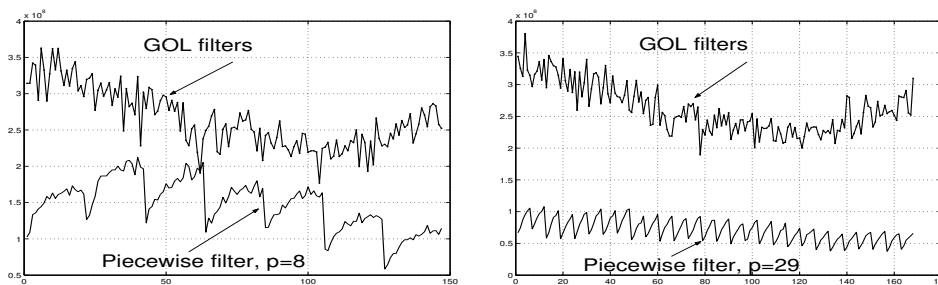


Figure 6. Illustration of errors associated with the piecewise interpolation filter $F^{(p-1)}$ of order p and the generic optimal linear (GOL) filters [5] applied to signals described in Examples 9 and 10.

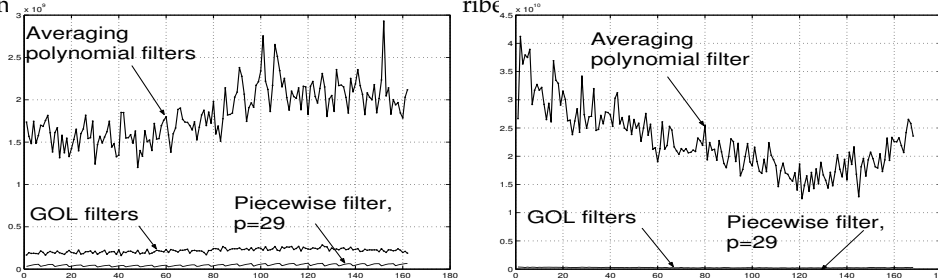


Figure 7. Illustration of errors associated with the averaging polynomial filters [10,12] in Examples 9 and 10.

Example 12. In this example, the reference signals $X^{(1)}, \dots, X^{(141)}$ are as those in Section 5.3, i.e. they are formed from the image ‘boat’. The observed signals are given by

$$Y^{(k)} = X^{(k)} + 1000 \times \text{randn}_{(k)}, \quad \text{for } k = 1, \dots, 141. \tag{51}$$

The piecewise interpolation filter $F^{(28)}$ and the GOL filters [5] estimate the reference signals with the associated errors represented in Figure 8 (b). As in Example 11 above, in this case, the errors associated with the averaging polynomial filter [10,12] are much greater (of order $\mathcal{O}(10^{10})$) and they are not presented in Figure 8 (b).

Examples 11 and 12 further demonstrate the advantages of the proposed piecewise interpolation filter.

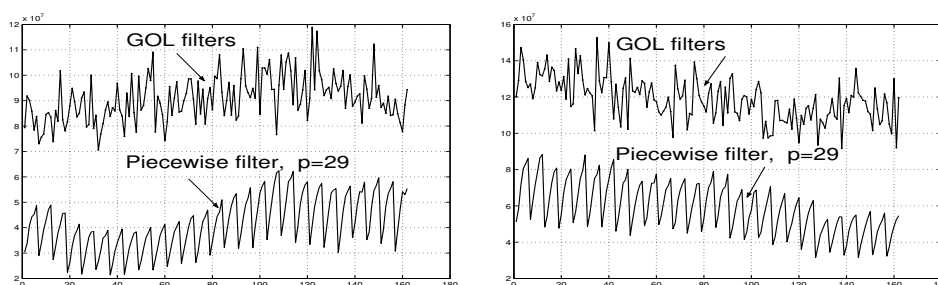


Figure 8. Illustration of errors associated with the piecewise interpolation filter $F^{(p-1)}$ and the generic optimal linear (GOL) filters [5] applied to signals described in Examples 11 and 12.

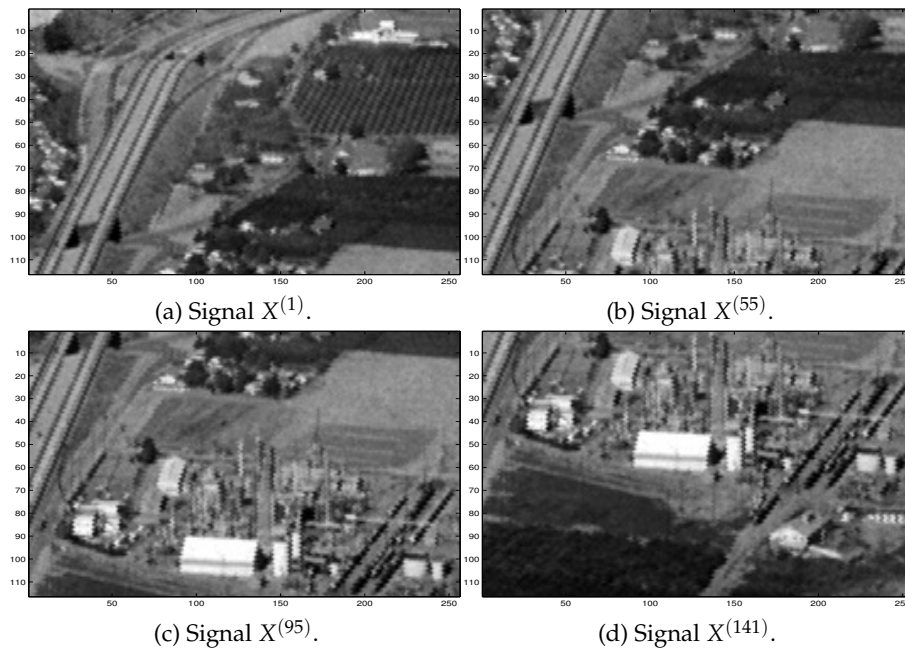
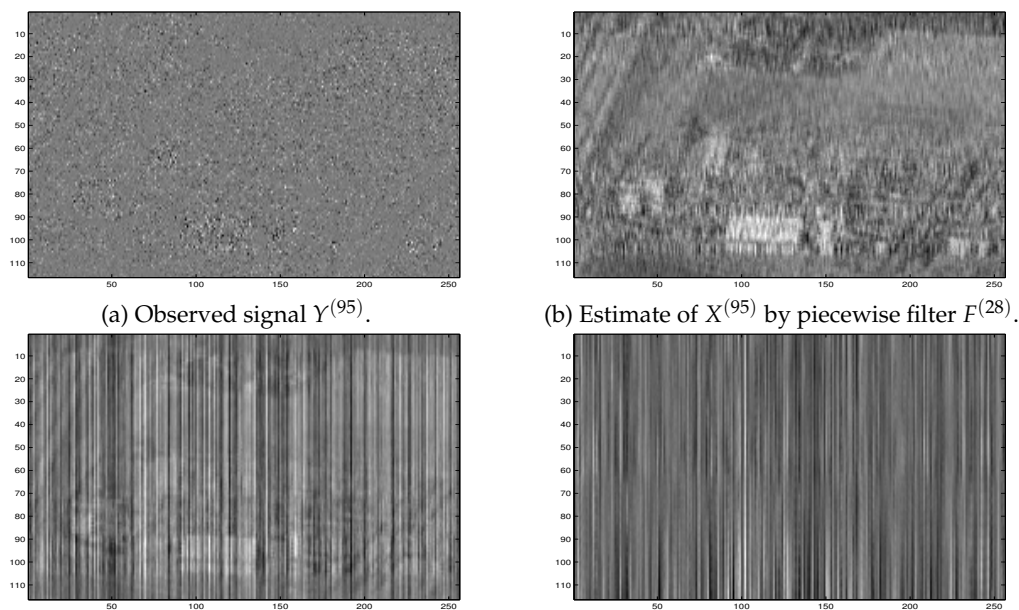


Figure 9. Examples of selected signals to be estimated from observed data.



(c) Estimate of $X^{(95)}$ by generic optimal linear filter. (d) Estimate of $X^{(95)}$ by averaging polynomial filter.

Figure 10. Examples of the observed signal and the estimates obtained by different filters.

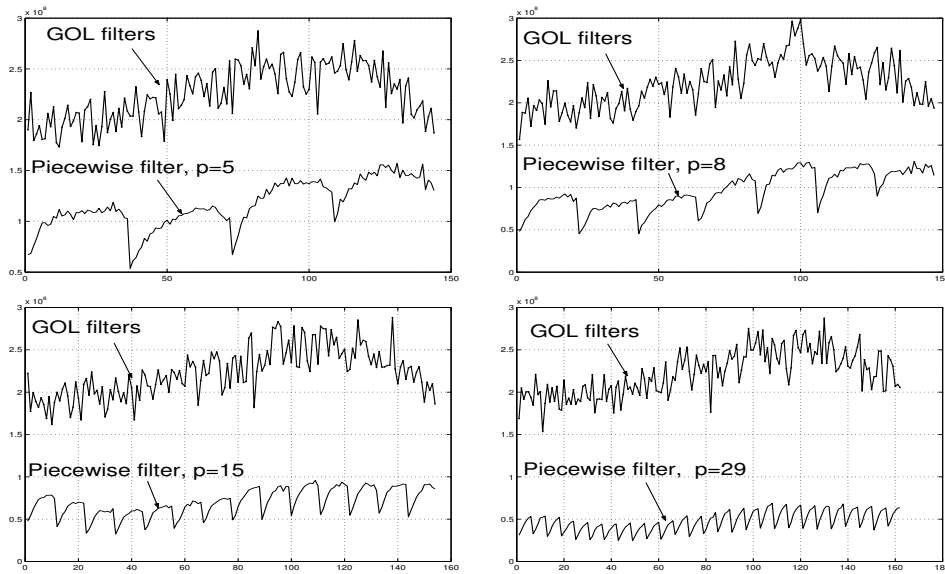


Figure 11. Illustration of the errors associated with the piecewise interpolation filters $F^{(p-1)}$ and the generic optimal linear (GOL) filters [5] applied to signals described in Examples 1–4.

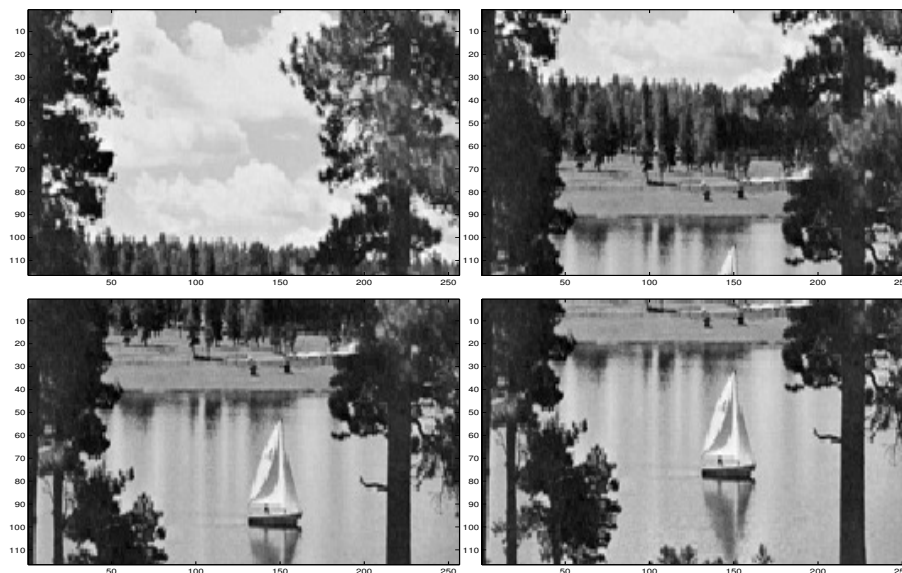


Figure 12. Examples of selected signals to be estimated from observed data considered in Example 5-9.

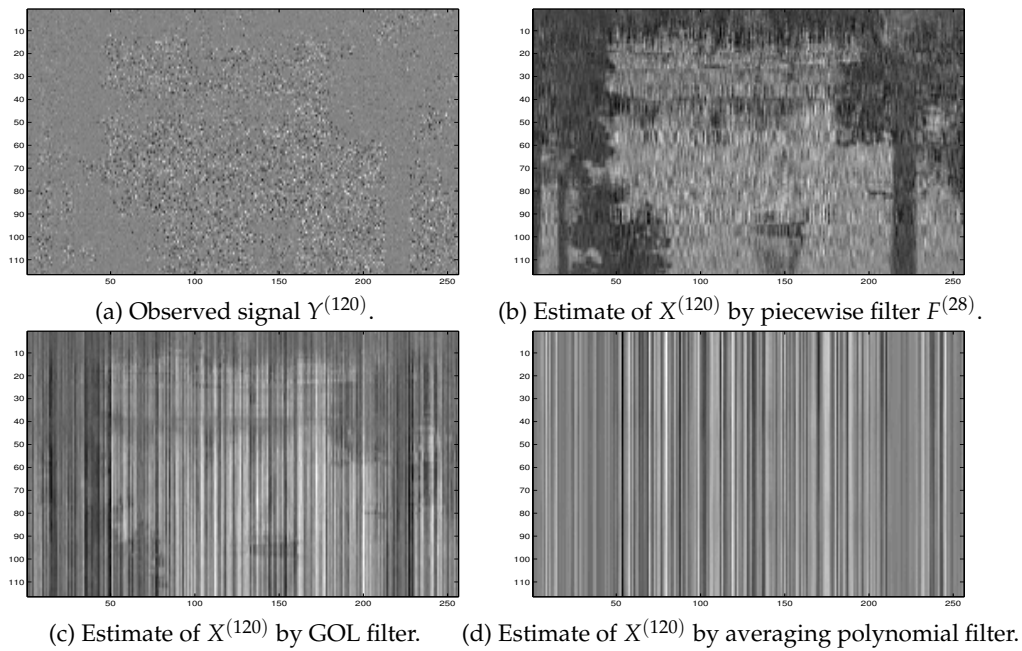


Figure 13. Examples of the observed signal and the estimates obtained by different filters.

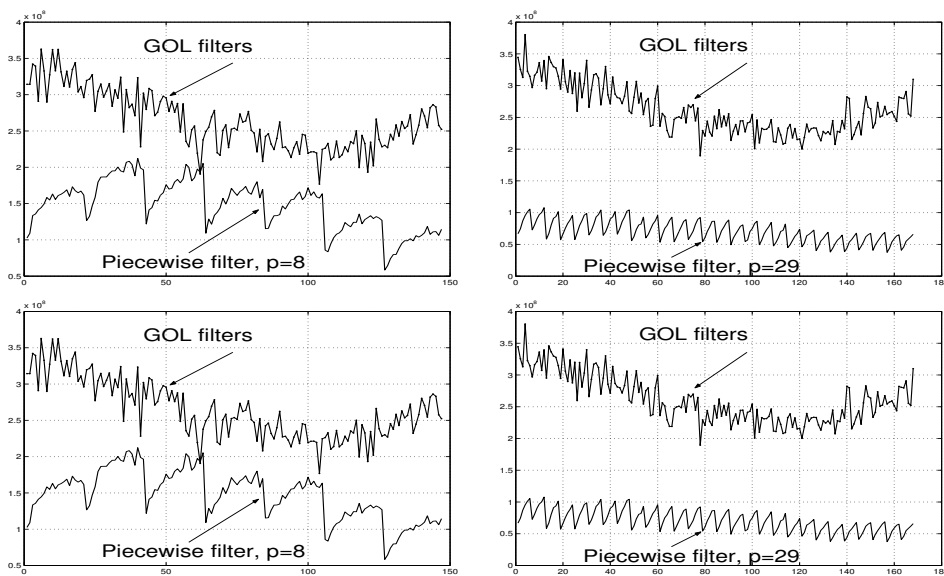


Figure 14. Illustration of errors associated with the piecewise interpolation filter $F^{(p-1)}$ of order p and the generic optimal linear (GOL) filters [5] applied to signals described in Examples 5–8.

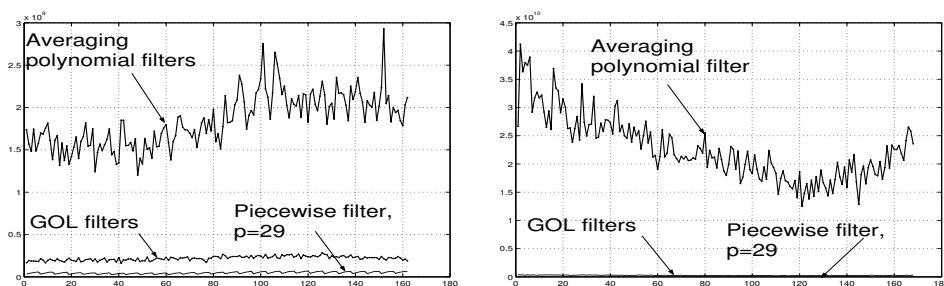


Figure 15. Illustration of errors associated with the averaging polynomial filters [10,12] in Examples 9 and 10.

5.6. Summary of Simulations

The above simulations confirm the theoretical results obtained in Theorems 1–3. In particular, Figs. 11 and 14 demonstrate that the error associated with the piecewise interpolation filter $F^{(p-1)}$ decreases when the number of sub-filters F_1, \dots, F_p , p , increases.

A comparison between the proposed filter $F^{(p-1)}$ and the known related filters [5,10,12,22,23] has been done. The filter $F^{(p-1)}$ estimates the reference signals with the accuracies that are much better than those of the generic optimal linear (GOL) filters [5] and the averaging polynomial filter [10,12]. Further, the filters proposed in [22,23] fail in processing the signals under consideration. This is because the observed signals in (40), (49), (50) and (51) are grossly corrupted and, therefore, the inverse matrices used in the filter structures in [23] do not exist. The technique in [22] requires the use of the reference signal in the proposed filter which is supposed to be unknown in the simulations above.

The filters have been applied to the different signal sets (presented in Sections 5.2 and 5.3), using different forms of noise (given in (40), (49), (50) and (51)).

computational work associated with the proposed filter $F^{(p-1)}$ is substantially less than that associated with the known filters discussed in Section 1 (in particular, with the filters in [11]–[23]). This is because, for the processing of a data set containing N signals, filter $F^{(p-1)}$ requires computation of p covariance matrices with $p \ll N$ while the known filters require computation of N matrices (in the above Examples, $p = 5, 7, 15, 28$, respectively, and $N = 141$).

6. Conclusions

The technique of the constructive approximation of a nonlinear operator $\mathcal{F} : K_Y \rightarrow K_X$ has been provided. Here, $K_X = \{\mathbf{x}(t, \cdot) \in L^2(\Omega, \mathbb{R}^m) \mid t \in T\}$ and $K_Y = \{\mathbf{y}(t, \cdot) \in L^2(\Omega, \mathbb{R}^n) \mid t \in T\}$ where $T := [a, b] \subseteq \mathbb{R}$ and $\Omega = \{\omega\}$ is the set of outcomes of a probability space. The device behind the proposed method is based on a special extension of the piecewise linear interpolation technique to the case of stochastic sets K_X and K_Y . The proposed methodology is motivated by the problems arising in signal processing where the nonlinear operator \mathcal{F} is interpreted as a nonlinear system (or a nonlinear filter) transforming a set of stochastic signals K_Y to the set of stochastic signals K_X . Therefore, the provided technique provides an effective way to transform large data sets.

Distinctive features of the approach are as follows.

(i) The proposed filter $\mathcal{F}^{(p-1)} : K_Y \rightarrow K_X$ is nonlinear and is presented in the form of a sum with $p - 1$ terms where each term, $\mathcal{F}_j : K_{Y,j} \rightarrow K_{X,j}$, is interpreted as a particular sub-filter. Here, $K_{Y,j}$ and $K_{X,j}$ are ‘small’ pieces of K_Y and K_X , respectively.

(ii) The prime idea is to exploit a prior information only on *few reference signals*, p , from the set K_X that contains $N \gg p$ signals (or even an infinite number of signals) and determine \mathcal{F}_j separately, for each pieces $K_{Y,j}$ and $K_{X,j}$, so that the associated error is minimal. In other words, the filter $\mathcal{F}^{(p-1)}$ is flexible to changes in the sets of observed and reference signals K_Y and K_X , respectively.

(iii) Due to the specific way of determining \mathcal{F}_j , the filter $\mathcal{F}^{(p-1)}$ provides a smaller associated error than that for the processing of the whole set K_Y by a filter which is not specifically adjusted to each particular piece $K_{Y,j}$. Moreover, the error associated with our filter decreases when the number of its terms, $\mathcal{F}_1, \dots, \mathcal{F}_{p-1}$, increases.

(iv) While the proposed filter $\mathcal{F}^{(p-1)}$ processes arbitrarily large (and even infinite) signal sets, the filter is nevertheless fixed for all signals in the sets.

(v) The filter $\mathcal{F}^{(p-1)}$ is determined in terms of pseudo-inverse matrices so that the filter always exists.

(vi) computational load associated with the filter $\mathcal{F}^{(p-1)}$ is less than that associated with other known filters applied to the processing of large signal sets.

Author Contributions: Conceptualization, methodology, writing original draft - A.T. Numerical simulations - A.T. and P.P. Algorithm - P.P. Matlab codes - P.P. English amelioration - P.P. The authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Proof of Theorem 1: It follows from (8) and (11) that α_j , for $j = 1, \dots, p - 1$, is given by

$$\alpha_j = \widehat{\mathbf{x}}(t_j, \omega) - B_j[\mathbf{y}(t_j, \omega)]. \quad (\text{A1})$$

Further, for α_j given by (A1),

$$\begin{aligned} & \|[\mathbf{x}(t_{j+1}, \cdot) - \alpha_j] - B_j[\mathbf{y}(t_{j+1}, \cdot)]\|_{\Omega}^2 \\ &= \|\mathbf{z}(t_j, t_{j+1}, \cdot) - B_j[\mathbf{w}(t_j, t_{j+1}, \cdot)]\|_{\Omega}^2 \end{aligned} \quad (\text{A2})$$

$$\begin{aligned} &= \text{tr}\{E_{z_j z_j} - E_{z_j w_j} B_j^T - B_j E_{w_j z_j} + B_j E_{w_j w_j} B_j^T\} \\ &= \|E_{z_j z_j}^{1/2}\|^2 - \|E_{z_j w_j} (E_{w_j w_j}^{1/2})^\dagger\|^2 + \|(B_j - E_{z_j w_j} E_{w_j w_j}^\dagger) E_{w_j w_j}^{1/2}\|^2 \\ &= \|E_{z_j z_j}^{1/2}\|^2 - \|E_{z_j w_j} (E_{w_j w_j}^{1/2})^\dagger\|^2 + \|E_{z_j w_j} (E_{w_j w_j}^{1/2})^\dagger - B_j E_{w_j w_j}^{1/2}\|^2, \end{aligned} \quad (\text{A3})$$

where $\|\cdot\|$ is the Frobenius norm. The latter is true because

$$E_{w_j w_j}^\dagger E_{w_j w_j}^{1/2} = (E_{w_j w_j}^{1/2})^\dagger$$

and

$$E_{z_j w_j} E_{w_j w_j}^\dagger E_{w_j w_j} = E_{z_j w_j} \quad (\text{A4})$$

by Lemma 24 in [5]. Thus, the second expression in (11) is reduced to the problem

$$\min_{B_j} \|E_{z_j w_j} (E_{w_j w_j}^{1/2})^\dagger - B_j E_{w_j w_j}^{1/2}\|^2. \quad (\text{A5})$$

It is known (see, for example, [5], p. 304) that the solution of problem (A5) is given by (19). The equation (17) follows from (6) and (A1).

Theorem 1 is proven. \square

Proof of Theorem 2: For $t \in [t_j, t_{j+1}]$ and F_j defined by (17)–(19),

$$\begin{aligned} & \mathbf{x}(t, \omega) - F[\mathbf{y}(t, \omega)] \\ &= \mathbf{x}(t, \omega) - F_j[\mathbf{y}(t, \omega)] \\ &= \mathbf{x}(t, \omega) - \widehat{\mathbf{x}}(t_j, \omega) + B_j \mathbf{y}(t_j, \omega) - B_j \mathbf{y}(t, \omega) \\ &= [\mathbf{x}(t, \omega) - \mathbf{x}(t_{j+1}, \omega)] + \mathbf{z}(t_j, t_{j+1}, \omega) - B_j \mathbf{w}(t_j, t_{j+1}, \omega) \\ & \quad + B_j[\mathbf{y}(t_{j+1}, \omega) - \mathbf{y}(t, \omega)]. \end{aligned} \quad (\text{A6})$$

Then (A6) implies

$$\begin{aligned} \|\mathbf{x}(t, \omega) - F[\mathbf{y}(t, \omega)]\|_{T, \Omega}^2 &\leq \|\mathbf{x}(t, \omega) - \mathbf{x}(t_{j+1}, \omega)\|_{T, \Omega}^2 \\ & \quad + \|\mathbf{z}(t_j, t_{j+1}, \omega) - B_j \mathbf{w}(t_j, t_{j+1}, \omega)\|_{\Omega}^2 \\ & \quad + \|B_j[\mathbf{y}(t_{j+1}, \omega) - \mathbf{y}(t, \omega)]\|_{T, \Omega}^2 \end{aligned} \quad (\text{A7})$$

where $\|\mathbf{z}(t_j, t_{j+1}, \omega) - B_j \mathbf{w}(t_j, t_{j+1}, \omega)\|_{\Omega}^2 = \|\mathbf{z}(t_j, t_{j+1}, \omega) - B_j \mathbf{w}(t_j, t_{j+1}, \omega)\|_{T, \Omega}^2$.

It follows from (A2) and (A3) that for B_j given by (19),

$$\|\mathbf{z}(t_j, t_{j+1}, \omega) - B_j \mathbf{w}(t_j, t_{j+1}, \omega)\|_{\Omega}^2 = \|E_{z_j z_j}^{1/2}\|^2 - \|E_{z_j w_j}(E_{w_j w_j}^{1/2})^\dagger\|^2. \quad (\text{A8})$$

Then (16)–(19), (23) and (A6)–(A8) imply that for all $t \in [a, b]$ and $\omega \in \Omega$, (24) is true. \square

Proof of Theorem 3: The relation (22) implies that

$$\|\mathbf{x}(t, \omega) - F[\mathbf{y}(t, \omega)]\|_{T, \Omega}^2 = \frac{1}{b-a} \sum_{j=1}^{p-1} \int_{t_j}^{t_{j+1}} \|\mathbf{x}(t, \omega) - F_j[\mathbf{y}(t, \omega)]\|_{\Omega}^2 dt, \quad (\text{A9})$$

where

$$\begin{aligned} & \|\mathbf{x}(t, \omega) - F_j[\mathbf{y}(t, \omega)]\|_{\Omega}^2 \\ &= \|\mathbf{x}(t, \omega) - \widehat{\mathbf{x}}(t_j, \omega) + B_j[\mathbf{y}(t_j, \omega) - B_j \mathbf{y}(t, \omega)]\|_{\Omega}^2 \\ &\leq \|\mathbf{x}(t, \omega) - \mathbf{x}(t_j, \omega)\|_{\Omega}^2 + \|\mathbf{x}(t_j, \omega) - \widehat{\mathbf{x}}(t_j, \omega)\|_{\Omega}^2 \\ &\quad + \|B_j[\mathbf{y}(t_j, \omega) - B_j \mathbf{y}(t, \omega)]\|_{\Omega}^2. \end{aligned}$$

Then

$$\int_{t_j}^{t_{j+1}} \|\mathbf{x}(t, \omega) - F_j[\mathbf{y}(t, \omega)]\|_{\Omega}^2 dt \quad (\text{A10})$$

$$\begin{aligned} &\leq \int_{t_j}^{t_{j+1}} \|\mathbf{x}(t, \omega) - \mathbf{x}(t_j, \omega)\|_{\Omega}^2 dt + \int_{t_j}^{t_{j+1}} \|\mathbf{x}(t_j, \omega) - \widehat{\mathbf{x}}(t_j, \omega)\|_{\Omega}^2 dt \\ &\quad + \|B_j\| \int_{t_j}^{t_{j+1}} \|\mathbf{y}(t_j, \omega) - \mathbf{y}(t, \omega)\|_{\Omega}^2 dt \\ &\leq \lambda_j (\Delta t_j)^2 + \|\mathbf{x}(t_j, \omega) - \widehat{\mathbf{x}}(t_j, \omega)\|_{\Omega}^2 \Delta t_j + \|B_j\| \gamma_j (\Delta t_j)^2 \quad (\text{A11}) \end{aligned}$$

Let us consider an estimate of $\|\mathbf{x}(t_j, \omega) - \widehat{\mathbf{x}}(t_j, \omega)\|_{\Omega}^2$, for $j = 1, \dots, p-1$. To this end, let us denote $\Delta t = \max_{j=1, \dots, p-1} \Delta t_j$.

For $j = 1$, i.e. for $t \in [t_1, t_2]$,

$$\begin{aligned} & \|\mathbf{x}(t, \omega) - F_1 \mathbf{y}(t, \omega)\|_{\Omega}^2 \\ &\leq \|\mathbf{x}(t, \omega) - \mathbf{x}(t_1, \omega)\|_{\Omega}^2 + \|\mathbf{x}(t_1, \omega) - \widehat{\mathbf{x}}(t_1, \omega)\|_{\Omega}^2 + \|B_1\| \|\mathbf{y}(t_1, \omega) - \mathbf{y}(t, \omega)\|_{\Omega}^2 \\ &\leq \lambda_1 \Delta t_1 + c_1 \Delta t_1 + \|B_1\| \gamma_1 \Delta t_1 \\ &\leq \beta_1 \Delta t, \end{aligned}$$

where $\beta_1 = \lambda_1 + c_1 + \|B_1\| \gamma_1$. In particular, the latter implies

$$\|\mathbf{x}(t_2, \omega) - \widehat{\mathbf{x}}(t_2, \omega)\|_{\Omega}^2 = \|\mathbf{x}(t_2, \omega) - F_1 \mathbf{y}(t_2, \omega)\|_{\Omega}^2 \leq \beta_1 \Delta t$$

For $j = 2$, i.e. for $t \in [t_2, t_3]$,

$$\begin{aligned} & \|\mathbf{x}(t, \omega) - F_2 \mathbf{y}(t, \omega)\|_{\Omega}^2 \\ &\leq \|\mathbf{x}(t, \omega) - \mathbf{x}(t_2, \omega)\|_{\Omega}^2 + \|\mathbf{x}(t_2, \omega) - \widehat{\mathbf{x}}(t_2, \omega)\|_{\Omega}^2 + \|B_2\| \|\mathbf{y}(t_2, \omega) - \mathbf{y}(t, \omega)\|_{\Omega}^2 \\ &\leq \lambda_2 \Delta t_2 + \beta_1 \Delta t + \|B_2\| \gamma_2 \Delta t_2 \\ &\leq \beta_2 \Delta t, \end{aligned}$$

where $\beta_2 = \lambda_2 + \beta_1 + \|B_2\| \gamma_2$. In particular, then it follows that

$$\|\mathbf{x}(t_3, \omega) - \widehat{\mathbf{x}}(t_3, \omega)\|_{\Omega}^2 = \|\mathbf{x}(t_3, \omega) - F_2 \mathbf{y}(t_3, \omega)\|_{\Omega}^2 \leq \beta_2 \Delta t.$$

On the above basis, let us assume that, for $j = k - 1$ with $k = 2, \dots, p - 1$, i.e. for $t \in [t_{k-1}, t_k]$,

$$\|\mathbf{x}(t_k, \omega) - \widehat{\mathbf{x}}(t_k, \omega)\|_{\Omega}^2 = \|\mathbf{x}(t_k, \omega) - F_{k-1}\mathbf{y}(t_k, \omega)\|_{\Omega}^2 \leq \beta_{k-1}\Delta t$$

where β_{k-1} is defined by analogy with β_2 .

Then, for $j = k$ with $k = 2, \dots, p - 1$, i.e. for $t \in [t_k, t_{k+1}]$,

$$\begin{aligned} & \|\mathbf{x}(t, \omega) - F_k\mathbf{y}(t, \omega)\|_{\Omega}^2 \\ & \leq \|\mathbf{x}(t, \omega) - \mathbf{x}(t_k, \omega)\|_{\Omega}^2 + \|\mathbf{x}(t_k, \omega) - \widehat{\mathbf{x}}(t_k, \omega)\|_{\Omega}^2 + \|B_k\| \|\mathbf{y}(t_k, \omega) - \mathbf{y}(t, \omega)\|_{\Omega}^2 \\ & \leq \lambda_k \Delta t_k + \beta_{k-1} \Delta t + \|B_k\| \gamma_2 \Delta t_k \\ & \leq \beta_k \Delta t, \end{aligned}$$

where $\beta_k = \lambda_k + \beta_{k-1} + \|B_k\| \gamma_k$. Thus, the following is true:

$$\|\mathbf{x}(t_{k+1}, \omega) - \widehat{\mathbf{x}}(t_{k+1}, \omega)\|_{\Omega}^2 = \|\mathbf{x}(t_{k+1}, \omega) - F_k\mathbf{y}(t_{k+1}, \omega)\|_{\Omega}^2 \leq \beta_k \Delta t. \quad (\text{A12})$$

Therefore, (A10), (A11) and (A12) imply

$$\begin{aligned} & \int_{t_j}^{t_{j+1}} \|\mathbf{x}(t, \omega) - F_j[\mathbf{y}(t, \omega)]\|_{\Omega}^2 dt \\ & \leq \lambda_j (\Delta t_j)^2 + \beta_{j-1} (\Delta t_j)^2 + \|B_j\| \gamma_j (\Delta t_j)^2 \\ & \leq \eta_j (\Delta t)^2 \end{aligned} \quad (\text{A13})$$

where $\eta_j = \lambda_j + \beta_{j-1} + \|B_j\| \gamma_j$, and then it follows from (A9)–(A11) and (A13) that for all $t \in [a, b]$,

$$\|\mathbf{x}(t, \omega) - F[\mathbf{y}(t, \omega)]\|_{T, \Omega}^2 \leq \frac{1}{b-a} \sum_{j=1}^{p-1} \eta_j (\Delta t)^2 = \frac{1}{b-a} \Delta t \sum_{j=1}^{p-1} \eta_j \Delta t. \quad (\text{A14})$$

Let us now choose $c \in \mathbb{R}$ and $d \in \mathbb{R}$ so that $\Delta t = \frac{d-c}{p}$ and partition interval $[c, d] \subset \mathbb{R}$ by points τ_1, \dots, τ_p so that $c = \tau_1$ and $\tau_j = \tau_1 + j\Delta t$ with $j = 1, \dots, p$. There exists an integrable (bounded) function $\varphi : [c, d] \rightarrow \mathbb{R}$ such that, for $\xi_j \in (\tau_j, \tau_{j+1})$, $\varphi(\xi_j) = \eta_j$. Then

$$\lim_{\Delta t \rightarrow \infty} \sum_{j=1}^{p-1} \eta_j \Delta t = \lim_{\Delta t \rightarrow \infty} \sum_{j=1}^{p-1} \varphi(\xi_j) \Delta t = \int_c^d \varphi(\tau) d\tau < +\infty. \quad (\text{A15})$$

Thus,

$$\frac{1}{b-a} \Delta t \sum_{j=1}^{p-1} \eta_j \Delta t \rightarrow 0 \quad \text{as } \Delta t \rightarrow 0. \quad (\text{A16})$$

As a result, (A14)–(A16) imply (26). \square

References

1. J. Chen, J. Benesty, Y. Huang, and S. Doclo, New Insights Into the Noise Reduction Wiener Filter, *IEEE Trans. on Audio, Speech, and Language Processing*, 14, No. 4, pp. 1218 - 1234, 2006.
2. M. Spurbeck and P. Schreier, Causal Wiener filter banks for periodically correlated time series, *Signal Processing*, 87, 6, pp. 1179-1187, 2007.
3. J. S. Goldstein, I. Reed, and L. L. Scharf, "A Multistage Representation of the Wiener Filter Based on Orthogonal Projections," *IEEE Trans. on Information Theory*, vol. 44, pp. 2943-2959, 1998.
4. Y. Hua, M. Nikpour, and P. Stoica, "Optimal Reduced-Rank estimation and filtering," *IEEE Trans. on Signal Processing*, vol. 49, pp. 457-469, 2001.

5. A. Torokhti and P. Howlett, *Computational Methods for Modelling of Nonlinear Systems*, Elsevier, 2007.
6. E. D. Sontag, *Polynomial Response Maps*, Lecture Notes in Control and Information Sciences, 13, 1979.
7. S. Chen and S. A. Billings, Representation of non-linear systems: NARMAX model, *Int. J. Control*, vol. 49, no. 3, pp. 1013–1032, 1989.
8. V. J. Mathews and G. L. Sicuranza, *Polynomial Signal Processing*, J. Wiley & Sons, 2001.
9. 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, No. 4, pp. 1431-1444, 2006.
10. A. Torokhti and P. Howlett, Filtering and Compression for Infinite Sets of Stochastic Signals, *Signal Processing*, 89, pp. 291-304, 2009.
11. J. Vesma and T. Saramaki, Polynomial-Based Interpolation Filters - Part I: Filter Synthesis, *Circuits, Systems, and Signal Processing*, Volume 26, Number 2, Pages 115-146, 2007.
12. A. Torokhti and J. Manton, Generic Weighted Filtering of Stochastic Signals, *IEEE Trans. on Signal Processing*, 57, issue 12, pp. 4675-4685, 2009.
13. A. Torokhti and S. Miklavcic, Data Compression under Constraints of Causality and Variable Finite Memory, *Signal Processing*, 90, Issue 10, pp. 2822-2834, 2010.
14. I. Babuska, U. Banerjee, J. E. Osborn, Generalized finite element methods: main ideas, results, and perspective, *International Journal of Computational Methods*, 1 (1), pp. 67-103, 2004.
15. S. Kang and L. Chua, A global representation of multidimensional piecewise-linear functions with linear partitions, *IEEE Trans. on Circuits and Systems*, 25 Issue:11, pp. 938 - 940, 1978.
16. L.O. Chua and A.-C. Deng, Canonical piecewise-linear representation, *IEEE Trans. on Circuits and Systems*, 35 Issue:1, pp. 101 - 111, 1988.
17. J.-N. Lin and R. Unbehauen, Adaptive nonlinear digital filter with canonical piecewise-linear structure, *IEEE Trans. on Circuits and Systems*, 37 Issue:3, pp. 347 - 353, 1990.
18. J.-N. Lin and R. Unbehauen, Canonical piecewise-linear approximations, *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, 39 Issue:8, pp. 697 - 699, 1992.
19. S.B. Gelfand and C.S. Ravishankar, A tree-structured piecewise linear adaptive filter, *IEEE Trans. on Inf. Theory*, 39, issue 6, pp. 1907–1922, 1993.
20. E.A. Heredia and G.R. Arce, Piecewise linear system modeling based on a continuous threshold decomposition, *IEEE Trans. on Signal Processing*, 44 Issue:6, pp. 1440 - 1453, 1996.
21. G. Feng, Robust filtering design of piecewise discrete time linear systems, *IEEE Trans. on Signal Processing*, 53 Issue:2, pp. 599 - 605, 2005.
22. F. Russo, Technique for image denoising based on adaptive piecewise linear filters and automatic parameter tuning, *IEEE Trans. on Instrumentation and Measurement*, 55, Issue:4, pp. 1362 - 1367, 2006.
23. J.E. Cousseau, J.L. Figueroa, S. Werner, T.I. Laakso, Efficient Nonlinear Wiener Model Identification Using a Complex-Valued Simplicial Canonical Piecewise Linear Filter, *IEEE Trans. on Signal Processing*, 55 Issue:5, pp. 1780 - 1792, 2007.
24. P. Julian, A. Desages, B. D'Amico, Orthonormal high-level canonical PWL functions with applications to model reduction, *IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications*, 47 Issue:5, pp. 702 - 712, 2000.
25. T. Wigren, Recursive Prediction Error Identification Using the Nonlinear Wiener Model, *Automatica*, 29, 4, pp. 1011–1025, 1993.
26. G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1996.
27. T. Anderson, *An Introduction to Multivariate Statistical Analysis*, New York, Wiley, 1984.
28. 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.
29. O. Ledoit and M. Wolf, A well-conditioned estimator for large-dimensional covariance matrices, *J. Multivariate Analysis* 88, pp. 365–411, 2004.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.