A Kalman Filter Framework for Simultaneous LTI Filtering and Total Variation Denoising

Arman Kheirati Roonizi and Ivan W. Selesnick, *Fellow, IEEE*

***Abstract*—This paper proposes a Kalman filter framework for signal denoising that simultaneously utilizes conventional linear time-invariant (LTI) filtering and total variation (TV) denoising. In this approach, the desired signal is considered to be a mixture of two distinct components: a band-limited (e.g., low-frequency component, high-frequency component) signal and a sparse-derivative signal. An iterative Kalman filter/smoother approach is formulated where zero-phase LTI filtering is used to estimate the band-limited signal and TV denoising is used to estimate the sparse-derivative signal.**

***Index Terms*—Sparse derivative signal, zero-phase filters, band-limited signal, total variation, Kalman filter/smoother.**

I. Introduction

THE success of linear time invariant (LTI) filtering and sparsity based denoising, i.e., total variation (TV) denoising, in a wide range of applications in signal/image processing, including signal restoration, denoising, deconvolution, compressed sensing, *etc*. are now widely recognized [1], [2], [3], [4], [5], [6], [7], [8], [9]. LTI filtering is particularly suited for signal filtering when the signal of interest is restricted to a known frequency band while the TV denoising is most suited for signals that are sparse derivative or admitting a sparse derivative representation. TV denoising as a regularization approach [10] forces the underlying signal to have a sparse derivative by enforcing the derivative to be small in the sense of ℓ_1-norm. This determines the jumps while coarsening the smooth regions. This property makes TV regularization suited for extraction of piecewise polynomial components. In contrast to TV denoising, zero-phase IIR filters (e.g., Butterworth, Chebyshev, etc) can be implemented using either forward-backward filtering or least-squares optimization approach [11]. For instance, a zero-phase Butterworth filter can be implemented as regularization approaches where the derivative of the signal enters as constraint in the sense of ℓ_2-norm. Therefore, the distinct difference between TV regularization and conventional linear filtering (e.g., zero-phase Butterworth filter) is the way that they define the regularization term. TV regularization is defined by a regularized optimization problem where the ℓ_1-norm of the signal derivative is used in the regularization term while in zero-phase Butterworth filtering, the ℓ_2-norm of the signal derivative is used in the regularization term. The ℓ_1-norm is non-differentiable while the ℓ_2-norm is differentiable. That is why TV regularization results in nonlinear filtering while zero-phase Butterworth filter does not. In summary, TV regularization is mainly used for signal filtering when the signal of interest is sparse-derivative while zero-phase LTI filters mostly used to extract a signal within a predetermined frequency band. However, in some signal processing applications, the signals are more complex. For instance, consider a situation in which a discrete event phenomenon is observed in the presence of a band-limited signal. In this case, the underlying signal can be modeled as mixture of a sparse and a band-limited component. For such signals, both approaches have some limitations that make them inefficient to reconstruct the signal of interest. Therefore, they must be combined to effectively reconstruct the desired components signal. In recent years, various methods have been developed for signal filtering/denoising which are based on the combination of TV regularization and other methods such as least-square polynomial signal smoothing, Tikhonov regularization and low-pass filtering [12], [13], [14]. In the following, we describe these methods, starting from TV regularization as it is commonly shared in all those algorithms.

A. Total Variation

TV denoising is an unconstrained optimization approach to estimate a signal x, having a (approximately) sparse derivative, from a noisy observation $y(t)$, i.e., $y(t) = x(t) + v(t)$, $t \in [a, b]$[1] and $v(t)$ is a zero mean white Gaussian noise. It is defined by the following optimization problems

$$\underset{x}{\operatorname{argmin}} \frac{1}{2} \|y - x\|_2^2 + \lambda \left\| x^{(i)} \right\|_1, \tag{1a}$$

$$\underset{x(t)}{\operatorname{argmin}} \frac{1}{2} \int_a^b [y(\tau) - x(\tau)]^2 \, d\tau + \lambda \int_a^b |x^{(i)}(\tau)| d\tau, \tag{1b}$$

where $x^{(i)}$ is the i-th order derivative of x with respect to t and λ is the regularization factor that controls the degree of sparsity of the solution. The optimization problem (1) formulates the sparsity based denoising as the problem of minimizing the ℓ_1 norm of the derivative of x subject to a data fidelity constraint.

[1] x is a continuous function whose domain contains the interval $[a, b]$

Manuscript received 11 September 2021; revised 26 June 2022; accepted 23 August 2022. Date of publication 2 September 2022; date of current version 27 September 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Hassan Mansour. *(Corresponding author: Arman Kheirati Roonizi.)*

Arman Kheirati Roonizi is with the Department of Computer Science, Fasa University, Fasa 74616-86131, Iran, and also with the Dipartimento di Informatica, Università degli Studi di Milano, 20131 Milano, Italy (e-mail: arman.kheirati@unimi.it).

Ivan W. Selesnick is with the Department of Electrical and Computer Engineering, Tandon School of Engineering, New York University, Brooklyn, NY 11201 USA (e-mail: selesi@nyu.edu).

Digital Object Identifier 10.1109/TSP.2022.3203852

The ℓ_1 norm is non-differentiable. Consequently, there is no analytic solution for the optimization problems defined in (1). A possible solution is to replace it with a sequence of simpler ones. The procedure is known as majorization-minimization (MM) method (also known as bound optimization or surrogate optimization method) [15]. An overview of MM algorithms in signal processing, machine learning and communications is presented in [16]. As an application of MM approach, it can be used to convert the optimization problem (1) to a simpler one. To this purpose, the MM approach proposes the following majorizer for the $|x(t)|$ [16], Example 6]:

$$|x(t)| \leq \frac{1}{2}\left(x(t)\Big/\sqrt{|x_k(t)|}\right)^2 + \frac{1}{2}|x_k(t)|.$$

With it, a majorizer of ℓ_1 norm is defined as

$$\left\|x^{(i)}\right\|_1 := \int_a^b \frac{1}{2}\left(x^{(i)}(\tau)\Big/\sqrt{|\hat{x}_k^{(i)}(\tau)|}\right)^2 d\tau + C, \tag{2}$$

where $\hat{x}_k(t)$ is the estimated signal after k iterations and

$$C = \frac{1}{2}\int_a^b |\hat{x}_k^{(i)}(\tau)| d\tau.$$

Note that C does not depend on x. Therefore, in order to solve (1), one can solve the following iterative optimization problem, instead:

$$\underset{x(t)}{\text{argmin}} \frac{1}{2}\int_a^b [y(\tau) - x(\tau)]^2 d\tau + \frac{\lambda}{2}\int_a^b \left(\frac{x^{(i)}(\tau)}{\sqrt{|\hat{x}_k^{(i)}(\tau)|}}\right)^2 d\tau + C \tag{3}$$

with an initialization [e.g., $\hat{x}_0(t) = y(t)$]. Note that the regularization term is now differentiable and x_k is considered as a constant with respect to x. By setting the derivative of (3) with respect to $x(t)$ to zero, we obtain the following linear time-variant (LTV) ordinary differential equation:

$$x(t) + \lambda(-1)^i \frac{x^{(2i)}(t)}{|\hat{x}_k^{(i)}(t)|} = y(t). \tag{4}$$

Proof: Considering the fact that $x^{(i)}(t)$ can be expressed as the convolution of $u_i(t)$ with the signal $x(t)$, where $u_i(t)$ is the i-th order derivative of the Dirac delta function, (3) is expressed as[a]

$$\underset{x(t)}{\text{argmin}} \frac{1}{2}\int_a^b [y(\tau) - x(\tau)]^2 d\tau + \frac{\lambda}{2}\int_a^b \frac{(u_i * x)^2(\tau)}{|\hat{x}_k^{(i)}(\tau)|} d\tau + C \tag{5}$$

where $*$ denotes the convolution operator. It is also proved in [11], see Lemma 2] that

$$\frac{\partial}{\partial x}(u_i * x)^2(t) = 2\left(u_i^r * u_i * x\right)(t),$$

[a]Note that $(u_i * x)(t)$ equals the i-th derivative of x, *i.e.*, $(u_i * x)(t) = x^{(i)}(t)$.

where $u_i^r(t) = u_i(-t)$. Taking the derivative of (5) with respect to x yields

$$x(t) + \lambda \frac{\left(u_i^r * u_i * x\right)(t)}{|\hat{x}_k^{(i)}(t)|} = y(t). \tag{6}$$

Note that Dirac delta function is an even function, *i.e.*, $\delta(-t) = \delta(t)$. Taking derivative of it, we find $-\delta'(-t) = \delta'(t)$, *i.e.*, $u_1(-t) = -u_1(t)$. Taking the derivative again and again, we find

$$u_i(-t) = (-1)^i u_i(t) \tag{7}$$

Convolving both sides of (7) with $u_i(t)$ yields

$$u_i(-t) * u_i(t) = (-1)^i u_{2i}(t) \Leftrightarrow u_i^r(t) * u_i(t) = (-1)^i u_{2i}(t)$$

Substituting $(u_i^r * u_i)(t)$ with $(-1)^i u_{2i}(t)$ and using the fact that $(u_i * x)(t) = x^{(i)}(t)$, (6) can be simplified to

$$x(t) + \lambda(-1)^i \frac{x^{(2i)}(t)}{|\hat{x}_k^{(i)}(t)|} = y(t).$$

□

It means that the MM approach proposes an iterative differential equation with order $2i$ to solve the TV regularization. Similarly, in discrete time (DT) domain, the i-th order TV regularization to estimate x_n from its noisy measurements $y_n = x_n + v_n$, $n = 1, \ldots, L$ is defined as

$$\underset{x_n}{\text{argmin}} \frac{1}{2}\sum_{m=1}^{L} [y_m - x_m]^2 + \frac{\lambda}{2}\sum_{m=1}^{L} \frac{[u_{i,m} * x_m]^2}{|u_{i,m} * \hat{x}_{k,m}|} + C, \tag{8}$$

where $\hat{x}_{k,m}$ is the estimated signal after k iterations, $u_{i,m} = \sum_{l=0}^{i}(-1)^l \binom{i}{l}\delta_{m-l}$ and δ_{m-l} is a shifted version of Dirac delta function. Setting the derivative of (8) with respect to x_n to zero, yields

$$x_n + \lambda \frac{u_{i,n}^r * u_{i,n} * x_n}{|u_{i,n} * \hat{x}_{k,n}|} = y_n, \tag{9}$$

where $u_{i,n}^r = u_{i,-n}$. It can be written as the following difference equation:

$$x_n + \lambda \frac{\sum_{l=0}^{2i}(-1)^l \binom{2i}{l} x_{n-l}}{\left|\sum_{l=0}^{i}(-1)^l \binom{i}{l} \hat{x}_{k,n-l}\right|} = y_n. \tag{10}$$

B. Total Variation and Polynomial Smoothing

The idea of combining the TV regularization and least-square polynomial signal smoothing into a unified problem formulation in order to reconstruct a local polynomial signal and a sparse signal from noisy measurements was first proposed by Selesnick et al. [12]. Assume that the measurement signal $y(t)$ is a noisy additive mixture of a low-order polynomial signal $p(t)$, a piecewise constant signal $x(t)$ and the Gaussian noise $v(t)$, i.e., $y(t) = p(t) + x(t) + v(t)$. Mathematically, $p(t)$ is defined as $p(t) = \sum_{l=0}^{r} a_l t^l$, where a_l is the polynomial coefficients and r is the polynomial order. If the observation signal contains no sparse signal, the signal $p(t)$ can be estimated by minimizing the following cost function:

$$\operatorname*{argmin}_{a_l}\Big\|y-\sum_{l=0}^{r}a_l t^l\Big\|_2^2. \tag{11}$$

However, when a discrete event phenomenon is observed in the presence of polynomial signal, it is preferred to combine (11) and (1) resulting in the following optimization approach [12]:

$$\operatorname*{argmin}_{a_l,x}\frac{1}{2}\Big\|y-\sum_{l=0}^{r}a_l t^l - x\Big\|_2^2+\lambda\Big\|x^{(i)}\Big\|_1. \tag{12}$$

The polynomial coefficients, a_l, and the signal x are unknown. They can be found by performing the minimization with respect to a_l and x.

C. Total Variation and Tikonov Regularization

Let us consider the problem of extraction of a piecewise-smooth signal form its noisy measurements using a combination of Tikhonov and TV regularization. Specially, we assume that the underlying signal comprises a piecewise constant component, $x(t)$, and a smooth component, $f(t)$. A solution to is to synthesize the total variation and Tikhonov regularization for reconstructing the desired signal (total variation for the first component and the ℓ_2 norm for the second component). The following optimization approach can be used:

$$\operatorname*{argmin}_{f,x}\frac{1}{2}\Big\|y-f-x\Big\|_2^2+\frac{\zeta}{2}\Big\|f^{(i)}\Big\|_2^2+\lambda\Big\|x^{(i)}\Big\|_1. \tag{13}$$

The desired components are found by setting the gradient of (13) with respect to f and x to zero. It results a system of non-linear equations with two unknowns f and x. Note that in Tikhonov and TV regularization, the regularization operator is typically an approximate of the i-th order derivative operator. Gholami and Hosseini suggested the following optimization approach [13]:

$$\operatorname*{argmin}_{f,x}\frac{1}{2}\left\|y-f-x\right\|_2^2+\frac{\zeta}{2}\left\|f''\right\|_2^2+\lambda\left\|x'\right\|_1. \tag{14}$$

D. Total Variation and Low-Pass Filtering

The third approach proposed by Selesnick et al. is designed for reconstruction of a mixture of a low-frequency signal and a sparse or sparse-derivative signal from the noisy measurements [14]. It combines LTI filtering and sparsity based denoising in a principled way. The observed signal is considered as a noisy additive mixture of a low-frequency signal $f_{lp}(t)$ and a sparse-derivative signal $x(t)$, i.e., $y(t)=f_{lp}(t)+x(t)+v(t)$. Since f_{lp} is a low-frequency signal, it can be obtained by filtering $y-\hat{x}$ with a low-pass filter, where $\hat{x}$ is an estimate of x. In other words, by filtering $y-\hat{x}$ with a high-pass filter, we can get a white Gaussian process, v. Using it, Selesnick et al. proposed the following optimization problem:

$$\operatorname*{argmin}_{x}\frac{1}{2}\left\|h*(y-x)\right\|_2^2+\lambda\Big\|x'\Big\|_1, \tag{15}$$

where $h(t)$ is the impulse response for a high-pass filter. An estimate of x is obtained by taking the derivative of (15) with respect to x, resulting in

$$(h^r*h*x)(t)-\lambda\frac{x''}{|\hat{x}_k'|}=(h^r*h*y)(t), \tag{16}$$

where $h^r(t)=h(-t)$. Once an estimate of x is computed, the low-frequency component f_{lp} is obtained by applying the low-pass filter to $y-x$, i.e., $\hat{f}_{lp}(t)=y(t)-\hat{x}(t)-h(t)*[y(t)-x(t)]$. Therefore, the components are estimated individually and sequentially.

II. Motivation

The above combination algorithms were proposed to reconstruct a picewise-polynomial, a picewise-smooth or a picewise low-frequency signal in white Gaussian noise. However, they cannot be used for reconstructing signals that comprise the sum of a band-limited component and a sparse derivative signal. Band-limited signals can be categorized in three special classes due to their common occurrence in applications: low-frequency component signal, high-frequency component signal and pass-band component signal. Therefore, there is a need for a new combination algorithm that provides a generic solution to reconstruct a piecewise band-limited signal. In this paper, we address a more general signal reconstruction where the underlying signal comprises a band-limited frequency component and a sparse-derivative component. Specifically, we assume that the observed signal $y(t)$ can be well modelled as $y(t)=f(t)+x(t)+v(t)$, where $f(t)$ is a band-limited signal, $x(t)$ is a sparse-derivative signal, and $v(t)$ is a white Gaussian noise. We seek to estimate the unknown signal components, f and x, from observation y and filter out the observation noise component, v. If the desired signal contains no sparse-derivative signal component (i.e., in the absence of the sparse-derivative signal, x) then the signal $\hat{f}$ can be estimated using conventional zero-phase filters (e.g., zero-phase low-pass, band-pass or high-pass filters). Depending on the predefined frequency band of the desired signal, a specific filter can be employed. If there is no band-limited signal in the desired system, and the signal is sparse-derivative (i.e., in the absence of the band-limited signal, f) then the signal $\hat{x}$ can be estimated using TV regularization. However, in the presence of both components, the former case reconstructs the band-limited signal well but fails to reconstruct the sparse-derivative signal, and it is the other way around for the latter case. Therefore, for such cases, neither conventional zero-phase filtering nor TV denoising is suitable. To overcome this problem, we develop a Bayesian filtering framework that combines zero-phase LTI filtering and TV denoising. The problem considered here is more general than those of [12], [13], [14] and our proposed approach has several advantages over the methods described there:

- The signal f considered in [13] is a smooth signal that requires only a few Fourier coefficients to represent; it is a low-frequency component signal in [14]. However, in this paper, we consider that f can be any class of band-limited signals such as low-frequency component signal, high-frequency component signal and so on.
- For the algorithms proposed in [12], [14], [17], there is a need for change of variable of the sparse-derivative

signal, x, and the change of variables is important, because otherwise the MM approach leads to a dense system of equations. This change of variables is no longer necessary with the approach proposed in this paper.

- Against the existing approaches, the proposed approach estimates the signal components simultaneously.
- One of the major advantages of the proposed method is that it treats continuous-time (CT) and discrete-time (DT) systems in parallel and brings CT and DT systems together in a unified way.
- The existing approaches [12], [13], [14], [17] are implemented non-causally and not suited for real-time applications. In the proposed approach, the Kalman smoother is suited for offline applications, while the Kalman filter can be used for online or real-time applications.

III. Problem Formulation

Let us consider the problem of observing a noisy additive mixture of a band-limited signal $f(t)$ and a sparse-derivative signal $x(t)$ in the model

$$y(t) = f(t) + x(t) + v(t), \tag{17}$$

where $v(t)$ is a white Gaussian noise which is assumed to be uncorrelated with other two signals, i.e., $f(t)$ and $x(t)$. Further, we assume that f is a band-limited signal whose Fourier spectra confined to a finite number of finite intervals (bands) along the real frequency axis. For instance, for a low-frequency signal, there exists a cutoff frequency ω_c such that for $|\omega| \geq \omega_c$, its Fourier transform vanishes, i.e.,

$$\exists\, \omega_c \text{ s.t. } \forall\, |\omega| \geq \omega_c,\ F(j\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt = 0,$$

where $j = \sqrt{-1}$. For a pass-band signal, there exist two cutoff frequencies ω_1 and ω_2 such that for $|\omega| \notin [\omega_1, \omega_2]$, $F(j\omega) = 0$. If for $|\omega| \in [\omega_1, \omega_2]$, we have $F(j\omega) = 0$, then the signal is stop-band. We seek a tracking algorithm that enables the estimation of f and x simultaneously.

IV. Solution

In [11], [18], we showed that forward-backward filtering (zero-phase IIR filters) can be presented as instances of penalized least-squares optimization (PLSO). Consider a zero-phase filter defined by the following Laplace transform

$$|H(s)|^2 = \frac{1}{1 + \Theta(s)\Theta^*(-s)}, \tag{18}$$

where * denotes the conjugate operator and $\Theta(s)$ is a rational function in s with numerators α_l and denominators β_l:

$$\Theta(s) = \frac{\sum_{l=0}^{p} \alpha_l s^l}{\sum_{l=0}^{q} \beta_l s^l}, \quad p > q. \tag{19}$$

The output of the above zero-phase filter, $|H(s)|^2$ to the input $y(t)$ can be computed through the convolution of the impulse response with the input signal: $f(t) = (h^r * h * y)(t)$, where $(h^r * h)(t)$ is the impulse response of the zero-phase filter. While the procedure is usually designed using forward-backward filtering, we showed in [11] that the output signal can be obtained using the following PLSO problem:

$$\underset{f}{\operatorname{argmin}} \|y - f\|_2^2 + \|\theta * f\|_2^2, \tag{20}$$

where $\theta(t)$ is the inverse Fourier transform of (19):

$$\sum_{l=0}^{q} \beta_l s^l \Theta(s) = \sum_{l=0}^{p} \alpha_l s^l \Leftrightarrow \sum_{l=0}^{q} \beta_l \theta^{(l)}(t) = \sum_{l=0}^{p} \alpha_l u_l(t),$$

where $u_l(t)$ is the l-th derivative of $\delta(t)$. In this paper, we seek to simultaneously estimate the signal f and x from the noisy measurements y in the model (17). To that end, we combine the ideas of zero-phase filtering based PLSO and TV regularization in a different way. A general Kalman filter and Kalman smoother approach to reconstruct the sparse derivative band-limited signal from its noisy measurements is presented. In the following, we derive the solution in continuous time (CT) domain, and also show the same path toward deriving the solution for its discrete time (DT) domain.

A. Continuous Time

Combining (1) and (20), the following least-squares optimization problem is proposed to estimate the desired components in (17):

$$\underset{f,x}{\operatorname{argmin}} \frac{1}{2}\|y - f - x\|_2^2 + \frac{1}{2}\|\theta * f\|_2^2 + \lambda \|x^{(i)}\|_1. \tag{21}$$

The first and second penalty terms control the frequency band of f and the sparsity of x, respectively. The cost function (21) is convex but difficult to minimize due to the last term as it is non-differentiable. In order to make it simpler, we employ the MM optimization procedure for the $\|x^{(i)}\|_1$. Using (2), a majorizer of (21) is given by

$$\underset{f,x}{\operatorname{argmin}} \frac{1}{2}\|y - f - x\|_2^2 + \frac{1}{2}\|\theta * f\|_2^2 + \frac{\lambda}{2}\int_a^b \left(x^{(i)} \Big/ \sqrt{|\hat{x}_k^{(i)}|} \right)^2 d\tau + \frac{\lambda}{2}\int_a^b |\hat{x}_k^{(i)}(\tau)| d\tau. \tag{22}$$

Using the following change of variables

$$f(t) = \sum_{l=0}^{q} \beta_l \phi^{(l)}(t), \tag{23}$$

(22) can be expressed as

$$\underset{\phi,x}{\operatorname{argmin}} \frac{1}{2}\Big\| y - \sum_{l=0}^{q} \beta_l \phi^{(l)} - x \Big\|_2^2 + \frac{1}{2}\Big\| \sum_{l=0}^{p} \alpha_l \phi^{(l)} \Big\|_2^2 + \frac{\lambda}{2} \frac{\left\| x^{(i)} \right\|_2^2}{\left\| \hat{x}_k^{(i)} \right\|_1} + \frac{\lambda}{2}\left\| \hat{x}_k^{(i)} \right\|_1 \tag{24}$$

The solution to a filtering problem with the linear system (25) is equivalent/similar to the solution of the optimization problem in (24) [19].

$$\begin{cases} \sum_{l=0}^{p} \alpha_l \phi^{(l)} = w_1(t) \\ x^{(i)} = g_k(t) w_2(t) \\ y(t) = \sum_{l=0}^{q} \beta_l \phi^{(l)}(t) + x(t) + v(t) \end{cases}, \tag{25}$$

where $g_k(t) = \sqrt{|\hat{x}_k^{(i)}|}$, $v(t)$ is the observation noise, $w_1(t)$ and $w_2(t)$ are the additive zero-mean random terms and known as the process (model) noise. It is more convenient to express (25) as

$$\begin{cases} \phi^{(p)} = \sum_{l=0}^{p-1} \zeta_l \phi^{(l)} + \frac{1}{\alpha_p} w_1(t) \\ x^{(i)} = g_k(t) w_2(t) \\ y_k = \sum_{j=0}^{q} \beta_l \phi^{(l)}(t) + x(t) + v(t) \end{cases}, \tag{26}$$

where $\zeta_l = -\alpha_l/\alpha_p$. In order to construct a Kalman filter for estimating ϕ and the sparse signal x, the dynamic equation in (26) needs to be converted to a state-space form. One of the possible state space models that is suitable for using in the Kalman filtering framework is as follows:

$$\begin{cases} \boldsymbol{x}'(t) &= \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}\boldsymbol{w}(t) \\ y(t) &= \boldsymbol{c}\boldsymbol{x}(t) + v(t) \end{cases}, \tag{27}$$

where $\boldsymbol{w} = [w_1, \ w_2]^T$,

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\Upsilon}_{p\times p} & \boldsymbol{0}_{p\times i} \\ \boldsymbol{0}_{i\times p} & \boldsymbol{\Gamma}_{i\times i} \end{bmatrix}, \boldsymbol{\Upsilon} = \begin{pmatrix} \zeta_{p-1} & \zeta_{p-2} & \cdots & \zeta_1 & \zeta_0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{\Gamma} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{x} = \begin{pmatrix} \phi^{(p-1)} & \phi^{(p-2)} & \cdots & \phi & x^{(i-1)} & x^{(i-2)} & \cdots & x \end{pmatrix}^T,$$

$$\boldsymbol{b} = \begin{pmatrix} \frac{1}{\alpha_p} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & g_k(t) & 0 & \cdots & 0 \end{pmatrix}^T,$$

$$\boldsymbol{c} = \begin{pmatrix} 0 & \cdots & 0 & \beta_q & \cdots & \beta_0 & 0 & \cdots & 0 & 1 \end{pmatrix},$$

and $\boldsymbol{0}$ is a null matrix. The CT Kalman filter equations for (27) is given as follows [20], [21]:

$$\begin{cases} \hat{\boldsymbol{x}}(0) = \mathbb{E}\{\boldsymbol{x}(0)\} \\ \boldsymbol{P}(0) = \mathbb{E}\{[\boldsymbol{x}(0) - \hat{\boldsymbol{x}}(0)][\boldsymbol{x}(0) - \hat{\boldsymbol{x}}(0)]^T\} \\ \boldsymbol{K} = \frac{\boldsymbol{P}\boldsymbol{c}^T}{r} \\ \hat{\boldsymbol{x}}' = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{K}(y - \boldsymbol{c}\hat{\boldsymbol{x}}) \\ \boldsymbol{P}' = \frac{-\boldsymbol{P}\boldsymbol{c}^T\boldsymbol{c}\boldsymbol{P}^T}{r} + \boldsymbol{A}\boldsymbol{P} + \boldsymbol{P}\boldsymbol{A}^T + \boldsymbol{b}^T\boldsymbol{Q}\boldsymbol{b} \end{cases}. \tag{28}$$

where r and $\boldsymbol{Q}$ are the covariance of the measurement noise and the covariance of the process noise, respectively and assumed to be known. Note that the Kalman filter estimates all the variable states $[\phi^{(p-1)}, \dots, \ \phi, \ x^{(i-1)}, \ \cdots, \ x]$. However, we are interested in f and x. Once an estimate of $[\phi^{(q)}, \dots, \phi', \phi]$ is computed, an estimate of f is obtained by (23). In the following, we tackle the same problem for DT systems by following the approach parallel to that used in the CT case.

B. Discrete Time Domain

In the previous section a continuous formulation was proposed for simultaneous zero-phase filtering and sparse derivative denoising. In practical applications a discretized version is needed for digitally processing the signal. In this section, we derive a discrete formulation for simultaneous sparse derivative denoising and zero-phase filtering. To this purpose, we assume that (29) is obtained by sampling (17), i.e.,

$$y_n = f_n + x_n + v_n, \qquad n = 1, 2, \dots, L, \tag{29}$$

where $y_n = y[nT_s]$ is the discrete-time samples of $y(t)$, L is the number of samples, T_s is the sampling period, f_n, x_n and v_n are respectively the sampled desired band-limited signal, sparse derivative signal and observation noise. Furthermore, we assume that the discrete time counterpart of (18) is

$$|H(z)|^2 = \frac{1}{1 + \Theta(z)\Theta(\frac{1}{z})}, \tag{30}$$

where $\Theta(z)$ is a polynomial function in z:

$$\Theta(z) = \frac{\sum_{l=0}^{p} \hat{\alpha}_l z^l}{\sum_{l=0}^{q} \hat{\beta}_l z^l}. \tag{31}$$

Consequently, in DT, (21), is expressed as

$$\underset{f_n, x_n}{\text{argmin}} \frac{1}{2} \sum_{n=1}^{L} (y_n - f_n - x_n)^2 + \frac{1}{2} \sum_{n=1}^{L} (\theta_n * f_n)^2 + \lambda \sum_{n=1}^{L} \left| \nabla^i x_n \right|, \tag{32}$$

where $\nabla^i x_n$ is the i-th order difference which is precisely defined by

$$\nabla^i x_n = \sum_{l=0}^{i} (-1)^l \binom{i}{l} x_{n-i+l}$$

and

$$\sum_{l=0}^{q} \hat{\beta}_l \theta_{n+l} = \sum_{l=0}^{p} \hat{\alpha}_l \delta_{n+l}.$$

A majorizer of (32) is given by

$$\underset{f,x}{\text{argmin}} \frac{1}{2} \sum_{n=1}^{L} (y_n - f_n - x_n)^2 + \frac{1}{2} \sum_{n=1}^{L} (\theta_n * f_n)^2 + \frac{\lambda}{2} \sum_{n=1}^{L} \left(\frac{\nabla^i x_n}{\sqrt{|\nabla^i \hat{x}_{k,n}|}} \right)^2 + \frac{\lambda}{2} \sum_{n=1}^{L} |\nabla^i \hat{x}_{k,n}| \tag{33}$$

Using the following change of variables

$$f_n = \sum_{l=0}^{q} \hat{\beta}_l \phi_{n+l}, \tag{34}$$

(33) can be expressed as

$$\frac{1}{2}\sum_{n=1}^{L}\left(y_n - \sum_{j=0}^{q}\hat{\beta}_j\phi_{n+j} - x_n\right)^2 + \frac{1}{2}\sum_{n=1}^{L}\left(\sum_{j=0}^{p}\hat{\alpha}_j\phi_{n+j}\right)^2 + \frac{\lambda}{2}\sum_{n=1}^{L}\left(\frac{\nabla^i x_n}{\sqrt{|\nabla^i \hat{x}_{k,n}|}}\right)^2 + \frac{\lambda}{2}\sum_{n=1}^{L}\left|\sqrt{|\nabla^i \hat{x}_{k,n}|}\right| \tag{35}$$

The linear state-space model corresponding to it can be represented as

$$\begin{cases} \sum_{l=0}^{p} \hat{\alpha}_l \phi_{n+l} = w_{1,n} \\ \nabla^i x_n = g_{k,n} w_{2,n} \\ y_n = \sum_{l=0}^{q} \hat{\beta}_l \phi_{n+l} + x_n + v_n \end{cases}, \tag{36}$$

where $g_{k,n} = \sqrt{|\nabla^i \hat{x}_{k,n}|}$, v_n is the observation noise, $w_{1,n}$ and $w_{2,n}$ are the process (model) noises. It is straightforward to show that (33) is a special case of Wiener smoothing filter (see [19] for more details). Note that for stationary processes, a stable Kalman filter/smoother converges to the smoothing Wiener filter in steady state. The convergence time depends on the covariances of the process and measurement noises or merely to their ratio, $\lambda = |u_{i,n} * \hat{x}_{k,n}|\sigma_v^2/\sigma_{w_2}^2$ [19]. It is more convenient to express (36) as

$$\begin{cases} \phi_{n+p} = \sum_{l=0}^{p-1} \hat{\zeta}_l \phi_{n+l} + \frac{1}{\hat{\alpha}_p} w_{1,n} \\ x_n = \sum_{l=0}^{i-1} \gamma_l x_{n-i+l} + g_{k,n} w_{2,n} \\ y_n = \sum_{l=0}^{q} \hat{\beta}_l \phi_{n+l} + x_n + v_n \end{cases}, \tag{37}$$

where $\gamma_l = (-1)^{l+1}\binom{i}{l}$ and $\hat{\zeta}_l = -\hat{\alpha}_l/\hat{\alpha}_p$. In order to construct a Kalman filter for estimating ϕ_n and the sparse signal x_n, the dynamic equation in (37) needs to be converted to a state-space form. One of the possible state space models is as follows:

$$\begin{cases} \boldsymbol{x}_n = \boldsymbol{A}\boldsymbol{x}_{n-1} + \boldsymbol{b}\boldsymbol{w}_n \\ y_n = \boldsymbol{c}\boldsymbol{x}_n + v_n \end{cases}, \tag{38}$$

where $\boldsymbol{w}_n = [w_{1,n},\ w_{2,n}]$,

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\Upsilon} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Gamma} \end{bmatrix}, \boldsymbol{\Upsilon} = \begin{pmatrix} \hat{\zeta}_{p-1} & \hat{\zeta}_{p-2} & \dots & \hat{\zeta}_1 & \hat{\zeta}_0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{\Gamma} = \begin{pmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{i-2} & \gamma_{i-1} \\ 1 & 0 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{x}_n = (\phi_{n+p}\ \phi_{n+p-1}\ \cdots\ \phi_n\ x_n\ x_{n-1}\ \cdots\ x_{n-i+1})^T,$$

$$\boldsymbol{b} = \begin{pmatrix} \frac{1}{\hat{\alpha}_p} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & g_{k,n} & 0 & \cdots & 0 \end{pmatrix}^T,$$

$$\boldsymbol{c} = \begin{pmatrix} 0 & \cdots & 0 & \hat{\beta}_q & \cdots & \hat{\beta}_0 & g_{k,n} & 0 & \cdots & 0 \end{pmatrix}$$

Note that $\boldsymbol{b}$ is a function of n. The Kalman filter for (38) is given as follows:

Time Update:

$$\begin{cases} \hat{\boldsymbol{x}}_{n+1}^- = \boldsymbol{A}\boldsymbol{x}_n^+ \\ \boldsymbol{P}_{n+1}^- = \boldsymbol{A}\boldsymbol{P}_n^+\boldsymbol{A}^T + \boldsymbol{b}^T\boldsymbol{Q}\boldsymbol{b} \end{cases} \tag{39}$$

Measurement update:

$$\begin{cases} \hat{\boldsymbol{x}}_n^+ = \hat{\boldsymbol{x}}_n^- + \boldsymbol{K}_n\left[y_n - \boldsymbol{c}\hat{\boldsymbol{x}}_n^-\right] \\ \boldsymbol{K}_n = \boldsymbol{P}_n^-\boldsymbol{c}^T\left(\boldsymbol{c}\boldsymbol{P}_n\boldsymbol{c}^T + r_n\right)^{-1} \\ \boldsymbol{P}_n^+ = \boldsymbol{P}_n^- - \boldsymbol{K}_n\boldsymbol{c}\boldsymbol{P}_n^- \end{cases}, \tag{40}$$

where $\boldsymbol{Q}_n \triangleq \mathbb{E}\{\boldsymbol{w}_n^2\}$, $r_n \triangleq \mathbb{E}\{v_n^2\}$, $\hat{\boldsymbol{x}}_n^- \triangleq \mathbb{E}\{\boldsymbol{x}_n|y_{n-1},\ldots,y_1\}$ is the a priori estimate of the state vector $\boldsymbol{x}_n$ in the n-th stage using the observation y_1 to y_{n-1}, and $\hat{\boldsymbol{x}}_n^+ \triangleq \mathbb{E}\{\boldsymbol{x}_n|y_n,\ldots,y_1\}$ is the a posteriori estimate of the state vector after using the n-th observation y_n. The matrices $\boldsymbol{P}_n^- \triangleq \mathrm{E}\{(\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n^-)(\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n^-)^T\}$ and $\boldsymbol{P}_n^+ \triangleq \mathrm{E}\{(\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n^+)(\boldsymbol{x}_n - \hat{\boldsymbol{x}}_n^+)^T\}$ are also defined as the *prior* and *posterior* state covariance matrices, while $\boldsymbol{K}_n$ is the Kalman gain. For smoother results, a Kalman smoother is usually employed after Kalman filter. It consists of a forward Kalman filter stage followed by a backward recursive smoothing stage. Since Kalman smoother uses information brought by "future" observations, it always outperforms the Kalman filter (with the exception of the estimate at the terminal time which is equivalent in the filtering and smoothing posteriors). Once an estimate of $[\phi_n,\ \cdots,\ \phi_{n+q}]$ is computed, an estimate of f_n is obtained by (34).

V. Examples

In this section, we present two examples to illustrate the application of the proposed framework for 1) simultaneous zero-phase low-pass filtering and TV denoising 2) simultaneous band-pass filtering and TV denoising.

A. Simultaneous Low-Pass Filtering and TV Denoising

Recall that a zero-phase low-pass Butterworth filter of order p is defined by [1]:

$$|H_{lf}(s)|^2 = \frac{1}{1 + (\frac{s}{j\omega_c})^{2p}} = \frac{1}{1 + (\frac{s}{\omega_c})^p(\frac{-s}{\omega_c})^p}, \tag{41}$$

where, the subscript lf indicates the low-pass filter and ω_c is the cut-off frequency. (41) can be expressed as

$$|H_{lf}(s)|^2 = \frac{1}{1+\Theta_{lf}(s)\Theta_{lf}(-s)}, \tag{42}$$

where $\Theta_{lf}(s) = (s/\omega_c)^p$. Assume that the low-frequency component signal $f(t)$ can be estimated using (41). It can be obtained using the following optimization problem [11]

$$\underset{f}{\text{argmin}}\, \|y-f\|_2^2 + \left\| f^{(p)} \big/ \omega_c^p \right\|_2^2. \tag{43}$$

According to the proposed framework, the following optimization problem is defined for simultaneous p-th order zero-phase low-pass Butterworth filtering and q-th order TV denoising:

$$\underset{f,x}{\text{argmin}}\, \|y-f-x\|_2^2 + \left\| f^{(p)} \big/ \omega_c^p \right\|_2^2 + \lambda \left\| x^{(i)} \Big/ \sqrt{|\hat{x}_k^{(i)}|} \right\|^2.$$

We consider the following linear state space model:

$$\begin{cases} f^{(p)} = \omega_c^p w_1(t) \\ x^{(i)} = g_k(t) w_2(t), \qquad g_k(t) = \sqrt{|\hat{x}_k^{(i)}|} \\ y(t) = f(t) + x(t) + v(t) \end{cases}$$

In the framework of Kalman filter, it is represented as

$$\begin{cases} \boldsymbol{x}'(t) = \boldsymbol{b}\boldsymbol{w}(t) \\ y(t) = \boldsymbol{c}\boldsymbol{x}(t) + v(t) \end{cases}$$

where $\boldsymbol{w} = [w_1,\ w_2]$, $\boldsymbol{c} = [0,\ \cdots,\ 0,\ 1,\ 0,\ \cdots,\ 0,\ 1]$,

$$\boldsymbol{x} = \begin{pmatrix} f^{(p-1)} & f^{(p-2)} & \cdots & f & x^{(i-1)} & x^{(i-2)} & \cdots & x \end{pmatrix}^T,$$

$$\boldsymbol{b} = \begin{pmatrix} \omega_c^p & 0 & \cdots & 0 \\ g_k(t) & 0 & \cdots & 0 \end{pmatrix}^T$$

In discrete time, the p-th order zero-phase Butterworth filter is defined by [11]

$$|H_{lf}(z)|^2 = \frac{1}{1+\frac{(1-z^{-1})^p(1-z)^p}{(2\sin\frac{\omega_c}{2})^{2p}}}. \tag{44}$$

Similarly, the following optimization problem is defined for simultaneously zero-phase low-pass Butterworth filtering and TV denoising in DT domain:

$$\underset{f,x}{\text{argmin}} \sum_n (y_n - f_n - x_n)^2 + \sum_n \left(\frac{(1-z^{-1})^p f_n}{(2\sin\frac{\omega_c}{2})^p} \right)^2 + \lambda \sum_n \left(\frac{(1-z^{-1})^i x_n}{\sqrt{|(1-z^{-1})^i \hat{x}_{k,n}|}} \right)^2.$$

We consider the following linear state space model:

$$\begin{cases} (1-z^{-1})^p f_n = (2\sin\frac{\omega_c}{2})^p w_{1,n} \\ (1-z^{-1})^i x_n = g_{k,n} w_{2,n}, \quad g_{k,n} = \sqrt{|(1-z^{-1})^i \hat{x}_{k,n}|} \\ y_n = f_n + x_n + v_n \end{cases},$$

which can be represented in the following standard form:

$$\begin{cases} \boldsymbol{x}_n = \boldsymbol{A}\boldsymbol{x}_{n-1} + \boldsymbol{b}\boldsymbol{w}_n \\ y_n = \boldsymbol{c}\boldsymbol{x}_n + v_n \end{cases} \tag{45}$$

where $\boldsymbol{c} = [1,\ 0,\ \cdots,\ 0,\ 1,\ 0,\ \cdots,\ 0]$,

$$\boldsymbol{x}_n = \begin{pmatrix} f_n & f_{n-1} & \cdots & f_{n-p} & x_n & x_{n-1} & \cdots & x_{n-i} \end{pmatrix}^T,$$

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{\Upsilon} & \boldsymbol{0}_{p\times i} \\ \boldsymbol{0}_{i\times p} & \boldsymbol{\Gamma} \end{bmatrix}, \boldsymbol{\Upsilon} = \begin{pmatrix} \zeta_{p-1} & \zeta_{p-2} & \cdots & \zeta_1 & \zeta_0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

$$\boldsymbol{\Gamma} = \begin{pmatrix} \xi_{i-1} & \xi_{i-2} & \cdots & \xi_1 & \xi_0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 & 0 \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}, \zeta_j = (-1)^j \binom{p}{j},$$

$$\xi_j = (-1)^j \binom{i}{j}, \boldsymbol{b} = \begin{pmatrix} (2\sin\frac{\omega_c}{2})^p & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & g_{k,n} \end{pmatrix}^T$$

B. Simultaneous Band-Pass Filtering and TV Denoising

Lets us consider the first order band-pass filter with frequency response defined by

$$H_{bf}(s) = \frac{1}{1+\alpha s + \beta/s}, \tag{46}$$

where subscript bf stands for band-pass filter. The parameters α and β are related to the filter's cutoff frequencies as $\alpha = 1/(\omega_2 - \omega_1)$, $\beta = \omega_1\omega_2/(\omega_2-\omega_1)$, where ω_1 and ω_2 are the cutoff frequencies [22]. Without loss of generality, we assume $\omega_1 < \omega_2$. It is interesting to note that a low-pass filter with cutoff frequency ω_2 can be obtained by setting ω_1 to zero. In this case, we have $\beta = 0$ and $\alpha = 1/\omega_2$. As a result, (46) becomes

$$H_{lf}(s) = \frac{1}{1+s/\omega_2}, \tag{47}$$

where the subscript lf stands for low-pass filter. A zero-phase band-pass filter (or band-pass smoothing filter) can be obtained by multiplying $H_{bf}(s)$ with its conjugate, $H_{bf}(-s)$, resulting in:

$$H_{bsf}(s) = \frac{1}{1-(\alpha s + \beta/s)^2}, \tag{48}$$

where subscript bsf denotes the band-pass smoothing filter. (48) can be expressed as

$$H_{bsf}(s) = \frac{1}{1+\Theta_{bsf}(s)\Theta_{bsf}(-s)}, \tag{49}$$

where $\Theta_{bsf}(s) = \frac{\alpha s^2+\beta}{s}$. Let $f(t)$ be the output of (48) to the input signal, $y(t)$. Using the change of variables $f = \phi'$, the optimal estimate of $\phi(t)$ can be obtained using the following PLSO problem [11]:

$$\underset{\phi}{\operatorname{argmin}}\frac{1}{2}\left\|y-\phi'\right\|_2^2+\frac{1}{2}\left\|\alpha\phi''+\beta\phi\right\|_2^2. \tag{50}$$

Note that by setting β to zero, the optimization problem (50) can be used for designing a low-pass smoothing filter. According to the proposed framework, the following optimization problem is defined for picewise zero-phase band-pass filtering:

$$\underset{\phi,x}{\operatorname{argmin}}\frac{1}{2}\left\|y-\phi'-x\right\|_2^2+\frac{1}{2}\left\|\alpha\phi''+\beta\phi\right\|_2^2+\lambda\left(x'\Big/\sqrt{|\hat{x}_k'|}\right)^2$$

The following linear state space model is considered:

$$\begin{cases}\alpha\phi''+\beta\phi=w_1(t)\\ x'=g_k(t)w_2(t)\\ y(t)=\phi'+x(t)+v(t)\end{cases},$$

where $g_k(t)=|x_k'(t)|$. We use the following state space model in the framework of Kalman filter

$$\begin{cases}\boldsymbol{x}'=A\boldsymbol{x}+\boldsymbol{b}\boldsymbol{w}(t)\\ y(t)=\boldsymbol{c}\boldsymbol{x}(t)+v(t)\end{cases}$$

where $\boldsymbol{x}=[\phi',\ \phi,x]^T$, $\boldsymbol{c}=[1,\ 0,\ 1]^T$,

$$A=\begin{pmatrix}0&-\frac{\beta}{\alpha}&0\\1&0&0\\0&0&0\end{pmatrix},\boldsymbol{b}=\begin{pmatrix}\frac{1}{\alpha}&0&0\\0&0&g_k(t)\end{pmatrix}^T$$

In discrete time, the zero-phase band-pass filter using bilinear transform is defined by [22]

$$|H_{bsf}(z)|^2=\frac{1}{1+\Theta_{bsf}(z)\Theta_{bsf}(\frac{1}{z})}, \tag{51}$$

where $\Theta_{bsf}(z)=\alpha\frac{1-z^{-1}}{1+z^{-1}}+\beta\frac{1+z^{-1}}{1-z^{-1}}$ and the parameters are $\alpha=1/|\tan\frac{\omega_1}{2}-\tan\frac{\omega_2}{2}|$ and $\beta=\tan\frac{\omega_1}{2}\tan\frac{\omega_2}{2}/|\tan\frac{\omega_1}{2}-\tan\frac{\omega_2}{2}|$ [22]. According to the proposed framework, using the change of variable $f_n=\phi_n-\phi_{n-2}$, the following optimization problem is defined for the first order picewise zero-phase band-pass filtering:

$$\begin{aligned}\underset{\phi,x}{\operatorname{argmin}}&\sum_k\left(y_n-\phi_n+\phi_{n-2}-x_n\right)^2+\sum_k([\alpha+\beta]\phi_n\\&-2[\alpha-\beta]\phi_{n-1}+[\alpha+\beta]\phi_{n-2})^2\\&+\lambda\sum_k\left(\frac{x_n-x_{n-1}}{\sqrt{|\hat{x}_{k,n}-\hat{x}_{k,n-1}|}}\right)^2\end{aligned} \tag{52}$$

The linear state space model for (52) is

$$\begin{cases}\phi_n=2\frac{\alpha-\beta}{\alpha+\beta}\phi_{n-1}-\phi_{n-2}+\frac{1}{\alpha+\beta}w_{1,n}\\ x_n=x_{n-1}+g_{k,n}w_{2,n},\quad g_{k,n}=\sqrt{|\hat{x}_{k,n}-\hat{x}_{k,n-1}|}\\ y_n=\phi_n-\phi_{n-2}+x_n+v_n\end{cases},$$

which can be represented in the following standard form:

$$\begin{cases}\boldsymbol{x}_n=A\boldsymbol{x}_{n-1}+\boldsymbol{b}\boldsymbol{w}_n\\ y_n=\boldsymbol{c}\boldsymbol{x}_n+v_n\end{cases}$$

where $\boldsymbol{x}_n=[\phi_n,\ \phi_{n-1},\ \phi_{n-2},\ x_n]^T$, $\boldsymbol{c}=[1,\ 0,\ -1,\ 1]^T$,

$$A=\begin{pmatrix}2\frac{\alpha-\beta}{\alpha+\beta}&-1&0&0\\1&0&0&0\\0&1&0&0\\0&0&0&1\end{pmatrix},\ \boldsymbol{b}=\begin{pmatrix}\frac{1}{\alpha+\beta}&0&0&0\\0&0&0&g_{k,n}\end{pmatrix}^T,$$

Finally, after estimating $[\phi_n,\ \phi_{n-1},\ \phi_{n-2},\ x_n]$, f_n is computed using $f_n=\phi_n-\phi_{n-2}$.

VI. Simulation Examples

Two examples are given in this section. As a preliminary remark, we mention that the value of λ is related to the ratio of observation and process noise covariances. Therefore, one can estimate the variance of the observation noise, σ_v^2, and compute the variance of process noise based on the arbitrary value of λ. There are also several ways to estimate σ_v^2. One can estimate the noise power from the deviations of the portions of the signal. Another approach is to use the online approaches suggested in [23]. Nevertheless, σ_v^2 represents the degree of reliability of the observation. In other words, when a rather precise measurement of the states of a system is valid r_n is small, and the Kalman filter gain is adapted so as to rely on the measurement. While for the epochs where the measurements are too noisy, r_n is large and the Kalman filter ends to follow its internal dynamics rather than tracking the observations.

A. Example 1

In the first example, we employ the Kalman filter/smoother proposed in Section V-A to simultaneously estimate a piecewise and a low-pass signal in white Gaussian noise. See Fig. 1, for an example, where Fig. 1(a)–(c) show a piecewise, a low-pass signal and the sum of those signals contaminated with white Gaussian noise. Fig. 1(d)–(f) show the result of simultaneous first order low-pass filtering and first order TV denoising using the proposed framework. In order to improve the low-frequency signal estimate, the order of low-pass filtering can be increased. Fig. 2 shows the results of simultaneous second order low-pass filtering and first order TV denoising. We see that the estimated low-frequency component using higher order is closer to the original signal. It is mentionable that the system (45) is not observable. Note that the objective of TV denoising is to estimate a piecewise polynomial signal in white Gaussian noise. A piecewise polynomial signal may contain a low frequency component with sharp edges. Consequently, when TV denoising and low-pass filtering are simultaneously used to estimate a piecewise polynomial and a low frequency component, the low frequency component is tracked by both TV denoising and low-pass filtering. In other words, the sparse state (x_n) and the low-pass state (f_n) in (45) are not distinguishable by only the measurement y_n. The result of this paper confirms this problem. Specially, the state-space model (45) is not observable.

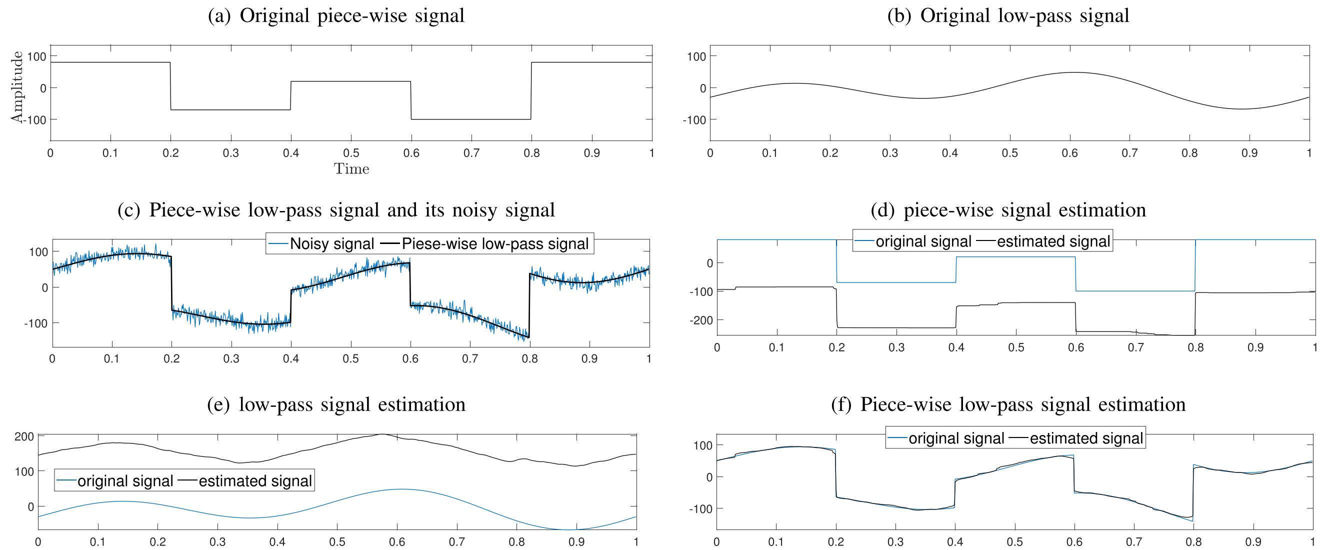

Fig. 1. Piece-wise low-pass signal smoothing using a combination of first order low-pass filtering and first order TV denoising (a) Original piece-wise signal. (b) Original low-pass signal. (c) Original piece-wise low-pass signal and its noisy signal (SNR = 15) (d) Estimated piece-wise signal. (e) Estimated low-pass signal. (f) Sum of the estimated piece-wise and low-pass signals.

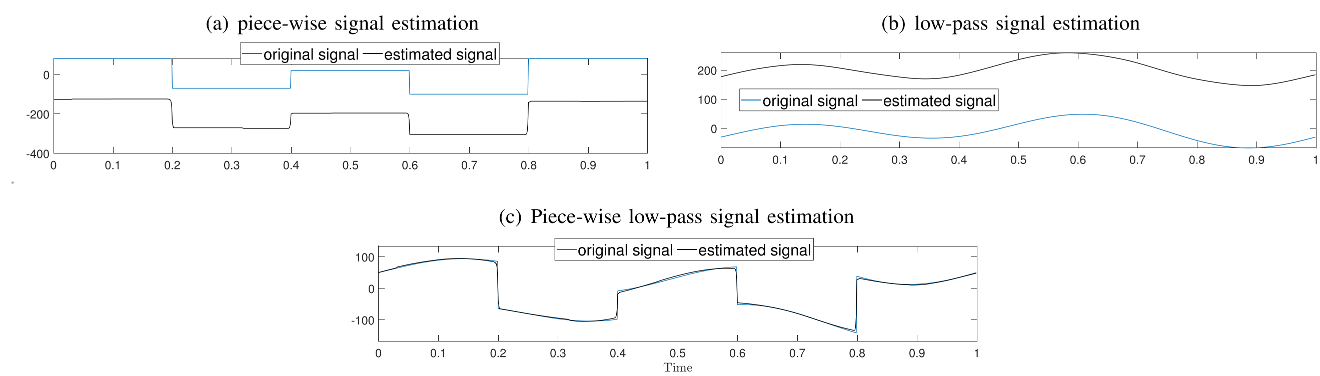

Fig. 2. Piece-wise low-pass signal smoothing using a combination of first order low-pass filtering and second order TV denoising (a) Estimated piece-wise signal. (b) Estimated low-pass signal. (c) Sum of the estimated piece-wise and low-pass signals.

B. Example 2: ECG Analysis

As a real application, the proposed framework is used to design a simultaneous band-pass smoothing and TV denoising to filter the real electrocardiogram (ECG) signal. The designed sparse band-pass smoothing filter is employed to reduce the influence of two noises in the ECG namely power-line interference (PLI) and high-frequency random noise. There are a lot of methods in the literature that can be used for canceling the PLI and removing the random noise in ECG signals [24], [25]. However, the PLI and random noise removal are obtained individually in most of the existing technologies. In this section, we design a sparse ban-pass smoothing filter that is a combination of a band-pass smoothing filter and a sparsity based denoising algorithm. In this method, the ECG signal is modelled as a sparse order-3 derivative (i.e., approximately piecewise polynomial with polynomial segments of order 2) [17], [26] and the PLI signal is modelled as a single tone with frequency of 50 or 60 Hz. Let us consider a continuous-time PLI signal which may be considered as a single sinusoid with an arbitrary amplitude, ρ, and phase, θ:

$$f(t) = \rho \cos(\omega_0 t + \theta), \tag{53}$$

where ω_0 is the PLI frequency.

Taking twice derivative of (53), we find

$$f'' = -\omega_0^2 f. \tag{54}$$

Combining the model of PLI and the model of the ECG, the following dynamical model is suggested for simultaneous band-pass smoothing filter and sparse denoising:

$$\begin{cases} f'' = -\omega_0^2 f + w_1(t) \\ x''' = \sqrt{|x_k'''(t)|} w_2(t) \\ y(t) = f(t) + x(t) + v(t) \end{cases} \tag{55}$$

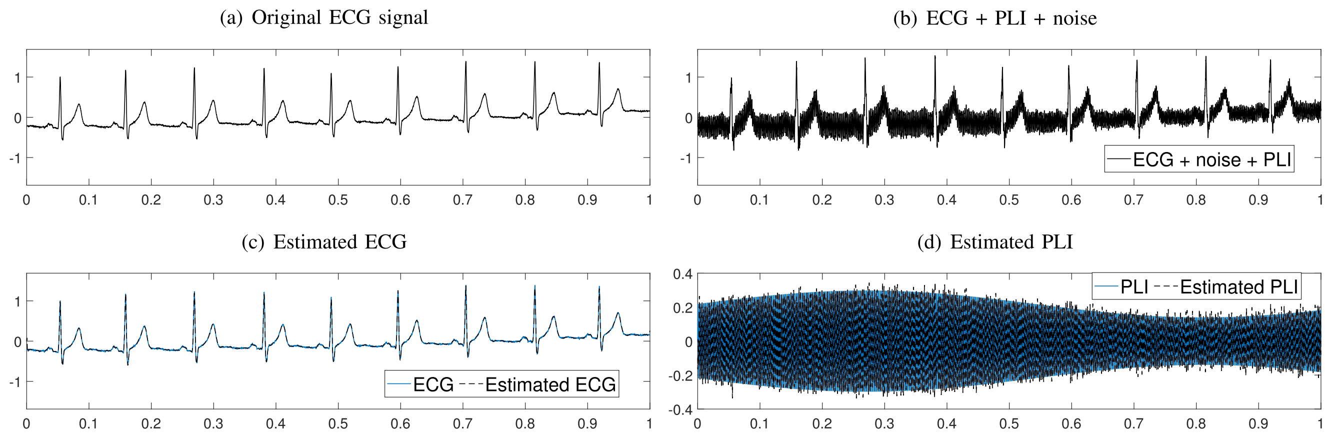

Fig. 3. Denoised ECG signal using the Kalman smoother (a) Original ECG signal. (b) Noisy ECG signal (ECG + PLI + Noise) (c) Estimated ECG signal. (d) Estimated PLI.

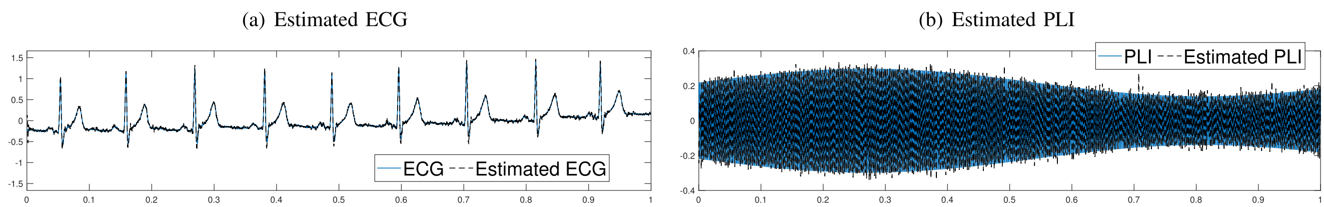

Fig. 4. Real-time ECG denoising using Kalman filter (a) Estimated ECG signal. (e) Estimated PLI.

where $f(t)$ is the PLI signal and $x(t)$ is the ECG signal. While (55) is continuous, we need a discrete-time model. A discrete-time PLI signal can be represented as

$$f_n = \rho\cos(\Omega_0 n + \theta). \tag{56}$$

Using trigonometric manipulation, (56) can be represented as

$$f_n = 2\cos(\Omega_0) f_{n-1} - f_{n-2}. \tag{57}$$

Putting (57) together with the TV denoising sate-space model, we have

$$\begin{cases} f_n = 2\cos(\Omega_0) f_{n-1} - f_{n-2} + w_{1,n} \\ x_n = 3x_{n-1} - 3x_{n-2} + x_{n-3} + \\ \sqrt{|\hat{x}_{k,n} - 3\hat{x}_{k,n-1} + 3\hat{x}_{k,n-2} - \hat{x}_{k,n-3}|}\, w_{2,n} \\ y_n = f_n + x_n + v_n \end{cases} \tag{58}$$

The hidden states of (58) is then estimated using Kalman filter/smoother approach. As an example, in Fig. 3, the proposed Kalman smoother is used to filter the PLI and high frequency noise from ECG, where the original signal is plotted in Fig. 3(a), the ECG diluted by PLI and high frequency noise is plotted in Fig. 3(b). The estimate ECG and PLI using the proposed Kalman smoother are respectively shown in Fig. 3(c) and (d). We mention again that the way TV denoising is usually implemented in the literature is in terms of the whole data, not as a causal system/algorithm which is needed for real-time applications. That is why the proposed approaches in [12], [13], [14] are implemented non-causally. In contrast to the above approaches, the proposed Kalman filter estimates the current states using a recursive estimator. In the recursive estimator, only the estimated state from the previous time step and the current measurement are needed to compute the estimate for the current state [see (39) and (40)]. Therefore, one of the main advantages of the proposed framework is that the Kalman filter can be used for the particular implementation of the proposed approaches to be real-time. As an illustration, the estimated ECG and PLI provided by Kalman filter for the previous example, is plotted in Fig. 4. From a practical point of view, Kalman smoother provides better estimate than Kalman filter. It is because the Kalman smoother uses all the measurements (past, current and future samples) while the Kalman filter only uses the past and current measurements to estimate the signal of interest. We tested the approach over the PhysioNet PTB Diagnostic ECG Database [27] which contains 549 records from 290 subjects. Each record consists of twelve conventional ECG leads plus the three Frank's ones, sampled at 1 kHz with 16-bit resolution. Synthetic PLI was generated and added to each ECG record. To represent the respiratory coupled changes in the PLI amplitude, we modulated the PLI

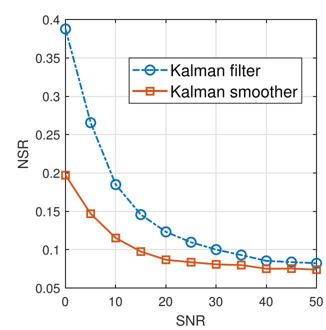

Fig. 5. Mean values of NSR for signal reconstruction by Kalman filter and Kalman smoother as a function of SNR.

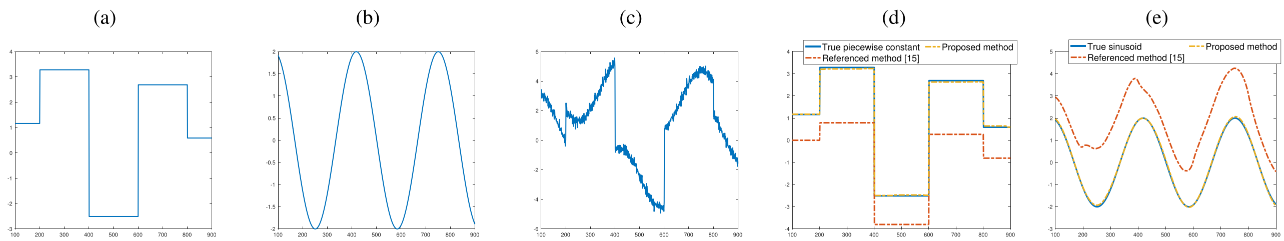

Fig. 6. Simultaneous piecewise constant and sinusoidal signal tracking (a) Piecewise constant signal (b) Sinusoidal signal (c) Noisy signal: piecewise constant plus sinusoidal signal plus white Gaussian noise (d) Piecewise constant signal tracking (e) Sinusoidal tracking.

amplitude with a 0.2 Hz sinusoid. Finally, the Gaussian noise was added with varying SNR (from 0 to 50 dB). The ECG and PLI were estimated from the noisy ECGs using the proposed approaches. We compared the proposed Kalman filter and Kalman smoother on this problem. To quantify the performance of the two methods, we employed NSR $= \sqrt{\sum_n (x_n - \hat{x}_n)^2 / \sum_k x_n^2}$ which is the ratio between the power of the reconstruction error and the power of the desired signal [28], [29]. The results of the ECG reconstruction procedures using this metric are reported in Fig. 5. As expected the proposed Kalman smoother outperformed Kalman filter. Finally, we mention that compared to the referenced ones [12], [13], [14], this paper proposes a more general framework. For instance, consider the problem of estimating a signal comprises of a piecewise constant and a sinusoid in white Gaussian noise. Fig. 6(a)–(c), respectively shows a piecewise constant, a low-frequency sinusoidal signal and sum of those signals contaminated with white Gaussian noise. Therefore, a combination of a low-pass filter and TV denoising (the referenced method in [14]) can be employed to estimate both signal components. However, the main problem with the method [14] is that the piecewise constant signal contains a low-frequency component (the DC component) with edges. Therefore, the DC component will be tracked with both TV and lowpass filtering. As a result, employing a combination of a lowpass filter and TV denoising is not the best solution. The best solution is to use a combination of a TV denoising and a bandpass filter as proposed in Section V-B. We used our method and the referenced method in [14] to estimate the piecewise constant and the sinusoidal signal. The results are shown in Fig. 6(d)–(e). The superiority of the proposed method is evident.

VII. Conclusion

This paper proposed a framework for the combination of zero-phase IIR filtering and total variation denoising. In the framework, while the desired signal is considered to be a mixture a band-limited component and a sparse-derivative component, its components are estimated using an iterative Kalman filter or Kalman smoother approach. The application of the proposed framework to design of sparse derivative low-pass, band-pass and band-stop smoothing filter was discussed in this paper. Although the techniques proposed in this paper focused on the combination of zero-phase filtering and TV denoising, their principles are easy to extend to the combination of autoregressive moving average (ARMA) signal smoothing filters [29] and TV denoising.

References

[1] A. Oppenheim and R. Schafer, *Discrete-Time Signal Processing*. London, U.K.: Pearson Education, 2011.

[2] J. Smith, *Introduction to Digital Filters: With Audio Applications*, Charleston, SC, USA: Booksurge Llc, 2006, vol. 2.

[3] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM J. Sci. Comput.*, vol. 22, pp. 503–516, 2000.

[4] Y. Hu and M. Jacob, "Higher degree total variation (HDTV) regularization for image recovery," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2559–2571, May 2012.

[5] Y. Hu, G. Ongie, S. Ramani, and M. Jacob, "Generalized higher degree total variation (HDTV) regularization," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2423–2435, Jun. 2014.

[6] A. Chambolle, D. Cremers, and T. Pock, "A convex approach to minimal partitions," *SIAM J. Imag. Sci.*, vol. 5, pp. 1113–1158, 2012.

[7] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 231–241, Feb. 2001.

[8] M. Vetterli, P. Marziliano, and T. Blu, "Generalized higher degree total variation (HDTV) regularization," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2014.

[9] L. Condat, "Discrete total variation: New definition and minimization," *SIAM J. Imag. Sci.*, vol. 10, pp. 1258–1290, 2017.

[10] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, 1992.

[11] A. Kheirati Roonizi and C. Jutten, "Forward-backward filtering and penalized least-squares optimization: A unified framework," *Signal Process.*, vol. 178, 2021, Art. no. 107796.

[12] I. W. Selesnick, S. Arnold, and V. R. Dantham, "Polynomial smoothing of time series with additive step discontinuities," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6305–6318, Dec. 2012.

[13] A. Gholami and S. M. Hosseini, "A balanced combination of TIKHONOV and total variation regularizations for reconstruction of piecewise-smooth signals," *Signal Process.*, vol. 93, no. 7, pp. 1945–1960, 2013.

[14] I. W. Selesnick, H. L. Graber, D. S. Pfeil, and R. L. Barbour, "Simultaneous low-pass filtering and total variation denoising," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1109–1124, Mar. 2014.

[15] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[16] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[17] I. W. Selesnick, *Sparsity-Assisted Signal Smoothing*. Cham, Switzerland: Springer, 2015, pp. 149–176.

[18] A. Kheirati Roonizi and C. Jutten, "Band-stop smoothing filter design," *IEEE Trans. Signal Process.*, vol. 69, pp. 1797–1810, 2021.

[19] A. Kheirati Roonizi, "ℓ_2 and ℓ_1 trend filtering: A Kalman filter approach," *IEEE Signal Process. Mag.*, vol. 38, no. 6, pp. 137–145, Nov. 2021.

[20] D. Simon, *Optimal State Estimation: Kalman, H. Infinity, and Nonlinear Approaches*. Hoboken, NJ, USA: Wiley, 2006.

[21] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

[22] A. Kheirati Roonizi and H. Pakniat, "A simple and effective feedback structure for variable-q filter design," *Analog Integr. Circuits Signal Process.*, vol. 107, pp. 463–473, 2021, [Online]. Available; https://doi.org/10.1007/s10470-020-01793-9

[23] P. Lander and E. Berbari, "Time-frequency plane wiener filtering of the high-resolution ECG: Background and time-frequency representations," *IEEE. Trans. Biomed. Eng.*, vol. 44, no. 4, pp. 247–255, Apr. 1997.

[24] S. Chatterjee, R. S. Thakur, R. N. Yadav, L. Gupta, and D. K. Raghuvanshi, "Review of noise removal techniques in ECG signals," *IET Signal Process.*, vol. 14, no. 9, pp. 569–590, 2020.
[25] C. Levkov, G. Mihov, R. Ivanov, I. Daskalov, I. Christov, and I. Dotsinsky, "Removal of power-line interference from the ECG: A review of the subtraction procedure," *Biomed. Eng. Online*, vol. 4, pp. 4–50, 2005.
[26] X. Wang, Y. Zhou, M. Shu, Y. Wang, and A. Dong, "ECG baseline wander correction and denoising based on sparsity," *IEEE Access*, vol. 7, pp. 31573–31585, 2019.
[27] A. L. Goldberger et al., "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. e215–e220, 2000.
[28] E. Kheirati Roonizi and R. Sassi, "A signal decomposition model-based bayesian framework for ECG components separation," *IEEE. Trans. Signal Process.*, vol. 64, no. 3, pp. 665–674, Feb. 2016.
[29] A. Kheirati Roonizi, "A new approach to ARMAX signals smoothing: Application to variable-q ARMA filter design," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4535–4544, Sep. 2019.

Arman Kheirati Roonizi received the B.Sc. degree in computer engineering, the M.Sc. degree in bioelectrical engineering from Shiraz University, Shiraz, Iran, in 2006 and 2011, respectively, and the Ph.D. degree in computer science from the University of Milan, Milan, Italy, in 2017. He is currently an Assistant Professor with the Department of Computer Science, Fasa University, Fasa, Iran, and is also a Research Fellow with the University of Milan. After graduating, he worked as a Postdoctoral Research Fellow with the Department of Image and Signal Processing, Gipsa-Lab, CNRS, University Grenoble Alpes, Grenoble, France. His research interests include statistical signal processing with applications to biomedical signal and especially for cardiac modeling and simulation, ventricular repolarization, sparse optimization algorithms, and Kalman filtering. He is working on aspects related to modeling, filtering, separation, and analysis of cardiac signal waveforms and multimodal recordings.

Ivan W. Selesnick (Fellow, IEEE) received the B.S., M.E.E., and Ph.D. degrees in electrical engineering from Rice University, Houston, TX, USA, in 1990, 1991, and 1996, respectively. He is with the Department of Electrical and Computer Engineering with New York University, New York, NY, USA, with the Tandon School of Engineering, where he is the Department Chair. His research interests include signal and image processing, wavelet-based signal processing, sparsity techniques, and biomedical signal processing. In 1997, he was a Visiting Professor with the University of ErlangenNurnberg, Erlangen, Germany. He was the recipient of the Jacobs Excellence in Education Award from Polytechnic University in 2003 and became an IEEE Fellow in 2016. He has been an Associate Editor for the several IEEE TRANSACTIONS and IEEE SIGNAL PROCESSING LETTERS.

Open Access provided by 'Università degli Studi di Milano' within the CRUI-CARE Agreement