PADOVA UNIVERSITY PRESS

Dolomites Research Notes on Approximation

Volume 12 · 2019 · Pages 73–85

An algorithm for model-based denoising of input-output data

Marta Gatto [a] · Fabio Marcuzzi [a]

Communicated by M. Vianello

Abstract

It is well known that experimental measurements contain noise that corrupts the (usually deterministic) real quantities to be measured. For this reason, many denoising algorithms have been proposed and analyzed in the literature. In this paper we focus on the case in which, given a physical dynamical system, its input/output variables are measured and these measurements are corrupted by a considerable amount of Gaussian noise, whose means and variances are supposed to be unknown. In this case, the noise not only corrupts each single piece of data but also the causal (input-output) relation between the data. The aim of this work is to use a deterministic model of the physical system, that we suppose described by linear differential equations, inside an iterative denoising algorithm, based on global optimization and smoothing, that allows also to estimate the means and covariances of the noise. Some numerical results will be shown.

1 Introduction

It is well known that experimental measurements usually contain noise that corrupts the (usually deterministic) real quantities to be measured. Many denoising algorithms have been proposed and analyzed in the literature, and differ from each other for the hypothesis on signal and noise properties and for the prior knowledge required about them [6]. Comparisons on these methods have been carried out, for example in [26] for signal denoising and [28] for image denoising, and different kind of noisy data and functions for the measure of denoising results have been considered. Some of the most important techniques are: moving average filters, linear filters, nonlinear filters, Fourier decomposition, Wavelet decomposition, Total Variation Minimization, Neural Networks (e.g. [8]).

Denoising methods are mainly studied for individual signals, isolated from the system from which they arise. When more variables are measured from the same physical dynamical system, the noise not only corrupts each single quantity but also the causal (input-output) relation between the data. Therefore, applying the methods previously mentioned separately to each noisy signal is not sufficient to recover the correct input-output relations: this paper aims at introducing a new algorithm that satisfies this requirement. In particular, we will consider the simultaneous denoising of input-output signals from a physical dynamical system, that we suppose can be modeled by a discrete linear time-invariant (DLTI) system. Taking into account the system relations we will obtain a linear denoising method that we therefore classify as "*DLTI model-based denoising*".

The availability of data with a precise input-output relation is of fundamental importance in real-life applications, e.g. when quantities of interest derived from input/output variables (like the mechanical/electric power) are required, or in the resolution of computational inverse problems. In such a context the use of a mathematical model of the system allows to recover unknown (and not measurable) quantities, for example model parameters and boundary conditions for distributed parameters systems [9], [29]. In particular, good input-output algebraic relations in the data are required to estimate continuous system parameters from discrete measures [2], [30].

The term *Model-Based* in the image and signal processing literature, is often used to refer not only to physical models, but also to (abstract) mathematical model structures. This is the case of Candy, who develops in [6] a general model-based approach to signal processing for different kinds of problems, using not only *physical-based* models but also "black-box" ones.

In this work we focus on model-based denoising where models arise from the physical equations that describe the system. Studies in this direction are, for example [40], where a "Physically Consistent Denoising" is described and an algorithm for the denoising and "missing data recovery" of 2D physically plausible vector fields is introduced, and [34], where a Model-Based Image Reconstruction that satisfies the physical Cahn-Hilliard equation is presented.

In this paper, we will consider signals corrupted by additive noise, also called in literature "*Error-in-variables (EIV) framework*" [13]. The studies in [10] and [31], [32], where the "*noisy I/O problem*" is introduced, are the nearest to the approach of this work. In this setting, the matrices of a DLTI system are supposed known, and the aim is the denoising of input-output data. The method is based on the a-priori knowledge of mean and covariances of the noise signals added to the inputs and the outputs. Unlike this approach, we are interested in the more general case in which these quantities are unknown, as it happens in many applications [13].

In the literature, a more general topic has also been investigated, in the case of additive noise with known properties: the denoising of input-output data and simultaneous identification of the DLTI system that links them. This problem shows a

[a] Department of Mathematics "Tullio Levi Civita", University of Padova, Via Trieste 63, 35131 Padova, Italy

certain affinity with the Total Least Squares (TLS) problem, see e.g. [37], [36], [39]. Moreover, the problem of denoising of the input-output data used for the identification of the model that relates them, has been studied with optimization methods, more precisely the minimization of the nuclear-norm of a matrix generated by data and model predictions, see e.g. [25].

The approach presented in this paper differs from the ones in the literature because it assumes a DLTI physical model to be available and aims at denoising the input-output signals, of which only noisy measurements are available with unknown values of means and variances, i.e. we generalize the problem of [31] and [32] in the case in which the statistical properties of the input and output noises are unknown. This comes at the price of a bigger computational effort that makes its applicability to real-time not straightforward, since we use global optimization and smoothing.

It is worth noticing that this generalization is important in applications; for example, when these properties are constant but good estimates of them are not available, or when the sensors (or estimators) generate noises with non constant properties, which variability cannot be described a-priori by a model.

We will show that in these cases the approach of [31] is not optimal and we will see in more details which are the additional conditions required when the estimates of noise means and covariances are unknown. In this work we propose a method based on the resolution of a linear system generated by the model equations (that we want to be satisfied precisely), and four other parameterized regularization terms (that are not required to be satisfied exactly). Hence, the problem is reduced to the choice of the regularization parameters of a multi-parametric linear system. With this approach, several methods have been presented in the literature and are divided in two categories: methods that use the noise variance value and, viceversa, *heuristic* methods also called *noise level free rules*. If the noise covariance is known (at least approximatively), the parameters can be determined with the discrepancy principle, and in the multi-parameter case with its generalization, introducing the *discrepancy hypersurface* ([12], [27]). The optimal parameters are found on that hypersurface through the optimization of a function of the parameters, for example the norm of the solution can be maximized [12] or the quasi-optimality criterion (introduced in [38]) can be considered [11].

On the other side, *heuristic* methods possess bad convergence properties. In fact, they don't converge in the "worst case scenario", i.e. it is not true that the regularized solution converges to the true one for every noise realization with noise level that tends to zero. Although this result, called *Bakushinskii veto* [1], convergence results have been demonstrated under appropriate conditions, that are usually satisfied in real situations (see [23] and references therein). Among heuristic methods for multi-parameter regularization, we can find the generalizations of the L-curve [3], of the Generalized Cross Validation (GCV) [5], a balancing principle [19] and parameter learning for denoising [24], [18].

Since we suppose the noise means and variances to be unknown, we rely upon an additional criterion based on other statistical properties of the noise. More precisely, we will use the Normalized Cumulative Periodogram (NCP), also known as Bartlett test, to measure the whiteness of the estimated noises, a well known method for the one-parameter regularization choice ([15], [16]).

The organization of the paper is as follows. In section 2 we will define the problem and the proposed approach to solve it; in Section 3 some general issues are discussed. In Section 4 an algorithm for the selection of parameters will be introduced and in Section 5 some numerical results will be shown. Concluding remarks and possible future work are provided in Section 6.

2 DLTI Model-Based denoising

Consider a linear dynamical system with deterministic and measurable input and output, described by a discrete linear time-invariant (DLTI) system [20] in state-space form:

$$\begin{cases} x(k+1) &= A\,x(k) + B\,u(k) \\ y(k) &= C\,x(k) + D\,u(k). \end{cases} \quad \text{per } k = 0,\dots,N-1, \tag{1}$$

where $x(k) \in \mathbb{R}^{n_x \times 1}$, $u(k) \in \mathbb{R}^{n_u \times 1}$, $y(k) \in \mathbb{R}^{n_y \times 1}$, and A, B, C, D are the system matrices, $A \in \mathbb{R}^{n_x \times n_x}$, $B \in \mathbb{R}^{n_x \times n_u}$, $C \in \mathbb{R}^{n_y \times n_x}$, $D \in \mathbb{R}^{n_y \times n_u}$.

The following definitions will be useful

$$\begin{aligned} u &:= [u^T(0) \cdots u^T(N-1)]^T && \in \mathbb{R}^{Nn_u \times 1} \\ y &:= [y^T(0) \cdots y^T(N-1)]^T && \in \mathbb{R}^{Nn_y \times 1} \\ x &:= [x^T(0) \cdots x^T(N)]^T && \in \mathbb{R}^{(N+1)n_x \times 1} \end{aligned} \tag{2}$$

and the superscript notation will be used to denote the components of the signals, i.e. for example $u(k) = [u^1(k), \dots, u^{n_u}(k)]^T$ and $u^i = [u^i(0) \cdots u^i(N-1)]^T$.

In this work we will consider the common choice of $D = 0_{n_y \times n_u}$ and a more restrictive, but often acceptable, condition that the matrix C be invertible so that the system reduces to

$$y(k+1) = CAC^{-1}\,y(k) + CB\,u(k), \qquad k = 0,\dots,N-2. \tag{3}$$

We suppose to know only noisy measurements of input and output signals u_e, y_e (defined in the same way as (2)), corrupted with white noise

$$\begin{cases} u_e(k) = u(k) + e_u(k) \\ y_e(k) = y(k) + e_y(k) \end{cases} \tag{4}$$

where the error vectors relative to each component of input and output signals e_{u^i}, e_{y^i} are white noise vectors that satisfy

$$\begin{cases} \mathbb{E}\{e_{u^i}\} = 0 \\ \mathbb{E}\{e_{u^i} e_{u^i}^T\} = \sigma_{eu^i}^2 I \end{cases} \quad \text{for } i = 1,\dots,n_u, \qquad \begin{cases} \mathbb{E}\{e_{y^i}\} = 0 \\ \mathbb{E}\{e_{y^i} e_{y^i}^T\} = \sigma_{ey^i}^2 I \end{cases} \quad \text{for } i = 1,\dots,n_y. \tag{5}$$

In the following we will consider also a more general hypothesis in which e_u and e_y are biased noises with mean values different from zero, i.e. we will consider Gaussian noises with biases $\mu_{eu^i} \neq 0, \mu_{ey^i} \neq 0$, $\forall i$:

$$\begin{cases} \mathbb{E}\{e_{u^i}\} = \mu_{eu^i} \\ \mathbb{E}\{e_{u^i} e_{u^i}^T\} = \sigma_{eu^i}^2 I \end{cases} \quad \text{for } i = 1, \dots, n_u, \qquad \begin{cases} \mathbb{E}\{e_{y^i}\} = \mu_{ey^i} \\ \mathbb{E}\{e_{y^i} e_{y^i}^T\} = \sigma_{ey^i}^2 I \end{cases} \quad \text{for } i = 1, \dots, n_y. \tag{6}$$

We observe that the noisy measurements u_e, y_e do not satisfy the model constraints. Hence, any other quantity calculated from them will be corrupted, even nonlinearly, by input/output noise. This fact motivates the need of an additional requirement for the denoising algorithm: it must be model-based, i.e. the denoised data must satisfy the deterministic model.

Our approach for the model-based denoising problem is the following: we impose the model constraints, together with other additional conditions on input-output data, as described in the following subsections.

2.1 Problem formulation

2.1.1 Model-based constraining

We define the fundamental constraint of model-based denoising in the following problem, that by itself is ill-posed and has infinitely many solutions, as will be shown later:

Problem 1 ("model-based constraining"). Given the vectors $y_e(k), u_e(k)$ with $k = 0, 1, \dots, N-1$, measures of signals with white (5) or Gaussian (6), additive noise (4), determine vectors $\hat{y}(k), \hat{u}(k)$ that satisfy the model (3).

Following [32], we can impose model equations (3) writing them in matrix form

$$\begin{bmatrix} \bar{A} & \bar{B} \end{bmatrix} \begin{bmatrix} y \\ u \end{bmatrix} = 0$$

where the matrices $\bar{A} \in \mathbb{R}^{(N-1)n_y \times N n_y}$ and $\bar{B} \in \mathbb{R}^{(N-1)n_y \times N n_u}$ are the following

$$\bar{A} = \begin{bmatrix} CAC^{-1} & -I_{n_u} & 0 & \dots & 0 \\ 0 & CAC^{-1} & -I_{n_u} & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & CAC^{-1} & -I_{n_u} \end{bmatrix} \qquad \bar{B} = \begin{bmatrix} CB & & \dots & 0 \\ 0 & CB & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & CB \end{bmatrix}.$$

We write the system with respect to noise variables, recalling (4), and substituting u with $u_e - e_u$ and y with $y_e - e_y$, so that we obtain the system:

$$\begin{bmatrix} \bar{A} & \bar{B} \end{bmatrix} \begin{bmatrix} e_y \\ e_u \end{bmatrix} = d \tag{7}$$

where

$$d = \begin{bmatrix} \bar{A} & \bar{B} \end{bmatrix} \begin{bmatrix} y_e \\ u_e \end{bmatrix}.$$

Calling

$$G = \begin{bmatrix} \bar{A} & \bar{B} \end{bmatrix} \in \mathbb{R}^{(N-1)n_y \times N(n_y + n_u)} \qquad \text{and} \qquad z = \begin{bmatrix} e_y \\ e_u \end{bmatrix} \in \mathbb{R}^{N(n_y+n_u)\times 1}$$

we obtain the linear system

$$Gz = d.$$

Since this system is underdetermined, and infinitely many solutions $\hat{z}$ exist, we can consider the least-squares solution

$$\begin{aligned} z_{ls} = \ & \underset{z}{\operatorname{argmin}} \, \|z\|_2^2 \quad = G^T (G G^T)^{-1} d \\ & \text{s.t.} \quad Gz = d \end{aligned}$$

that is the minimum norm solution. It is worth noticing that, if the variances of the noise signals e_u and e_y are unitary and their covariance as well as the means μ_{eu^i}, μ_{ey^i} $\forall i$ are zero, then z_{ls} is the solution of the problem given by equation (5) of [32] (reported in Section 2.3, Problem 2), therein demonstrated to be optimal. We will deal with the case of unknown and not necessarily unitary variances and not necessarily zero means.

2.1.2 White noise and bias separation

Since we are interested not only in the case of white noise (5), but also Gaussian one (6), we reformulate the problem separating the noisy signals in the sum of a zero mean signal (that we will call $\tilde{e}_{u,y}(k)$) and its mean ($\bar{e}_{u,y}$):

$$\begin{cases} \hat{e}_y(k) = \tilde{e}_y(k) + \bar{e}_y, \\ \hat{e}_u(k) = \tilde{e}_u(k) + \bar{e}_u. \end{cases}$$

Because of this separation we must add to the system the zero mean conditions on the vectors $\tilde{e}_{u^i}$ and $\tilde{e}_{y^i}$ for every i :

$$\sum_{k=0}^{N-1} \tilde{e}_{u^i}(k) = 0, \quad \text{for } i = 1, \dots, n_u, \qquad \sum_{k=0}^{N-1} \tilde{e}_{y^i}(k) = 0, \quad \text{for } i = 1, \dots, n_y.$$

We call the new unknown vector

$$\tilde{z} = [\tilde{e}_u, \tilde{e}_y, \bar{e}_u, \bar{e}_y]^T$$

and consider the new system with respect to the new variables and call it with the same notation

$$G\tilde{z} = d \tag{8}$$

where now $G \in \mathbb{R}^{(Nn_y+n_u)\times(N+1)(n_y+n_u)}$ and $d \in \mathbb{R}^{(Nn_y+n_u)\times 1}$.

2.2 Regularization

The considered system (8) is underdetermined, and the minimum norm solution is not able to determine a sufficiently good denoising, as we will see in Section 2.3. In fact, the optimal solution is the weighted least squares solution with weights that depends on the covariance values, as explained in [32] and recalled in Section 2.3. Therefore, since we are assuming unknown values of variances, and bad estimates of these values can produce bad denoised signals, the minimum norm solution (obtained assuming unitary variances) is not optimal.

For these reasons, we must consider additional conditions that regularize the problem, in order to obtain a good, unique denoised solution. Since there are no other known conditions to be minimized exactly, we must add adequately weighted regularization terms.

2.2.1 Single Signal Denoising

One of the main signal denoising methods is based on the Tikhonov regularization ([16], [21], [4]): given a signal $u_e \in \mathbb{R}^{N\times 1}$ corrupted with white noise, the denoised signal $\hat{u} \in \mathbb{R}^{N\times 1}$ is given by

$$\hat{u} = \min_u \|u - u_e\|_2^2 + \|\lambda L u\|_2^2 \tag{9}$$

for a certain value of the regularization parameter λ, with L a regularization matrix that represent the discrete approximation of a derivative operator. In the case of the second derivative, the matrix L is

$$L = \begin{bmatrix} 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(N-2)\times N}$$

and the obtained method is called Hodrick-Prescott filter [17] in economics and statistics, and is used for trend estimation of time series.

This technique is used for many applications, e.g. for the denoising of ECG signals [35]. In that work in particular, it is shown how the parameter λ is linked with the value of the variance of the noise signal: if its value is known it is possible to determine the optimal parameter λ with eq. (30) of that paper.

As pointed out in [22], the solution of problem (9) is smooth with respect to the regularization parameter λ that varies in the interval $[0, \infty)$:

- for λ that tends to zero, the solution tends to the measured signal u_e,
- for λ that tends to infinity, the solution converges to the affine subspace that best approximate the time series ("ba"= best affine):

$$u^{ba}(t) = \alpha^{ba} + \beta^{ba} t.$$

Now we generalize the problem above to multiple signals with $u \in \mathbb{R}^{Nn_u\times 1}$ defined as in Section 2:

$$u := [u^T(0) \cdots u^T(N-1)]^T \quad \in \mathbb{R}^{Nn_u\times 1}.$$

In this case the matrix L becomes

$$L = \begin{bmatrix} I_{n_u} & -2I_{n_u} & I_{n_u} & & & & \\ & I_{n_u} & -2I_{n_u} & I_{n_u} & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & I_{n_u} & -2I_{n_u} & I_{n_u} & \\ & & & & I_{n_u} & -2I_{n_u} & I_{n_u} \end{bmatrix} \in \mathbb{R}^{(N-2)n_u \times N n_u}$$

and we must consider n_u parameters $\lambda_{u1}, \ldots, \lambda_{u_{n_u}}$, one for each input variable, so that the minimization problem (9) becomes

$$\hat{u} = \min_u \|u - u_e\|_2^2 + \|\Lambda L u\|_2^2 \tag{10}$$

with the matrix $\Lambda \in \mathbb{R}^{(N-2)n_u\times(N-2)n_u}$ as follows

$$\Lambda = I_{N-2} \otimes \operatorname{diag}(\lambda_{u1}, \ldots, \lambda_{u_{n_u}}) = \begin{bmatrix} \operatorname{diag}(\lambda_{u1}, \ldots, \lambda_{u_{n_u}}) & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ddots & \operatorname{diag}(\lambda_{u1}, \ldots, \lambda_{u_{n_u}}) \end{bmatrix}$$

where ⊗ is the Kronecker product.

We note also that, setting $\hat{e}_u = u_e - \hat{u}$, problem (10) is equivalent to the following problem

$$\hat{e}_u = \min_{e_u} \|e_u\|_2^2 + \|\Lambda\, L\,(u_e - e_u)\|_2^2. \tag{11}$$

2.2.2 DLTI Model-Based Denoising

In this subsection we apply the regularization just considered to the underdetermined model-based denoising problem (8). Differently from the previous case, we must add two regularization terms for each input and output signal, and we obtain:

$$\min_{\tilde{z}} \left(\|G\tilde{z} - d\|_2^2 + \|\Lambda_{eu}^{min}\tilde{e}_u\|_2^2 + \|\Lambda_{ey}^{min}\tilde{e}_y\|_2^2 + \|\Lambda_{eu}^{curv} L_{n_u}(u_e - \hat{e}_u)\|_2^2 + \|\Lambda_{ey}^{curv} L_{n_y}(y_e - \hat{e}_y)\|_2^2 \right) \tag{12}$$

where the matrices $\Lambda_{<eu,ey>}^{<min,curv>}$ have the form of diagonal matrices, as described in the multiple signal case of previous section, more precisely

$$\Lambda_{eu}^{min} \in \mathbb{R}^{Nn_u \times Nn_u}, \qquad \Lambda_{ey}^{min} \in \mathbb{R}^{Nn_y \times Nn_y},$$
$$\Lambda_{eu}^{curv} \in \mathbb{R}^{(N-2)n_u \times (N-2)n_u}, \qquad \Lambda_{ey}^{curv} \in \mathbb{R}^{(N-2)n_y \times (N-2)n_y},$$

and where $L_{n_u} \in \mathbb{R}^{(N-2)n_u \times Nn_u}$ and $L_{n_y} \in \mathbb{R}^{(N-2)n_y \times Nn_y}$ are rectangular matrices generated by the discretization of the second derivative operator, as defined previously.

We call G_{reg} the matrix of the least-squares regularized problem (12)

$$G_{reg} \in \mathbb{R}^{(N(n_y+n_u)+N(n_u+n_y)+(N-2)(n_u+n_y))\times(N+1)(n_y+n_u)}$$

$$G_{reg} \in \mathbb{R}^{(3N-2)(n_y+n_u)\times(N+1)(n_y+n_u)}$$

and $\tilde{z}^* = [\tilde{e}_u^*, \tilde{e}_y^*, \bar{e}_u^*, \bar{e}_y^*]$ its solution.

Note that in the case of single input-single output (SISO), i.e. the scalar case with $n_u = n_y = 1$, $A = a \in \mathbb{R}$, $B = b \in \mathbb{R}$, we obtain

$$\min_{\tilde{z}} \left(\|G\tilde{z} - d\|_2^2 + \|\lambda_{eu}^{min}\tilde{e}_u\|_2^2 + \|\lambda_{ey}^{min}\tilde{e}_y\|_2^2 + \|\lambda_{eu}^{curv} L_{n_u}(u_e - \hat{e}_u)\|_2^2 + \|\lambda_{ey}^{curv} L_{n_y}(y_e - \hat{e}_y)\|_2^2 \right). \tag{13}$$

Analogously to the single signal denoising case, we can highlight the following cases:

- for $\lambda_{eu^i}^{min}, \lambda_{ey^j}^{min} \to \infty\ \forall i, j$ the vectors $\tilde{e}_u^*$ and $\tilde{e}_u^*$ of the solution $\tilde{z}^*$ of the problem will be trivially zero. In this case the model is not satisfied since the denoised signals u_{den}^*, y_{den}^* coincide with the measured signals but for additive constants
$$u_{den}^* = u_e - \hat{e}_u^* = u_e - \bar{e}_u^*, \qquad y_{den}^* = y_e - \hat{e}_y^* = y_e - \bar{e}_y^*.$$
- for $\lambda_{eu^i}^{curv}, \lambda_{ey^j}^{curv} \to \infty\ \forall i, j$ the solution $\tilde{z}^*$ of the problem is such that the denoised signals are
$$u_{den}^* = \alpha^{ba} + \beta^{ba} t, \quad y_{den}^* = \alpha^{ba} + \beta^{ba} t$$
i.e. are the best affine approximations of the time series. Hence the noise signals estimates tend to
$$e_u^* = u_e - u_{den}^*, \quad e_y^* = y_e - y_{den}^*.$$
In this case, likewise the previous one, the model constraints are not satisfied.
- for $\lambda_{eu^i}^{curv} = 0, \lambda_{ey^j}^{curv} = 0\ \forall i, j$ and parameters $\lambda_{<eu^i,ey^i>}^{min}$ sufficiently small so that the model constraints are accurately satisfied, the problem is equivalent to the problem in [31],[32] in the case in which the covariance of the input and output noise is zero: the parameters $\lambda_{<eu,ey>}^{min}$ have the same role of the variances of the two noise signals.
- for $\lambda_{eu^i}^{min} = 0, \lambda_{ey^j}^{min} = 0\ \forall i, j$ and $\lambda_{<eu^i,ey^i>}^{curv}$ sufficiently small so that the model constraints are accurately satisfied, we obtain a regularized problem that doesn't have conditions on the norm of the solution that could explode.

Although each single regularization term makes the problem determined-overdetermined, they are not sufficient to generate an acceptable solution. Therefore, it is necessary to find the right trade-off of the parameters and consider a method for this multi-parameter choice problem.

The parameters cannot be determined by the minimization of the residual of problem (12), because that condition gives bad parameters of the kind listed above. This happens because the real solution is not the one that gives the minimum of the objective function in (12). For this reason we consider the Normalized Cumulative Periodogram ([15], [16]), also known as Bartlett test, as the quantity to be minimized. We use this method to test the whiteness of the vectors $\tilde{e}_u^*$ and $\tilde{e}_y^*$, i.e. the distance of these vectors from white noises, as are supposed to be in this work. None of the degenerate cases listed above gives a good value of whiteness of the solution, hence this method is able to keep the solution away from those critical situations and find acceptable solutions.

In this paper we propose an iterative algorithm that minimizes the residual and the whiteness of the solution of (12), with iterative optimizations on single parameters $\lambda_{<eu^i,ey^i>}^{<min,curv>}$. We will call this algorithm "Whiteness and Minimum-Curvature Model-Based Denoising" (WMC-MBD) and we will describe it in detail in Section 4.

2.3 Comparison with previous works

In [31] and [32] a deterministic DLTI system in state-space form is considered (analogous to (1)) with noisy inputs and outputs, and is referred to as "noisy I/O model". The problem of estimating the state and the denoised input and output signals is then addressed, in the online and offline cases, defined respectively *filtering* and *smoothing* problems. In our case we are interested in the second of these problems that we quote here:

Problem 2 ("Optimal noisy I/O smoothing problem"). Call $\tilde{u}, \tilde{y}$ measurement errors random, centered, normal, uncorrelated and white, with known covariance matrices

$$cov(\tilde{u}) =: V_{\tilde{u}}(t), \quad cov(\tilde{y}) =: V_{\tilde{y}}(t)$$

and suppose that the initial condition $x_0 = \hat{x}(0)$ is known. Then, given the matrices A, B, C, D, the optimal noisy I/O smoothing problem is defined as

$$\begin{aligned} &\min_{\hat{u},\hat{y},\hat{x}} \left\| \begin{bmatrix} V_{\tilde{u}} & \\ & V_{\tilde{y}} \end{bmatrix}^{-1/2} \begin{bmatrix} \hat{u} - u_d \\ \hat{y} - y_d \end{bmatrix} \right\|_2^2 \\ &\text{s.t. } \ \hat{x}(t+1) = A\hat{x}(t) + B\hat{u}(t) \\ &\qquad \hat{y}(t) = C\hat{x}(t) + D\hat{u}(t) \qquad \text{for } \ t = 0, 1, \ldots, t_f - 1. \end{aligned} \tag{14}$$

The *optimal smoothed state estimate* $\hat{x}(\cdot, t_f)$ is the solution of (14).

In the notation of this paper $t_f = N$. The estimate $\hat{x}(\cdot, t_f)$ is the optimal smoothed state estimate with a time horizon t_f, i.e. with the input and output signals supposed known in all the time instants. Hence, in this case, the resolution of the underdetermined system given by the model equations is found imposing the minimum of a weighted norm of the solution.

We observe here that our formulation have the model-constraining equations in common with the more general one considered in [32]. On the other hand we are assuming a particular case of the problem with C invertible and $D = 0$ and we are not interested in the state estimation. Moreover, in the approach of [32] the regularization terms considered are the norms of the noises, weighted with the covariance matrices known values. In this work, we don't suppose variance values to be known and we add regularization terms based on the curvature of these signals (second derivative).

3 General issues

3.1 Uncertainty of the noise bias values

Equations of system (8) only determines the ratio of the biases $\bar{e}_u, \bar{e}_y$, but not their independent values. In particular, every couple $(e_{y_{offset}}, e_{u_{offset}})$ such that

$$e_{y_{offset}} - M_0 \, e_{u_{offset}} = 0 \tag{15}$$

satisfies the model equations (7), where M_0 is the static gain of the model (steady-state value of the unit step response): $M_0 = (I - A)^{-1} B$ (where we recall that we are considering the case $C = I$ and $D = 0$).

Moreover, none of the regularization terms introduced in paragraph 2.2.2 gives any additional information to resolve this uncertainty, that remains in the regularized problem (12). This situation can be dangerous since the bias estimates calculated can explode, especially when the conditioning number of the regularized matrix G_{reg} obtained from (12) is big.

For this reason, we introduce an additional constraint equation, that weights the proximity of the bias $\bar{e}_u$ to an a-priori estimate of it, obtained with the truncated SVD method applied to the regularized matrix G_{reg}.

In fact, the bias of the noise e_u can be estimated as follows. Solving the regularized system with zero $\lambda^{curv}_{<eu,ey>}$ values and relatively big $\lambda^{min}_{<eu,ey>}$ values, truncating at the first singular value, i.e. considering only the first principal component, the solution $\hat{e}_u, \hat{e}_y$ is almost constant (i.e. $\tilde{e}_u, \tilde{e}_y$ are almost zero).

3.2 Singular values and ill-conditioning

If we consider small regularization parameters, $\lambda^{curv}_{<eu,ey>}$ and $\lambda^{min}_{<eu,ey>}$, the singular values of the regularized matrix G_{reg} have a constant trend for different kind of noise signals. As shown in Figure 1, this trend is characterized by some singular values of bigger amplitude, relative to the model-constraining, a gap, a group of smaller singular values, another gap, and a group of nearly zero singular values; this last group is the responsible for the uncertainty on the bias estimation of Section 3.1.

The condition number of the matrix with small values of the regularization parameters is high and for this reason the matrix is numerically singular, in double precision. Since regularization parameters are chosen a-posteriori, with a criterion based on the whiteness of the estimated noise vector, it is necessary to solve the least-squares problem derived from (12) with a regularization approach to remain in the dynamic range of double precision, in particular we choose the Truncated SVD (TSVD).

4 The "Whiteness and Minimum Curvature Model-Based Denoising" (WMC-MBD) algorithm

We propose here a denoising algorithm based on the considerations in Section 3.

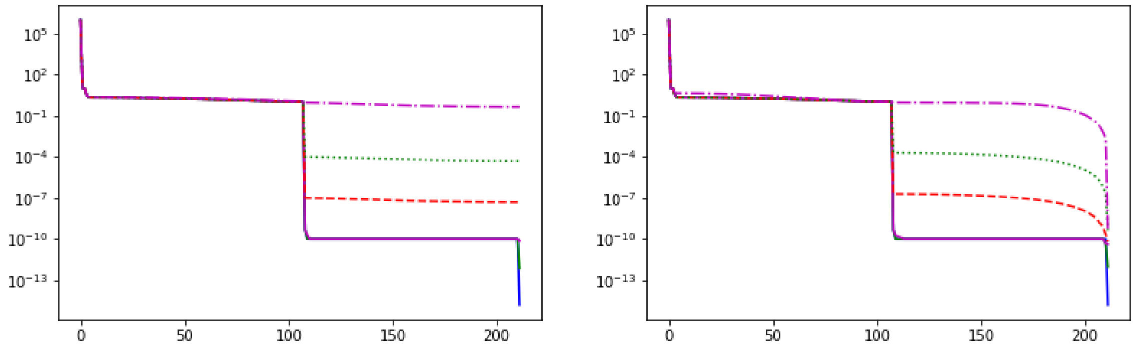


Figure 1: Singular values of the regularized matrix G_{reg} for different values of the regularization parameters: increasing λ^{min} (left) we see, going from the solid to the dashed line, a Tikhonov effect on singular values, while increasing λ^{curv} (right) we notice that the condition number remains the same.

4.1 Iterative Algorithm

We consider an algorithm in which we iterate single-parameter optimizations to maximize the whiteness of the estimated noise signals. Given an estimated noise signal e, we minimize its deviation from white noise using the following measure of whiteness loss:

Definition 1 (Whiteness Loss). Given an estimated noise signal $e \in \mathbb{R}^N$, we call *Whiteness Loss* of e

$$w_e := \|l - NCP(e)\|_2 \tag{16}$$

where $l \in \mathbb{R}^N$ is the vector of equispaced values from 0 to 1 and $NCP(e)$ is the Normalized Cumulative Periodogram of the signal e, defined (see [16]) as the vector

$$NCP(e)_i := \frac{(p_e)_2 + (p_e)_3 + \cdots + (p_e)_{i+1}}{(p_e)_2 + (p_e)_3 + \cdots + (p_e)_{q+1}} \qquad \text{for } i = 1, \ldots, q = \lfloor N/2 \rfloor$$

where

$$p_e = [|(f_e)_1|^2, |(f_e)_2|^2, \ldots, |(f_e)_N|^2]^T$$

is the power spectrum density and $f_e = dft(e) = [(f_e)_1, (f_e)_2, \ldots, (f_e)_N]^T \in \mathbb{C}^N$ is the discrete Fourier transform of e.

Definition 2 (Curvature). Given a signal $u \in \mathbb{R}^{N \times 1}$, we call *curvature of* u the signal

$$Lu \quad \in \mathbb{R}^{(N-2)\times 1}$$

where

$$L = \begin{bmatrix} 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 & \\ & & & & 1 & -2 & 1 \end{bmatrix} \in \mathbb{R}^{(N-2)\times N}.$$

In the iterative algorithm we will use definition (16) as a distance measure of the estimated noise signals from white noise. Hence, we will consider the functions

$$\begin{cases} \omega_{eu^i} = \|l - NCP(eu^i)\|_2 \\ \omega_{ey^i} = \|l - NCP(ey^i)\|_2 \end{cases}$$

where eu, ey are obtained from solving problem (12) for certain values of the parameters. The algorithm aims at finding such parameters $\Lambda^{<min,curv>}_{<eu,ey>}$ that produce estimated noise signals with a minimum value of the whiteness loss.

We can now introduce the "Whiteness and Minimum Curvature Model-Based Denoising" algorithm:

Algorithm 1. "Whiteness and Minimum Curvature Model-Based Denoising" (WMC-MBD)

A-priori estimate of the input noise bias $\bar{e}_u$ as described in subsection 3.1

Initialization of the current regularization parameters $\lambda^{min}_{<eu^i,ey^i>}$, $r_{<eu^i,ey^i>} = \frac{\lambda^{curv}_{<eu^i,ey^i>}}{\lambda^{min}_{<eu^i,ey^i>}}$ and initial guess of their optimal values $r^*_{eu^i}, \lambda^{min*}_{eu^i}, r^*_{ey^i}, \lambda^{min*}_{ey^i}$ to the initial mid-range value 10^{-7}

for $k = 1, \ldots, K_{maxiter}$ **do**

  for $i = 1, \ldots, n_u$ **do**

    1) $\bar{r}_{eu^i} = \underset{r_{eu^i}}{\operatorname{argmin}}\, w_{eu^i}$

  end for

for $i = 1, \ldots, n_u$ **do**
  1) $\bar{\lambda}^{min}_{eu^i} = \underset{\lambda^{min}_{eu^i}}{\text{argmin}}\, w_{eu^i}$
end for
for $i = 1, \ldots, n_y$ **do**
  1) $\bar{r}_{ey^i} = \underset{r_{ey^i}}{\text{argmin}}\, w_{ey^i}$
end for
for $i = 1, \ldots, n_y$ **do**
  1) $\bar{\lambda}^{min}_{ey^i} = \underset{\lambda^{min}_{ey^i}}{\text{argmin}}\, w_{ey^i}$
end for
Curvature check: if the curvature of both I/O denoised signals is less than the previous iteration, update the optimal values $r^*_{eu^i}, \lambda^{min*}_{eu^i}, r^*_{ey^i}, \lambda^{min*}_{ey^i}$ to their current values
end for

end

where w_{eu} and w_{ey} are respectively the functions that calculate the whiteness loss values (equation (16)) of the estimated signals e_u and e_y obatined from solving (12) with the current parameters $\Lambda^{<min,curv>}_{<eu,ey>}$. The variables r_{eu} and r_{ey} are respectively the ratios between the parameters of each signal of input and output

$$r_{eu} = \frac{\lambda^{curv}_{eu}}{\lambda^{min}_{eu}}, \qquad r_{ey} = \frac{\lambda^{curv}_{ey}}{\lambda^{min}_{ey}}.$$

We consider adaptive grids to perform each global minimization in Algorithm 1. This is because the whiteness loss function is not a convex function of the parameters, hence an optimization algorithm (such as `scipy.optimize.minimize` of the Scipy library for Python) can easily stop on local minima.

We consider grids obtained in the following way: at each iteration we calculate the grid by multiplying the actual value of the considered parameter to the following vector

$$10^{k_{exp}} * [1, \frac{1}{2^1}, \frac{1}{2^2}, \frac{1}{2^3}, \ldots, \frac{1}{2^j}, 1, \frac{1}{2^{j+1}}, \ldots, \frac{1}{2^{j_{max}+1}}]$$

where $k_{exp} = (K_{maxiter} + 1 - k)$ for each iteration k, $K_{maxiter}$ is the maximum number of iterations, and $j_{max} = log_2(10^{k_{exp} - (-k_{exp})})$ so that this vector span values from $10^{k_{exp}}$ to $10^{-k_{exp}}$. Moreover, such grids are restricted so that the values considered are only the ones for which the parameters belong to the interval $[10^{-15}, 10^{-2}]$: on one hand the value of the parameters must be sufficiently bigger than machine precision, on the other hand a value of the parameters too big would compromise the model constraints.

After each group of four kinds of minimizations, the optimal parameters are conditionally updated, so that the tuple that generates a solution of minimum curvature among the ones obtained during the iteration is kept as the final solution. This is done because among the tuples that minimize the whiteness function, we are interested in the smoothest one, and it allows us to exclude over-whitened solutions which curvature can be high.

As already introduced, the assumption of unknown value of the norm of the noise implies that the proposed method belongs to the heuristic class. Moreover as we observed, the true signals do not satisfy the minimum of the regularization terms of equation (12). Therefore Algorithm 1 can reach satisfying denoised signals but not, in general, the original ones.

At each grid point of the optimization the following computations are performed: an SVD of matrix G_{reg}, the TSVD solution of the system and one FFT for the calculation of the whiteness loss, with a total computational cost

$$\mathcal{O}(\min(n_{reg}\, m^2_{reg}, n^2_{reg}\, m_{reg})) + \mathcal{O}(N\ \log(N))$$

where we call n_{reg}, m_{reg} the dimensions of matrix G_{reg}.

In the end, we observe that, in this work, both biases and variances of the noise signals are supposed to be unknown. However, in case these quantities are available, the algorithm can take advantage of that. In fact, if one of the biases of the noises is known, there's no need to calculate the a-priori estimate of the input bias. Moreover, if one of the variance values is available, it is possible to determine the ratio between the parameters $r_{<eu,ey>} = \lambda^{curv}_{<eu,ey>} / \lambda^{min}_{<eu,ey>}$ as in [35] and skip the calculation of these parameters in Algorithm 1.

5 Numerical experiments

In this section we will see the numerical comparison between the denoising carried out with the "modified Kalman filter" of [32] and the WMC-MBD (1). We will show the effectiveness of Algorithm 1 through a Monte Carlo simulation, likewise [10].

5.1 Results and numerical comparisons of algorithms

We will show in the first two paragraphs the results for two particular examples, with sinusoidal and piecewise constant input, and after these we will show the results obtained for a bigger group of tests. For ease of presentation, to test the proposed method we reduce to the single input-single output (SISO) case, hence $n_u = n_y = 1, A = a \in \mathbb{R}, B = b \in \mathbb{R}$.

5.1.1 Example 1: sinusoidal input

In this example we consider a sinusoidal input $u(t) = Asin(3\,t)$ with amplitude $A_u = 10$, sampled with sampling time $dt = 0.1$ in the interval $[0, 10]$. The additive Gaussian noises of this example have the following statistical properties: standard deviations $\sigma_{eu} = 5$, $\sigma_{ey} = 10$, and means $\mu_{eu} = 0$, $\mu_{ey} = 1$, and are generated with a Python function from the Numpy library: `numpy.random.normal(bias, std, size=N)`. In Figure 2 we can see the noisy Input and Output signals and the signal obtained as their product, that is an example of a quantity of interest derived from the system I/O signals. In these graphs the sinusoidal trends are altered and difficult to recognize.

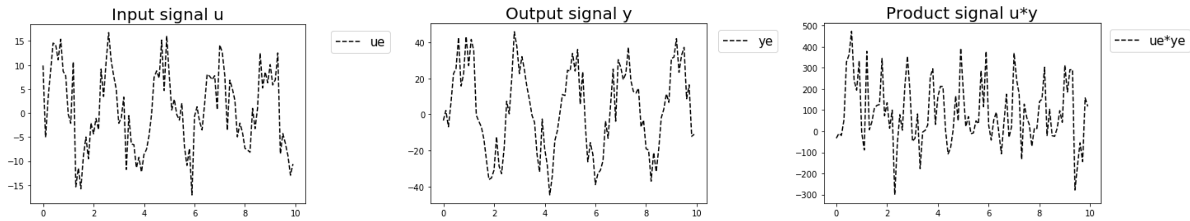
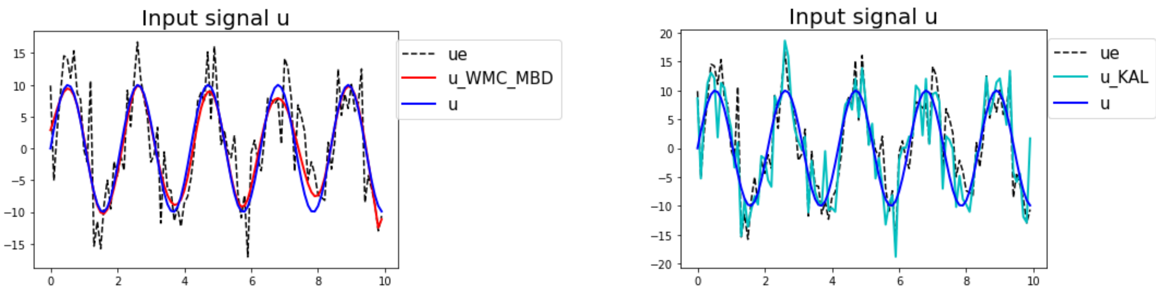
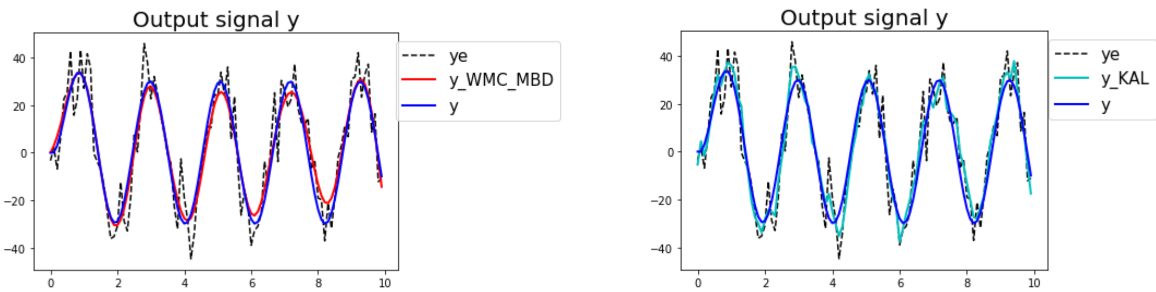


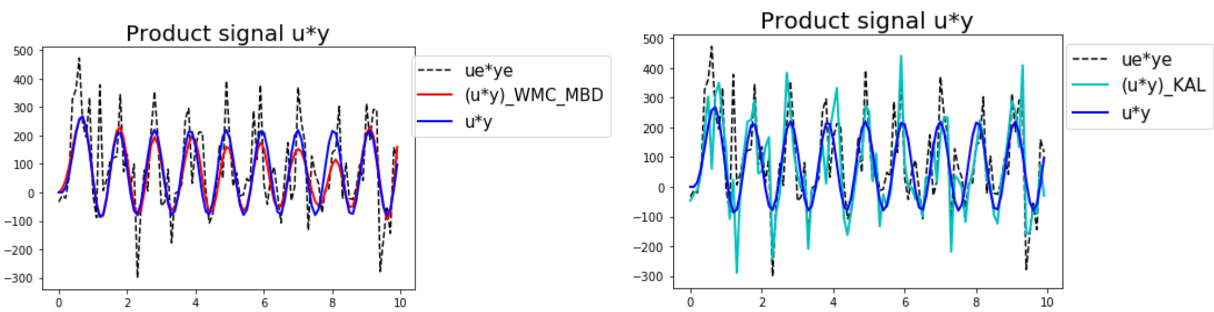
Figure 2: Example 1: Noisy Input, Output and Product signals.



(a) Input signals comparison



(b) Output signals comparison



(c) Product signals comparison

Figure 3: Comparison of the Input, Output and product signals for Example 1 respectively in figures 3a, 3b and 3c: in the left figures the noisy signal (dashed line), and the comparison of the true signal with the denoised signal from Algorithm 1 (WMC-MBD), in which the statistical properties of the noise are supposed unknown; in the right figures the noisy signal (dashed line), and the comparison of the true signal with the denoised signal with the "modified Kalman filter" of [32], in the case in which correct variances values are used, but biases are assumed zero (since it is an assumption of the method).

In Figures 3a, 3b and 3c we show the noisy and denoised input, output and product signals respectively, as follows:

- in the left figures the noisy signal (dashed line), and the comparison of the true signal with the denoised signal from Algorithm 1 (WMC-MBD), in which the statistical properties of the noise are supposed unknown. For this test 8 iterations of the proposed algorithm are used;
- in the right figures the noisy signal (dashed line), and the comparison of the true signal with the denoised signal with the "modified Kalman filter" of [32], in the case in which correct variances values are used, but biases are assumed zero (since it is an assumption of the method).

5.1.2 Example 2: piecewise constant input

In our second example we consider a piecewise constant input, with additive Gaussian noises with the same properties as the previous example. In Figure 4 we can see the noisy Input, Output and product signals, in which the discontinuous trend is altered and difficult to recognize. Analogously to the previous case we show in Figure 5a, 5b and 5c respectively the results for the input, output and product signals. As in the previous example, eight iterations of the proposed algorithm are used $K_{maxiter} = 8$.

We considered this example since for these kind of signals other type of norms and regularizations are usually preferred, for example Total Variation ([14], [33], [41], [7]). However the obtained results show that this method is able to perform a useful denoising also in this critical situation.

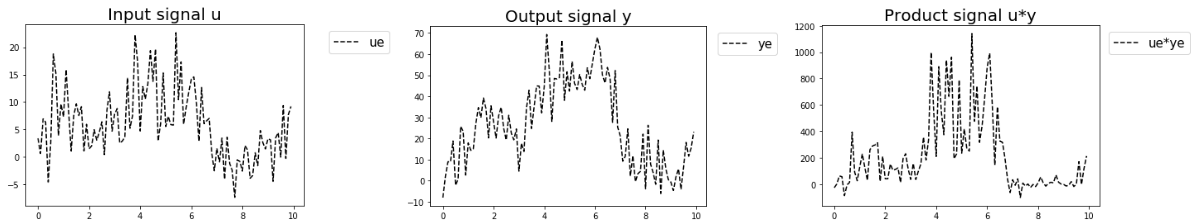


Figure 4: Example 2: Noisy Input, Output and Product signals.

5.1.3 Numerical results of the comparisons

In this paragraph we show the results of the Algorithm 1 on a total of 100 tests for different kinds of input signals and different values of variances and biases of I/O signals.

We consider 4 different kinds of input signals: a sinusoid, a piecewise constant signal (as in the two previous examples) and other two kinds of sum of two sinusoids with different amplitude and frequency values.

The couple of standard deviation values (square root of the variances) considered are the following

$$(\sigma_{eu}, \sigma_{ey}) = [(5, 10), (5, 12), (2.5, 15), (2.5, 7), (1.5, 10)].$$

For each of these, the following couples of biases have been tested:

$$(\mu_{eu}, \mu_{ey}) = [(0, 0), (0, 1), (1, 0), (1, -1), (-2, 3)].$$

Hence 25 kinds of input-output noises have been tested for each of the 4 input signals.

In Tables 1, 2, 3 and 4, the values of mean and standard deviation of the relative errors for input and output signals are shown, for each type of input:

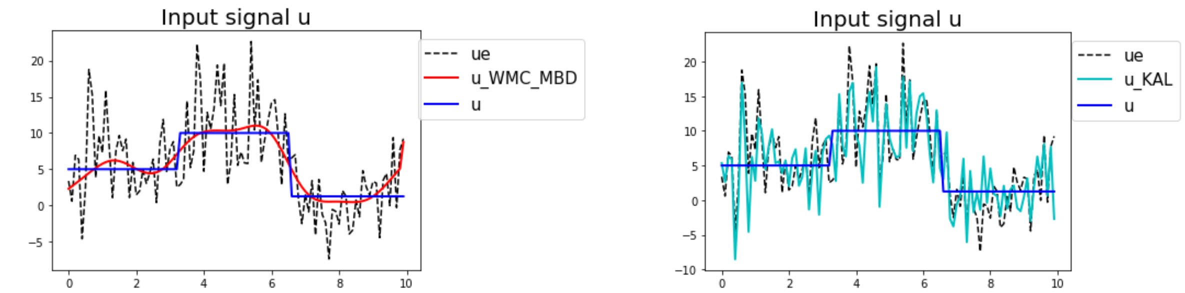
$$re_u = \frac{\|u - u^*\|}{\|u\|}, \qquad re_y = \frac{\|y - y^*\|}{\|y\|}.$$

Relative errors are shown in order for noisy measurements ("Noisy"), for I/O signals obtained with the "modified Kalman filter" of [32], in the case in which the variances are supposed known ("KAL exact vars") and unknown with hypothetic unitary values ("KAL vars=1"), and for I/O signals denoised with Algorithm 1 ("WMC-MBD"). As in the previous examples, 8 iterations of the proposed algorithm are used.

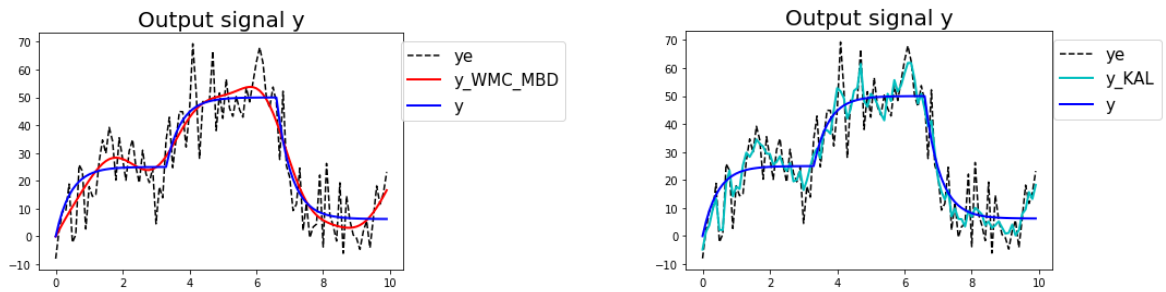
data	re_u mean	re_u std	re_y mean	re_y std
Noisy	0.493	0.194	0.515	0.126
KAL (exact vars)	0.484	0.157	0.216	0.059
KAL (vars=1)	0.743	0.166	0.293	0.073
WMC-MBD	0.231	0.109	0.197	0.07

Table 1: Sinusoid Input Comparisons of relative errors re

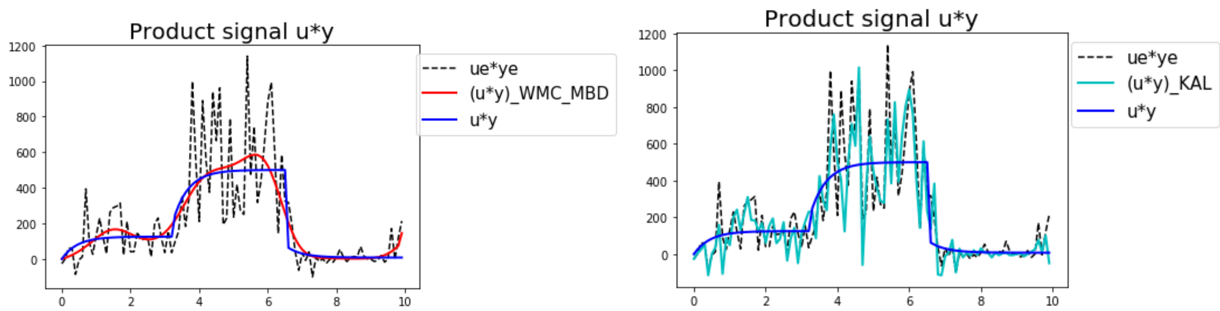
Table 5 contains the results for all the tests in which the noise biases are zero, for all the four kinds of inputs. From these results, it is possible to compare the denoising obtained with the "modified Kalman filter" with known and unknown variances values. In this second case we suppose hypothetic unitary values, and the results show that the denoising errors in this case are higher.



(a) Input signals comparison



(b) Output signals comparison



(c) Product signals comparison

Figure 5: Comparison of the Input, Output and product signals for Example 2 respectively in figures 5a, 5b and 5c: in the left figures the noisy signal (dashed line), and the comparison of the true signal with the denoised signal from Algorithm 1 (WMC-MBD), in which the statistical properties of the noise are supposed unknown; in the right figures the noisy signal (dashed line), and the comparison of the true signal with the denoised signal with the "modified Kalman filter" of [32], in the case in which correct variances values are used, but biases are assumed zero (since it is an assumption of the method).

data	re_u mean	re_u std	re_y mean	re_y std
Noisy	0.462	0.189	0.338	0.09
KAL (exact vars)	0.432	0.157	0.141	0.036
KAL (vars=1)	0.671	0.166	0.189	0.05
WMC-MBD	0.286	0.104	0.139	0.045

Table 2: Sum of sinusoids ($u = (A_u/2)sin(6t) + A_u sin(t)$) Input Comparisons of relative errors re

data	re_u mean	re_u std	re_y mean	re_y std
Noisy	0.442	0.168	0.532	0.139
KAL (exact vars)	0.433	0.14	0.229	0.063
KAL (vars=1)	0.66	0.15	0.309	0.079
WMC-MBD	0.308	0.104	0.242	0.085

Table 3: Sum of sinusoids ($u = A_u sin(6t) + (A_u/2)sin(t)$) Input Comparisons of relative errors re

data	re_u mean	re_u std	re_y mean	re_y std
Noisy	0.525	0.196	0.349	0.086
KAL (exact vars)	0.508	0.17	0.151	0.039
KAL (vars=1)	0.791	0.187	0.201	0.048
WMC-MBD	0.246	0.104	0.132	0.053

Table 4: Discontinuous Input Comparisons of relative errors re

data ($\mu_{<eu,ey>} = 0$)	re_u mean	re_u std	re_y mean	re_y std
Noisy	0.454	0.198	0.424	0.143
KAL (exact vars)	0.462	0.168	0.182	0.064
KAL (vars=1)	0.71	0.193	0.242	0.084
WMC-MBD	0.255	0.098	0.169	0.076

Table 5: Comparisons of relative errors re for the tests with zero bias noises ($\mu_{eu} = 0, \mu_{ey} = 0$) for all the different kind of signals

6 Conclusions

In this article we have presented a new algorithm, called "Whiteness Minimum-Curvature Model-Based Denoising" (WMC-MBD), for the denoising of data affected by white Gaussian noise whose statistical description, in terms of mean and variance, is not known. The algorithm produces denoised data that satisfies the input/output relations of the model used to describe the physical system.

The algorithm has been tested with success on various kind of signals: sinusoids, sums of sinusoids at distant frequencies and discontinuous signals as well.

Natural extensions of this work could be to consider different norms used for the regularization. For example, the norms used to regularize u and y could be different: u may be discontinuous, therefore the Total Variation could attain better results than the 2-norm, while y may be substantially smoother, being the system's output. Moreover, in the case of discontinuous signals it could be more useful to consider norms ℓ_p with $0 < p < 1$ and, therefore, methods that approximate such norms, like the Iterative Reweighting Least Squares or the Robust Least Squares.

7 Acknowledgments

The authors gratefully acknowledge the doctoral grant funded by Electrolux Italia s.p.a. "Calcolo ad alte prestazioni per il Model Based Design applicato alla progettazione di elettrodomestici" and in particular the application specific contribution given by eng. Alberto Moro.

References

[1] A. Bakushinsky. Remarks on choosing a regularization parameter using the quasioptimality and ratio criterion. *U.S.S.R. Comput. Math. Math. Phys.*, 24: 181–182,1984.

[2] A. Beghi, F. Marcuzzi, P. Martin, F. Tinazzi, M. Zigliotto. Virtual prototyping of embedded control software in mechatronic systems: A case study. *Mechatronics*, 43:99 –111, 2017.

[3] M. Belge, M. E. Kilmer, E. L. Miller Efficient determination of multiple regularization parameters in a generalized L-curve framework. *Inverse Problems, IOP Publishing*, 18(4): 1161–1183, jul 2002.

[4] S. Boyd, A. Vandenberghe, Faybusovich. Convex Optimization. *IEEE Transactions on Automatic Control*, 51: 1859-1859, 2006.

[5] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, S. Seatzu. Multi-parameter regularization techniques for ill-conditioned linear systems. *Numerische Mathematik*, 94: 203-228, 2003.

[6] J. V. Candy. Model-Based Signal Processing (Adaptive and Learning Systems for Signal Processing, Communications and Control Series). *Wiley-IEEE Press*, 2005.

[7] A. Chambolle. An Algorithm for Total Variation Minimization and Applications. *Journal of Mathematical Imaging and Vision*, 20(1):89–97, Jan 2004.

[8] Y. Chen, M. Akutagawa, Y. Kinouchi, Q. Zhang. Neural Network Based Audio Signal Denoising. *Proceedings of the 2008 International Conference on Advanced Infocomm Technology, ICAIT '08*, 54: 1–6, 2008.

[9] G. Deolmi, F. Marcuzzi A parabolic inverse convection-diffusion-reaction problem solved using space-time localization and adaptivity. *Applied Mathematics and Computation*, 219(16):8435 – 8454, 2013.

[10] R. Diversi, R. Guidorzi, U. Soverini. Algorithms for optimal errors-in-variables filtering. *Systems & Control Letters*, 48(1): 1–13, 2003.

[11] M. Fornasier, V. Naumova, S. V. Pereverzyev. Parameter Choice Strategies for Multipenalty Regularization. *SIAM J. Numerical Analysis*, 52:1770–1794,2014.

[12] S. Gazzola, L. Reichel. A new framework for multi-parameter regularization. *BIT Numerical Mathematics*, 56(3): 919–949, 9 2016.

[13] R. Guidorzi, R. Diversi, U. Soverini. Optimal errors-in-variables filtering. *Automatica*, 39(2): 281 – 289, 2003.

[14] A. Haddad, Y. Meyer. An improvement of Rudin-Osher-Fatemi model. *Applied and Computational Harmonic Analysis*, 22(3):319–334, 2007.

[15] P. C. Hansen, M. E. Kilmer, R. H. Kjeldsen. Exploiting Residual Information in the Parameter Choice for Discrete Ill-Posed Problems. *BIT Numerical Mathematics*, 46(1):41–59, Mar 2006.

[16] P. C. Hansen. Discrete Inverse Problems: Insight and Algorithms. *Society for Industrial and Applied Mathematics*, 2010.

[17] R. J. Hodrick, E. C. Prescott. Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit and Banking*, 29(1):1–16, 1997.

[18] G. Holler, K. Kunisch, R. C. Barnard. A bilevel approach for parameter learning in inverse problems. *Inverse Problems, IOP Publishing*, 34(11): 115012, sep 2018.

[19] K. Ito, B. Jin, T. Takeuchi. Multi-Parameter Tikhonov Regularization. *Methods and Applications of Analysis*, 18:31–46,2011.

[20] T. Kailath. Linear Systems. *Prentice-Hall*, 1980.

[21] M. E. Kilmer, P. C. Hansen, M. I. Espanol A projection-based approach to general-form Tikhonov regularization. *SIAM Journal on Scientific Computing*, 29(1): 315–330, 2007.

[22] S. Kim, K. Koh, S. Boyd, D. Gorinevsky. ℓ_1 Trend Filtering. *SIAM Review*, 51(2):339-360, 2009.

[23] S. Kindermann. Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems. *Electronic Transactions on Numerical Analysis*, 38:233–257, 2011.

[24] K. Kunisch, T. Pock. A Bilevel Optimization Approach for Parameter Learning in Variational Models. *SIAM J. Imaging Sciences*, 6:938–983, 2013.

[25] Z. Liu, A. Hansson, L. Vandenberghe Nuclear norm system identification with missing inputs and outputs. *Systems & Control Letters*, 62:605–612, 2013.

[26] D. Lorenz, T. Köhler. A Comparison of Denoising Methods for One Dimensional Time Series. *DFG-Schwerpunktprogramm 1114, Mathematical methods for time series analysis and digital image processing, Zentrum für Technomathematik*, 2005.

[27] S. Lu, S. Pereverzev, Y. Shao, et al. Discrepancy curves for multi-parameter regularization. *Journal of Inverse and Ill-posed Problems*, 18(6): 655-676, 2010.

[28] M. Mafi, H. Martin, M. Cabrerizo, J. Andrian, A. Barreto, M. Adjouadi. A comprehensive survey on impulse and Gaussian denoising filters for digital images. *Signal Processing*, 157:236–260, 2019.

[29] F. Marcuzzi. Space and time localization for the estimation of distributed parameters in a finite element model. *Computer Methods in Applied Mechanics and Engineering*, 198(37):3020 – 3025, 2009.

[30] F. Marcuzzi. Linear estimation of physical parameters with subsampled and delayed data. *Journal of Computational and Applied Mathematics*, 331:11–22, 2018.

[31] I. Markovsky, B. De Moor. Technical Communique: Linear Dynamic Filtering with Noisy Input and Output. *Automatica*, 41(1): 167–171, Jan. 2005.

[32] I. Markovsky, B. De Moor. Linear Dynamic Filtering with Noisy Input and Output. *IFAC Proceedings Volumes*, 36(16): 1711 – 1716, 2003.

[33] L. I. Rudin, S. Osher, E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[34] M. U. Sadiq, J. P. Simmons, C. A. Bouman. Model based image reconstruction with physics based priors. *2016 IEEE International Conference on Image Processing (ICIP)*, 2016: 3176–3179, 2016.

[35] R. Sameni. Online filtering using piecewise smoothness priors: Application to normal and abnormal electrocardiogram denoising. *Signal Processing* , 133:52–63, 2017.

[36] T. Söderström, U. Soverini, K. Mahata. Perspectives on errors-in-variables estimation for dynamic systems. *Signal Processing*, 82(8):1139 - 1154, 2002.

[37] T. Söderström, R. Diversi, U. Soverini. A unified framework for EIV identification methods in the presence of mutually correlated noises. *IFAC Proceedings Volumes*, 47(3): 4644 –4649, 2014.

[38] A.N. Tikhonov, V.B. Glasko. Use of the regularization method in non-linear problems. *USSR Computational Mathematics and Mathematical Physics*, 5(3): 93–107, 1965.

[39] S. Van Huffel, I. Markovsky, R. J. Vaccaro, T. Söderström. Total least squares and errors-in-variables modeling. *Signal Processing*, 87(10): 2281 - 2282, 2007.

[40] A. Vlasenko, C. Schnorr. Physically Consistent and Efficient Variational Denoising of Image Fluid Flow Estimates. *IEEE Transactions on Image Processing*, 19(3): 586–595, March 2010.

[41] M. Zhu, S. J. Wright, T. F. Chan. Duality-based algorithms for total-variation-regularized image restoration. *Computational Optimization and Applications*, 47(3):377–400, Nov 2010.