Band-stop Smoothing Filter Design

Arman Kheirati Roonizi, Christian Jutten

HAL Id: hal-03147947
https://hal.science/hal-03147947

Band-stop Smoothing Filter Design

Arman Kheirati Roonizi*, and Christian Jutten, *Fellow, IEEE*

Abstract—Smoothness priors and quadratic variation (QV) regularization are widely used techniques in many applications ranging from signal and image processing, computer vision, pattern recognition, and many other fields of engineering and science. In this contribution, an extension of such algorithms to band-stop smoothing filters (BSSFs) is investigated. For designing a BSSF, the most important parameters are the order and the cut-off frequencies. In this paper, we show that with the optimization approaches (smoothness priors or QV regularization), the cutoff frequencies are related to the regularized parameters and the order can be directly (and easily) controlled with the number of derivatives. We describe two ways to implement the BSSFs using these approaches. First, we present a parallel structure to BSSF and then illustrate why it is less than ideal. Next, we present a novel approach regarding parallel structure to produce BSSFs with very sharp transition bands for high-performance applications. An improved optimization-based approach to BSSF design is introduced. The performance of the new BSSFs is nearly ideal.

Index Terms—Smoothness priors, Quadratic variation regularization, Least-squares optimization, Band-stop smoothing filter, Parallel structure

I. Introduction

THERE is no doubt that smoothing filters (i.e., smoothness priors and quadratic variation) have a respectable place in signal and image processing, computer vision and pattern recognition, statistics, time series analysis, economics and many other fields of engineering and science. Many papers and a number of books have appeared. In the following, some major contributions that have been developed in the design of smoothing filters are highlighted.

The conceptual structure of smoothing filter dates from 1899 with Bohlmann's work who studied the application of regularized method for time series smoothing [1]. In 1923, Whittaker addressed the problem of estimating a smooth trend embedded in white noise and spoke about the method of graduating data [2]. Therefore, his method is known as the method of graduating data [3]. At the same time, Henderson had presented a different solution which was widely extended in North America [4]. So the method of graduating data is also known as Whittaker-Henderson (WH) graduation [5]. In 1973, Shiller worked on the same problem and introduced the notation of "smoothness priors" [6]. It seems that he was not aware of either Whittaker-Henderson graduation or Bohlmann's work. Subsequently, smoothness priors became popular in research communities. It was used in many applications.

In time series analysis, smoothness priors was used for modeling nonstationary time series [7], [8], generalizing a nonparametric estimator [9], [10] and exploring large-scale time series [11]. Some developments, extensions and the least-squares computational framework of smoothness priors in time series analysis are presented in [12], [13].

In signal and image analysis, smoothness priors was used for signal smoothing and detrending [14]–[18], image smoothing and restoration [19], [20], spectral estimation [21] and spline smoothing [9], [22].

In pattern recognition, the second-order priors was used for surface reconstruction [23], [24], dense stereo [25] and global stereo reconstruction [26].

Some other connections to smoothness priors are quadratic variation (QV) regularization [27], Hodrick-Prescott filtering [28], Savitzky-Golay filter [29] and the ill-posed inverse problems of statistical Tikhonov regularization [30].

The smoothing approach based QV regularization or smoothness priors is based on a penalized least-squares optimization which depends on a regularization factor that needs to be designed and weighted properly for getting good performance. There are several methods for designing the regularization factor and its weight (usually denoted as an hyperparameter). Some of them are briefly presented below.

- The L-curve method is a well-known heuristic method for choosing the regularization parameter for ill-posed problems by selecting it according to the maximal curvature of the L-curve [31].
- Generalized cross validation is a popular tool for calculating the parameters in the context of inverse problems and regularization techniques. It was initially used to tune the smoothness parameter in ridge regression and smoothing splines [32].
- Discrepancy principle is a posteriori parameter choice strategies for Tikhonov regularization for solving non-linear ill-posed problems [33].
- Maximum likelihood estimation (MLE) algorithm which is based on the concept of Bayesian likelihood, was first introduced by Akaike to determine the smoothness trade-off parameter [34].
- Stein's unbiased risk estimate (SURE) regularization is another approach that uses the Jacobian matrix of the reconstruction operator with respect to the data [35].
- Regularization factor design in terms of the cutoff frequency. Recently, we have derived a rather simple closed-form expression for the frequency response of the smoothness priors and proposed a closed-form expression for the regularization factor. The design parameter was calculated in terms of the cutoff frequency [36].

For extracting a signal within a predetermined frequency band, conventional linear filtering (digital filtering) techniques are

*A. Kheirati Roonizi is with the Department of Computer Science, Fasa University, Daneshjou blvd, 74616-86131 Fasa, Iran (e-mail: ebad.kheirati.roonizi@gmail.com)

Christian Jutten is with Gipsa-Lab, CNRS, Univ. Grenoble Alpes, France

commonly used [37]. There are many papers and many ways for designing digital filters developed over the past decades. This technique is performed through linear time-invariant (LTI) systems which are characterized in time/frequency domain by their impulse/frequency response. The frequency response is the Fourier transform of the system's impulse response.

In [17], [38], [39], it is shown that smoothness priors or QV regularization can be designed for linear filtering or smoothing. In other word, there is an equivalence between penalized least-squares optimization and zero-phase filtering. Especially, in a series of our recent papers, a general approach based on a regularized least-squares optimization framework has been introduced to signal denoising/smoothing [36], [38], [40] that shows there is a strong connection between linear filtering techniques and the current approach based on regularized least-squares optimization. We note that the main difference between them is that the designed filters based on regularized least-squares optimization handles smoothing/denoising directly in the time domain while in the conventional linear filtering techniques, the desired filter is first designed in the frequency domain and then implemented in the time domain using convolution operator. In the following, we briefly present the main results regarding the connection between penalized least-squares optimization and filtering reported in [36], [38], [40].

In [40], a general framework based on a regularized least-squares optimization has been introduced to signal smoothing when the signal is represented by an autoregressive moving average (ARMA) model. As an application, the framework was used for ARMA filter design. It was shown that the framework can be driven from a forward-backward filtering, which is accomplished through LTI system. In [40], the design of a variable-Q [1] notch filter design has been discussed. Recent comparisons with many classical methods such as simple notch filter or narrow-band notch filter using feedback structure proposed by Pei *et. al.* [41] showed the superiority of the current approach based on optimization with smoothing constraints.

In [38], a framework for unification of the penalized least-squares optimization and forward-backward filtering scheme was presented. We showed that forward-backward filtering (zero-phase IIR filters) can be presented as instances of penalized least-squares optimization. Against conventional linear filtering that uses an IIR filter in its forward pass (followed by time reversing, filtering the reversed signal with the same IIR filter and finally flipping the result), the approach presented in [38] uses an FIR matrix filter in its forward pass which has the advantage of being inherently stable. Comparison with classical linear filtering showed the superiority of the regularized approach. Therefore, the regularized least-squares optimization techniques (e.g., smoothness priors or QV regularization strategies) and zero-phase IIR filters can be conflated into one topic. In this paper, we propose a regularized least-squares optimization approach for BSSF design. The

optimization problem used in this work is a linear least-squares optimization (or a convex optimization) problem. It can be solved by equating its derivative with respect to each unknown variable to zero. The approach is different from linear time invariant (LTI) filter and matrix filter which can be used to design BSSF. For more details on matrix filters, we refer the interested readers to [42]. In [42], the design and implementation of causal recursive filters in terms of banded matrices are discussed. It is based on the fact that a discrete-time filter is described by the difference equation. In this paper, we consider the problem of filtering finite-length signals, since signal processing problems are generally formulated in terms of finite-length signals, and the developed algorithms are targeted for the finite-length case. In [36], we have shown that the regularized least-squares optimization techniques are also suitable for signal recovering even if the signal of interest is (approximately) restricted to a known frequency band. We introduced a new way to the design and implementation of smoothness priors and QV regularization and discussed its application to low-pass, high-pass and band-pass smoothing filter design. Building on that result, in this paper, we mainly focus on the design of band-stop smoothing filter (BSSF). In [36], we have proposed a simple parallel structure for band-pass smoothing filter (BPSF) or BSSF design. The simple band-stop smoothing filter is made out of a low-pass smoothing filter and a high-pass smoothing filter by connecting the two smoothing filter sections in parallel with each other. As it will be shown in Section III, the simple parallel structure has some limitations for BSSF design. Especially, its cutoff frequencies are not exactly mapped on the arbitrary cutoff frequencies and the amplitude at its center frequency is not exactly zero. Furthermore, it has a shallow frequency transition band which is another limitation of the simple parallel structure. It is known that the filters with steep frequency transition bands can better separate signal and noise components in adjacent frequency bands than filters with a shallow frequency transition band. To solve these issues, we propose an improved parallel structure to BSSF design. In the proposed approach, we consider a modified model and investigate the role of noise correlation which represents the link between two smoothing filter sections. Our results suggest that the noise correlation increases the steepness of the role-off and boosts the performance. First we briefly review the simple method described in [36] for BSSF design and then show its limitations. Next, we present an improved parallel structure which yields nearly ideal BSSF: its cutoff frequencies are exactly mapped on the arbitrary cutoff frequencies and it has a sharp frequency transition band.

The remainder of the paper is organized as follows. In Section II, the relevant background on smoothness priors or QV regularization is reviewed. In Section III, we present a naive BSSF using smoothness priors or QV regularization approach and show its limited performance. Section IV presents an extension of such algorithms to BSSFs design, which can achieve nearly optimal performance. In Section V, the proposed method is used to electrocardiogram (ECG) signal denoising and compared with conventional filters such as zero-phase Butterworth filter. It is shown that the proposed

---

[1]The quality factor or Q-factor of a filter is defined as the ratio of its center frequency over its bandwidth.

BSSF outperforms band-stop Butterworth filter. Finally, some concluding remarks are presented in Section VI-B.

Throughout this paper, we use the following notations: boldface uppercase/lowercase letters are used to denote matrices/vectors and lowercase letters for scalars. The subscript k stands for discrete time index while $(\cdot)^T$, $(\cdot)^{-1}$ and $(\cdot)^{\boxdot}$ denote the matrix transpose, matrix inverse and deconvolution, respectively. Symbol D^n denotes the n-th order derivative with respect to t, i.e., $D^n = \frac{d^n}{dt^n}$ and $*$ denotes the convolution.

II. BACKGROUND

Formally, the objective of univariate smoothness priors and QV regularization strategies is to find an estimate of signal $x_{lp}(t)$, where the subscript lp stands for low-pass signal, from its noisy measurement $y(t)$ in the model

$$y(t) = x_{lp}(t) + v_{lp}(t), \qquad (1)$$

where $v_{lp}(t)$ is the unwanted additive noise signal, by solving the following optimization problem

$$\hat{x}_{lp}(t) = \underset{x_{lp}(t)}{\operatorname{argmin}} \int [y(\tau) - x_{lp}(\tau)]^2 \, d\tau + \alpha^2 \int \left[\frac{d^n}{dt^n} x_{lp}(\tau)\right]^2 d\tau, \qquad (2)$$

where $\frac{d^n}{dt^n}$ is the n-th order derivative and α denotes a smoothness tradeoff which balances the error variance and the output smoothness. The subscript lp is used as a low-pass in the above equations since the mentioned methods act as low-pass smoothing filter [36]. In order to digitally process the signal at discrete time k, one needs to consider $y[k] = y[kT_s]$, the discrete-time samples of $y(t)$, where T_s is the sampling period, $x_{lp}[k]$ the sampled desired signal and $v_{lp}[k]$ the sampled observation noise:

$$y[k] = x_{lp}[k] + v_{lp}[k], \quad k = 1, \cdots, L.$$

The smoothing procedure (2) can be then expressed as

$$\hat{x}_{lp}[k] = \underset{x_{lp}[k]}{\operatorname{argmin}} \sum_{j=1}^{L} (y[j] - x_{lp}[j])^2 + \alpha^2 \sum_{j=1}^{L} (\nabla^n x_{lp}[j])^2, \qquad (3)$$

for all $k = 1, \cdots, L$, where $\nabla^n x_{lp}[k]$ is the n-th order difference approximation of the derivative. There are several methods that can be used for the difference approximation of the derivative. Some of them are forward, backward, central difference rule and bilinear transform. Without loss of generality, we consider the backward/forward difference rule for the difference approximation of the derivative. In this case, the first order difference is defined by $\nabla x_{lp}[k] = x_{lp}[k] - x_{lp}[k \pm 1]$, where one with "+" holds for forward difference rule and one with "−" holds for backward difference rule. The n-th order difference is then computed as $\nabla^n x_{lp}[k] = \nabla(\nabla^{n-1}) x_{lp}[k]$. The n-th order difference, $\nabla^n x_{lp}[k]$, can also be computed as

$$\nabla^n x_{lp}[k] = \sum_{i=0}^{n} (-1)^i \binom{n}{i} x_{lp}[k \pm i], \qquad (4)$$

where $\binom{n}{i}$ is the binomial coefficient. In the following, we use the backward difference rule. It then means that, from now,

we used $x_{lp}[k - i]$ in (4). Hence, (3) can be written in the following matrix notation [36]

$$\hat{x}_{lp} = \|\boldsymbol{y} - \boldsymbol{x}_{lp}\|^2 + \alpha^2 \|\boldsymbol{D}_n \boldsymbol{x}_{lp}\|^2, \qquad (5)$$

where $\boldsymbol{y} = (y[1], \cdots, y[L])^T$, $\boldsymbol{x}_{lp} = (x_{lp}[1], \cdots, x_{lp}[L])^T$, $\|\cdot\|^2$ denotes the Euclidean norm and $\boldsymbol{D}_n$ is the $L \times L$ Toeplitz matrix, built from the backward difference operator $\boldsymbol{d}_n$:

$$\boldsymbol{d}_n \triangleq \left[1, \ -\binom{n}{1}, \cdots, \ (-1)^{n-1}\binom{n}{n-1}, \ (-1)^n\binom{n}{n}\right].$$

The difference operator $\boldsymbol{d}_n$ can also be computed by the following recursion:

$$\begin{cases} \boldsymbol{d}_1 \triangleq [+1, \ -1] & n = 1 \\ \boldsymbol{d}_n = \boldsymbol{d}_{n-1} * \boldsymbol{d}_1 & n > 1 \end{cases}, \qquad (6)$$

where $*$ denotes the convolution operator. Throughout the paper, the second-order smoothness priors is of special interest. As a special case, for data on L samples, D_2 is a $L \times L$ matrix defined as:

$$\boldsymbol{D}_2 = \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 & 0 \\ -2 & 1 & \cdots & 0 & 0 & \\ 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}_{L \times L}. \qquad (7)$$

The solution that follows from the minimization of (5) is

$$\hat{\boldsymbol{x}}_{lp} = (I + \alpha^2 \boldsymbol{D}_n^T \boldsymbol{D}_n)^{-1} \boldsymbol{y}.$$

Any component $\hat{x}_{lp}[k]$ of the vector $\boldsymbol{x}_{lp}$ can be written in the following convolution form [36]:

$$\hat{x}_{lp}[k] = \left(\delta[k] + \alpha^2 \boldsymbol{d}_n[-k] * \boldsymbol{d}_n[k]\right)^{\boxdot} * y[k], \qquad (8)$$

where $(\cdot)^{\boxdot}$ denotes the deconvolution. (8) can be implemented as

$$\hat{x}_{lp}[k] = Z^{-1}\left\{\frac{1}{1 + \alpha^2 d_n(\frac{1}{z})d_n(z)}\right\} * y[k],$$

where Z^{-1} denotes the inverse Z-transform. The above regularization method can be stated in terms of a zero-phase, linear time-invariant (LTI) smoothing filter (for more details see [17], [36]). It can be analyzed in the frequency domain. To this purpose by substituting $y[k]$ with $\delta[k]$ in (8), and taking the Z-transform, the frequency response of the smoothing filter is obtained as follows [36]

$$\begin{cases} H_n^{lp}(z) = \dfrac{1}{1 + \alpha^2 d_n(z)d_n(\frac{1}{z})} = \dfrac{1}{1 + \alpha^2(1 - z^{-1})^n(1 - z)^n} \\ H_n^{lp}(e^{j\omega}) = \dfrac{1}{1 + \alpha^2(2\sin\frac{\omega}{2})^{2n}} \end{cases}. \qquad (9)$$

The transfer function of the smoothing filter, i.e., (9), is real and its phase is zero. It has a low-pass characteristic which makes it suited for extracting the low frequency components of the signal, hence the subscript lp in the previous equations. The regularization factor α determines at which frequency the
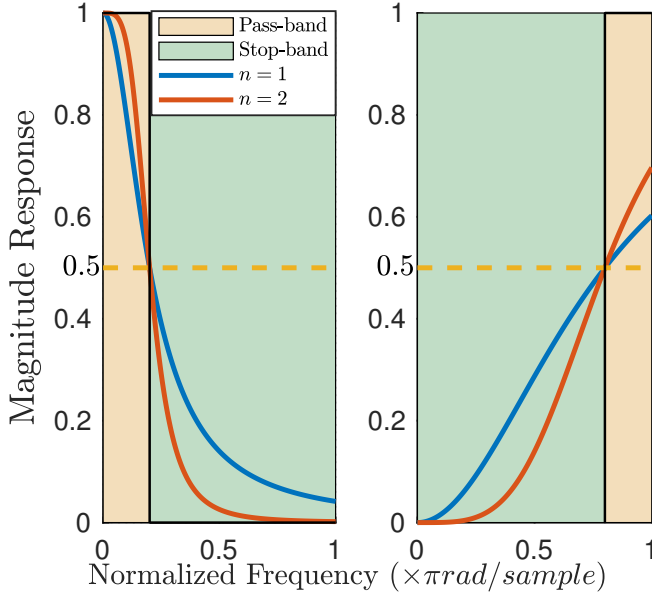
Fig. 1. The amplitude response of LPSF with cutoff frequency $\omega_1 = 0.2$ (left panel) and HPSF with cutoff frequency $\omega_2 = 0.8$ (right panel) for different values of n.

transition of pass-band to stop-band occurs. By equating the Fourier frequency response, $H_n^{lp}(e^{j\omega})$, to $1/2$, the value of α is obtained as

$$\alpha_o = \frac{1}{(2 \sin \frac{\omega_{cut}}{2})^n},$$

where ω_{cut} is the cutoff frequency of the smoothing filter. We refer to this approach as low-pass smoothing filter (LPSF). Note that for designing a filter, the most important parameters are the order and the cutoff frequency. So, with the optimization approach, we can directly (and easily) control the order with the number of derivatives, and the cutoff frequency is related to the hyperparameter α. The method can also be used for extracting the high frequency components: by extracting the low frequency components and subtract it from the observed signal:

$$\hat{x}_{hp}[k] = y[k] - \hat{x}_{lp}[k]. \qquad (10)$$

In this case, the transfer function of the high-pass smoothing filter is $H_n^{hp}(z) = 1 - H_n^{lp}(z)$, i.e.,

$$\begin{cases} H_n^{hp}(z) = \dfrac{\left[(1 - z^{-1})(1 - z)\right]^n}{(2 \sin \frac{\omega_{cut}}{2})^{2n} + \left[(1 - z^{-1})(1 - z)\right]^n} \\[4mm] H_n^{hp}(e^{j\omega}) = \dfrac{(2 \sin \frac{\omega}{2})^{2n}}{(2 \sin \frac{\omega_{cut}}{2})^{2n} + (2 \sin \frac{\omega}{2})^{2n}} \end{cases},$$

where the superscript hp refers to high-pass smoothing filter (HPSF). The frequency responses of an LPSF with cutoff frequency $\omega_1 = 0.2$ and an HPSF with cutoff frequency $\omega_2 = 0.8$ for the first and second order are depicted in Figure 1. The transition band of LPSF and HPSF increases as the order of the smoothing filter (i.e., the order of the derivative) increases. In Section III, we present a simple approach to
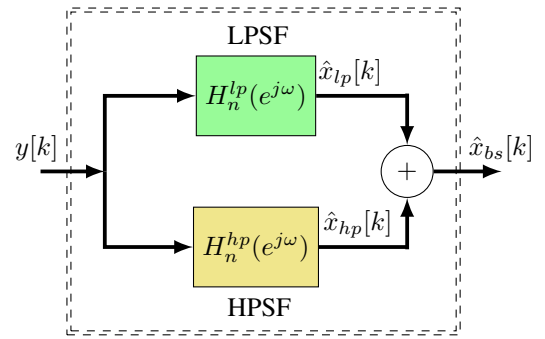

Fig. 2. The simple parallel structure to band-stop smoothing filter (BSSF) design.

BSSF design which is based on the combination of an LPSF in parallel with an HPSF. But, we show that the simple BSSF is less than ideal. Next, in Section IV, we show a new approach to modify that smoothing filter, making its performance nearly ideal.

III. A SIMPLE DESIGN OF BAND-STOP SMOOTHING FILTER

Let us consider the problem of signal denoising where the desired signal $x_{bs}[k]$ (subscript bs stands for band-stop) contains a low-frequency and a high-frequency component. We assume that the noisy signal $y[k]$ can be represented as

$$y[k] = \underbrace{x_{lp}[k] + x_{hp}[k]}_{x_{bs}[k]} + v_{bs}[k], \qquad (11)$$

where $v_{bs}[k]$ is the unwanted additive noise. We assume that the noise $v_{bs}[k]$ and the signal $x_{bs}[k]$ are on disjoint frequency ranges. A simple approach to recover $x_{bs}[k]$ is to extract low and high frequency components (i.e., $x_{lp}[k]$ and $x_{hp}[k]$) separately and combine them to extract the desired signal (i.e., $x_{bs}[k]$). Therefore, a simple approach to BSSF design is to combine an LPSF in parallel with an HPSF [36]. The block diagram of a simple BSSF is given in Figure 2. Suppose that a BSSF with cutoff frequencies ω_1 and ω_2 ($\omega_1 < \omega_2$) is desired. By extracting the low and high frequency components using an LPSF with cutoff frequency ω_1 and an HPSF with cutoff frequency ω_2 and combining them, one can extract a desired signal with frequency components in range of $\omega \in [0, \ \omega_1] \cup [\omega_2, \ 1]$. Note that LPSF and HPSF have the following transfer functions

$$\begin{cases} H_n^{lp}(e^{j\omega}) = \dfrac{1}{1 + \left(\frac{\sin \frac{\omega}{2}}{\sin \frac{\omega_1}{2}}\right)^{2n}} \\[4mm] H_n^{hp}(e^{j\omega}) = \dfrac{1}{1 + \left(\frac{\sin \frac{\omega_2}{2}}{\sin \frac{\omega}{2}}\right)^{2n}} \end{cases}. \qquad (12)$$

Therefore, the simplest BSSF that we can create has the frequency response function:

$$H_n^{bs}(e^{j\omega}) = \frac{1}{1 + \left(\frac{\sin \frac{\omega}{2}}{\sin \frac{\omega_1}{2}}\right)^{2n}} + \frac{1}{1 + \left(\frac{\sin \frac{\omega_2}{2}}{\sin \frac{\omega}{2}}\right)^{2n}}.$$

The BSSF has a center frequency at $\omega_c = \sqrt{\omega_1 \omega_2}$ and its bandwidth is $\omega_2 - \omega_1$. The quality factor (Q-factor) of the
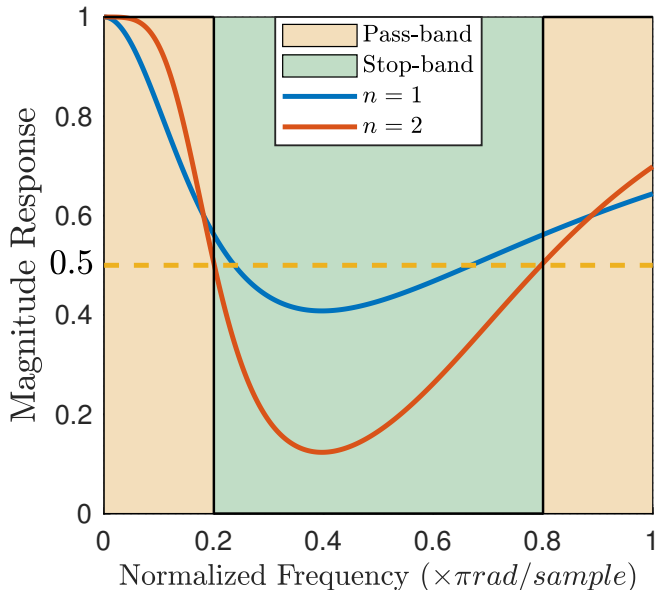
Fig. 3. The amplitude response of a simple BSSF with cutoff frequencies $\omega_1 = 0.2$ and $\omega_2 = 0.8$ for different values of n.

Fig. 4. The amplitude response of a simple NSF with cutoff frequencies $\omega_1 = 0.47$ and $\omega_2 = 0.53$ for different values of n.

smoothing filter that shows how fast the roll-off is, is defined by

$$Q = \frac{\omega_c}{|\omega_2 - \omega_1|}.$$

If $Q < 1$, then the BSSF is a wide-band BSSF, and if $Q > 1$, it is a narrow-band BSSF. The BSSF attenuates the signals whose frequencies are in the range $\omega \in (\omega_1, \omega_2)$ and allows other frequencies to pass. The magnitude responses of a simple BSSF, $H_n^{bs}(e^{j\omega})$ with cutoff frequencies $\omega_1 = 0.2$ and $\omega_2 = 0.8$ are shown in Figure 3. We find that the first order BSSF shown in Figure 3 does not have cutoff frequencies at ω_1 and ω_2. Another drawback is that the simple BSSF does not have zero gain at center frequency ($\omega_c = 0.4$) and its vicinity. Moreover, its transition band is not narrow. The problem even becomes worse when we need to design a BSSF with high Q-factor. These types of BSSF are known as notch smoothing filters (NSFs). For example, consider a simple NSF with center frequency $\omega_c = 0.5$, $\omega_1 = 0.47$ and $\omega_2 = 0.53$. The magnitude response plot of this NSF is shown in Figure 4. There, we find the worst case: its cutoff frequencies are not exactly the planned-for the desired cutoff frequencies, the amplitude at the center frequency is not zero and its transition band is not enough narrow to eliminate the unwanted frequency components and pass desired frequency components. Therefore, extracting low and high frequency components separately and combining them to construct BSSF is simple but naive and weak. We propose to establish the conditions for achieving optimal performance, in extracting the low and high frequency parts simultaneously and considering the interaction between LPSF and HPSF. In the following section, we propose an improved optimization-based approach to BSSF design.
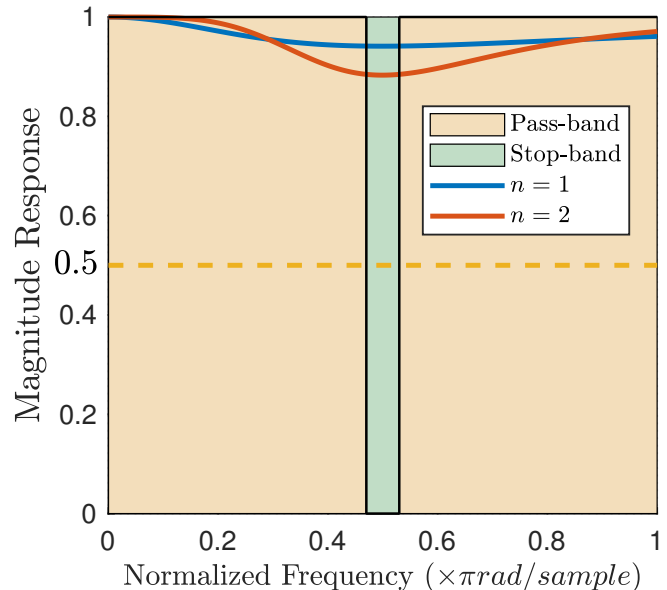
IV. IMPROVED DESIGN OF BAND-STOP SMOOTHING FILTER

In order to construct a BSSF, in the previous section, we naively combined an LPSF with an HPSF. However, the HPSF was designed by subtracting the output of LPSF from the observed signal (see (10)). As a result, its formulation depends on the LPSF. Therefore, it is not suitable for studying the interaction between the low-pass and high-pass smoothing filters. In the following, we propose a new approach to implement the regularization method (i.e., smoothness priors or QV regularization) which makes it suited for HPSF design.

To this end, by replacing the n-th order derivative using d_n following (4) and (6), we represent (3) in the following regularized form:

$$\hat{x}_{lp}[k] = \underset{x_{lp}[k]}{\mathrm{argmin}} \sum_{j=1}^{L} (y[j] - x_{lp}[j])^2 + \alpha^2 \sum_{j=1}^{L} (d_n[j] * x_{lp}[j])^2, \tag{13}$$

where

$$\begin{cases} d_1[k] = \delta[k] - \delta[k+1] \\ d_n[k] = d_{n-1}[k] * d_1[k] \end{cases}$$

The actual minimization problem (13) is related to the state model

$$\begin{cases} d_n[k] * x_{lp}[k] = e_{lp}[k] \\ y[k] = x_{lp}[k] + v_{lp}[k] \end{cases}, \tag{14}$$

According to (13), it can be realized that the traditional smoothness priors or QV regularization is an adaptive estimation algorithm which assumes that the state of the system at a time k evolved from the prior state at previous times according to the equation (14). The stabilizing kernel used in the QV regularization $d_n[k]$ is chosen to be some differential operator. It imposes smoothness constraints on the solution.

Therefore, the obtained filter acts as a low-pass filter. In the following, we suggest a new state space model for the signal which the stabilizing kernel $u[k]$, the unit step function, is chosen to be some integral operation. Using the new state space model, we propose a modified approach to smoothness priors which is suited for HPSF design.

A. A new approach to HPSF design

Let us consider the following state space model for the desired signal:

$$\begin{cases} b_n[k] * x_{hp}[k] = e_{hp}[k] \\ y[k] = x_{hp}[k] + v_{hp}[k] \end{cases}, \quad (15)$$

where

$$\begin{cases} b_1[k] = u[k] \\ b_n[k] = b_{n-1}[k] * b_1[k] \end{cases}.$$

Choosing $u[k]$ as the stabilizing kernel is equivalent to imposing the irregularity constraints on the solution. In the following, we will see that when the stabilizing kernel $u[k]$ is used, the obtained filter has a high-pass characteristic. According to the state space model (15), we propose the following regularized approach:

$$\hat{x}_{hp}[k] = \underset{x_{hp}[k]}{\mathrm{argmin}} \sum_{j=1}^{L} (y[j] - x_{hp}[j])^2 + \beta^2 \sum_{j=1}^{L} (b_n[j] * x_{hp}[j])^2, \quad (16)$$

where β is the regularization factor. The regularization term in (16) is expressed with power of the integral of the signal[2], against the regularization term in (3) which was explicitly the smoothness expressed with power of signal derivatives. (16) can be expressed in the following matrix notation

$$\hat{\boldsymbol{x}}_{hp} = \underset{\boldsymbol{x}_{hp}}{\mathrm{argmin}} \|\boldsymbol{y} - \boldsymbol{x}_{hp}\|^2 + \beta^2 \|\boldsymbol{B}_n \boldsymbol{x}_{hp}\|^2, \quad (17)$$

where $\boldsymbol{B}_n$ is defined as a lower triangular matrix and $b_n[k]$ is truncated to construct the matrix filter $\boldsymbol{B}_n$. For instance, $\boldsymbol{B}_1$ is defined by

$$\boldsymbol{B}_1 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 1 & \dots & 1 & 1 \end{pmatrix},$$

which represents a cumulative sum. $\boldsymbol{B}_2$ is defined by

$$\boldsymbol{B}_2 = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ n & \dots & 2 & 1 \end{pmatrix}. \quad (18)$$

The solution that follows from the minimization of (17) is

$$\hat{\boldsymbol{x}}_{hp} = (I + \beta^2 \boldsymbol{B}_n^T \boldsymbol{B}_n)^{-1} \boldsymbol{y}.$$

---

[2]In continuous time, we have $u(t) * x(t) = \int_{-\infty}^{t} x(\tau)d\tau$ and in discrete time, we have $u[k] * x[k] = \sum_{j=-\infty}^{k} x[j]$.

The parameter β can be computed in terms of cutoff frequency. To this purpose, we write the component $\hat{x}_{hp}[k]$ of the vector $\boldsymbol{x}_{hp}$ in the following convolution form

$$\hat{x}_{hp}[k] = \left(\delta[k] + \beta^2 \boldsymbol{b}_n[k] * \boldsymbol{b}_n[-k]\right)^{\square} * y[k]. \quad (19)$$

In the Z-domain, the amplitude response of (19) is

$$H_n^{hp}(z) = \frac{1}{1 + \beta^2 b_n(z) b_n(\frac{1}{z})} = \frac{1}{1 + \dfrac{\beta^2}{[(1 - z^{-1})(1 - z)]^n}}. \quad (20)$$

Eq. (20) becomes in the Fourier domain

$$H_n^{hp}(e^{j\omega}) = \frac{1}{1 + \dfrac{\beta^2}{\left(2 \sin \dfrac{\omega}{2}\right)^{2n}}}. \quad (21)$$

By equating (21) to $1/2$, the value of β is obtained in terms of cutoff frequency:

$$\beta_o = (2 \sin \frac{\omega_c}{2})^{2n}.$$

Finally, the amplitude response of (20) is expressed as

$$\begin{cases} H_n^{hp}(z) = \dfrac{[(1 - z^{-1})(1 - z)]^n}{(2 \sin \frac{\omega_c}{2})^{2n} + [(1 - z^{-1})(1 - z)]^n} \\ H_n^{hp}(e^{j\omega}) = \dfrac{(2 \sin \frac{\omega}{2})^{2n}}{(2 \sin \frac{\omega_c}{2})^{2n} + (2 \sin \frac{\omega}{2})^{2n}} \end{cases}. \quad (22)$$

Comparing (22) and (12), we conclude that the frequency response of the new regularization model is equivalent to the HPSF described in the previous section. It acts as an HPSF. The advantage of the new HPSF is that its model does not depend on the LPSF model. Therefore, it can be used for studying the interaction between the LPSF and HPSF when we combine them to construct a BSSF. In the following section, the new HPSF is combined with the LPSF for designing the BSSF.

B. A new model to band-stop smoothing filter design

In order to overcome the drawbacks of the simple parallel approach discussed in Section III, in the following, we consider the interactions between the LPSF and HPSF. The proposed method is as follows:

We are interested in the estimation of the signal $x_{bs}[k]$ from its noisy measurements $y[k]$ in the model (11). Figure 5 depicts our proposed scheme for BSSF, and the following equations give its mathematical representation

$$\begin{cases} d_n[k] * x_{lp}[k] = e_{lp}[k] \\ b_n[k] * x_{hp}[k] = e_{hp}[k] \\ y[k] = \underbrace{x_{lp}[k] + x_{hp}[k]}_{x_{bs}[k]} + v_{bs}[k] \end{cases}$$

This model assumes that the lp and hp part of the signal of interest are modelled by filtering a white noise. We assume
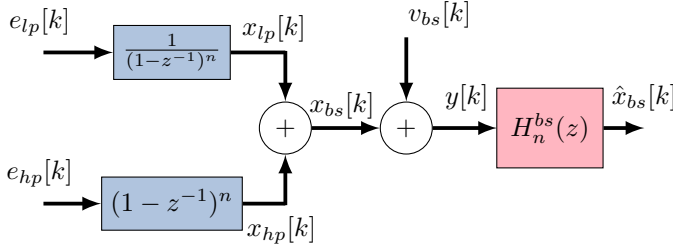
Fig. 5. The proposed scheme for band-stop smoothing filter design.

that the inputs $e_{lp}[k]$ and $e_{hp}[k]$ are each zero-mean white noise, with

$$
\mathbb{E}\left\{\begin{bmatrix} e_{lp}[k] \\ e_{hp}[k] \end{bmatrix} \begin{bmatrix} e_{lp}^*[m] & e_{hp}^*[m] \end{bmatrix}\right\}
$$
$$
= \begin{bmatrix} \sigma_{lp}^2 & \rho\sigma_{lp}\sigma_{hp} \\ \rho\sigma_{lp}\sigma_{hp} & \sigma_{hp}^2 \end{bmatrix} \delta_{k,m},
$$

(23)

where $\mathbb{E}\{\}$ refers to the expected value operation, σ_{lp}^2 and σ_{hp}^2 denote the noises variance and ρ is the noise correlation.

Considering the fact that the system noises can be correlated, we propose to estimate the desired signal $\hat{x}_{bs}[k]$ by solving the following least-squares optimization problem

$$
\hat{x}_{lp}[k], \hat{x}_{hp}[k] = \underset{x_{lp}[j], x_{hp}[j]}{\mathrm{argmin}} \sum_{j=1}^{L} (y[j] - x_{lp}[j] - x_{hp}[j])^2
$$
$$
+ \alpha^2 \sum_{j=1}^{L} (d_n[j] \star x_{lp}[j])^2
$$
$$
+ \beta^2 \sum_{j=1}^{L} (b_n[j] \star x_{hp}[j])^2
$$
$$
+ 2\rho\alpha\beta \sum_{j=1}^{L} (d_n[j] \star x_{lp}[j]) (b_n[j] \star x_{hp}[j]).
$$

(24)

The first regularization term is used to control the smoothness of x_{lp}, the second regularization term controls the sharpness of x_{hp} and the third regularization term is used to control the interaction between the smoothness of x_{lp} and the sharpness of x_{hp}.

Equation (24) can be expressed in the following matrix notation

$$
\hat{x}_{lp}, \hat{x}_{hp} = \underset{x_{lp}, x_{hp}}{\mathrm{argmin}} \|y - x_{lp} - x_{hp}\|^2 + \alpha^2 \|D_n x_{lp}\|^2
$$
$$
+ \beta^2 \|B_n x_{hp}\|^2 + 2\rho\alpha\beta (D_n x_{lp})^T (B_n x_{hp}),
$$

(25)

where $(\cdot)^T$ denotes the matrix transpose operator. Note that $D_n^T B_n = B_n^T D_n = I$. Hence, (25) can be expressed as

$$
\hat{x}_{lp}, \hat{x}_{hp} = \underset{x_{lp}, x_{hp}}{\mathrm{argmin}} \|y - x_{lp} - x_{hp}\|^2 + \alpha^2 \|D_n x_{lp}\|^2
$$
$$
+ \beta^2 \|B_n x_{hp}\|^2 + 2\rho\alpha\beta x_{lp}^T x_{hp}.
$$

(26)

By setting the derivatives of (26) with respect to x_{lp} and x_{hp} equal to zero, after some simplifications, we get the following system of equations for x_{lp} and x_{hp}:

$$
\begin{cases} M_{11} x_{lp} + M_{12} x_{hp} = y, \\ M_{21} x_{lp} + M_{22} x_{hp} = y \end{cases},
$$

(27)

where

$$
\begin{cases} M_{11} = I + \alpha^2 D_n^T D_n \\ M_{12} = M_{21} = (1 + \rho\alpha\beta) I \\ M_{22} = I + \beta^2 B_n^T B_n \end{cases}
$$

Solving the set of equations (27), we find

$$
\begin{cases} \hat{x}_{lp} = [M_{11} M_{22} - M_{12} M_{21}]^{-1} [M_{22} - M_{12}] y \\ \hat{x}_{hp} = [M_{11} M_{22} - M_{12} M_{21}]^{-1} [M_{11} - M_{21}] y \end{cases}
$$

Finally, the desired signal, $\hat{x}_{bs}$, is obtained as

$$
\hat{x}_{bs} = \hat{x}_{lp} + \hat{x}_{hp}
$$
$$
= [M_{11} M_{22} - M_{12} M_{21}]^{-1} [M_{11} + M_{22}
$$
$$
- (M_{12} + M_{21})] y.
$$

(28)

The only unknown parameters are the regularization factors α and β. In the following section, the design parameters are calculated in terms of the BSSF cutoff frequencies.

C. Frequency domain analysis: parameter design based on cutoff frequencies

The BSSF developed in (28), was derived from a regularized least-squares optimization in form of block-wise matrix formulation. It can also be stated in terms of a zero-phase, forward filtering backward smoothing scheme. Therefore, the regularization factors α and β can be calculated in terms of the BSSF cutoff frequencies. Any component $\hat{x}_{lp}[k]$, $\hat{x}_{hp}[k]$ and $\hat{x}_{bs}[k]$ of the vector x_{lp}, x_{hp} and x_{bs} can be written in the convolution form represented in (29), on the top of the next page, where

$$
\begin{cases} m_{11}[k] = \delta[k] + \alpha^2 d_n[-k] * d_n[k] \\ m_{12}[k] = m_{21}[k] = (1 + \rho\alpha\beta)\delta[k] \\ m_{22}[k] = \delta[k] + \beta^2 b_n[-k] * b_n[k] \end{cases}
$$

The frequency response of (29) is

$$
H_n^{bs}(z) = \frac{M_{11}(z) + M_{22}(z) - [M_{12}(z) + M_{21}(z)]}{M_{11}(z) M_{22}(z) - M_{12}(z) M_{21}(z)},
$$

(30)

where $M_{ij}(z)$ is the Z-transform of $m_{ij}[k]$:

$$
\begin{cases} M_{11}(z) = 1 + \alpha^2 d_n(\frac{1}{z}) d_n(z) \\ M_{12}(z) = M_{21}(z) = 1 + \rho\alpha\beta \\ M_{22}(z) = 1 + \beta^2 b_n(\frac{1}{z}) b_n(z) \end{cases}
$$

Equation (30) in Fourier domain becomes

$$
H_n^{bs}(e^{j\omega}) = \frac{\kappa}{(\alpha\beta)^2(1 - \rho^2) + \kappa},
$$

(31)

$$\begin{cases} \hat{x}_{lp}[k] = (m_{11}[k] * m_{22}[k] - m_{12}[k] * m_{21}[k])^{\square} * (m_{22}[k] - m_{21}[k]) * y[k] \\ \hat{x}_{hp}[k] = (m_{11}[k] * m_{22}[k] - m_{12}[k] * m_{21}[k])^{\square} * (m_{11}[k] - m_{12}[k]) * y[k] \\ \hat{x}_{bs}[k] = (m_{11}[k] * m_{22}[k] - m_{12}[k] * m_{21}[k])^{\square} * (m_{11}[k] + m_{22}[k] - m_{21}[k] - m_{12}[k]) * y[k] \end{cases} \quad (29)$$

where

$$\kappa = \alpha^2 (2 \sin \frac{\omega}{2})^{2n} - 2\rho\alpha\beta + \frac{\beta^2}{(2 \sin \frac{\omega}{2})^{2n}}.$$

The magnitude response of the BSSF, (31), has two cutoff frequencies at ω_1 and ω_2 which are satisfying $|H_n^{bs}(e^{j\omega})| = \frac{1}{2}$. Substituting each of these two points into equation (31) results a system of equations which contains two equations that share two unknowns α and β:

$$\alpha^2 (2 \sin \frac{\omega_1}{2})^{2n} - 2\rho\alpha\beta + \frac{\beta^2}{(2 \sin \frac{\omega_1}{2})^{2n}} = \alpha^2\beta^2(1 - \rho^2)$$

$$\alpha^2 (2 \sin \frac{\omega_2}{2})^{2n} - 2\rho\alpha\beta + \frac{\beta^2}{(2 \sin \frac{\omega_2}{2})^{2n}} = \alpha^2\beta^2(1 - \rho^2)$$

$$(32)$$

Subtracting the second equation from the first one, we find

$$\beta = \alpha \left[(2 \sin \frac{\omega_1}{2})(2 \sin \frac{\omega_2}{2}) \right]^n \quad (33)$$

Substituting (33) in the first equation of (32), the resulting equation is

$$\alpha^2 2^{2n} [(\sin \frac{\omega_1}{2})^{2n} - 2\rho(\sin \frac{\omega_1}{2} \sin \frac{\omega_2}{2})^n + (\sin \frac{\omega_2}{2})^{2n}] \quad (34)$$
$$= \alpha^2\beta^2(1 - \rho^2)$$

After some simplifications, we find

$$\beta_o = 2^n \sqrt{\frac{(\sin \frac{\omega_1}{2})^{2n} - 2\rho(\sin \frac{\omega_1}{2} \sin \frac{\omega_2}{2})^n + (\sin \frac{\omega_2}{2})^{2n}}{1 - \rho^2}}$$

$$(35)$$

From (33) and (35), the value of α is obtained as

$$\alpha_o = \frac{\beta_o}{\left[(2 \sin \frac{\omega_1}{2})(2 \sin \frac{\omega_2}{2}) \right]^n} \quad (36)$$

It means that by setting the value of α and β respectively to (36) and (35), the proposed optimization approach, i.e., (28), acts as a BSSF with cutoff frequencies ω_1 and ω_2. As an example, the amplitude response of the new BSSF for $\omega_1 = 0.2$ and $\omega_2 = 0.8$, $n = 1, 2$ and different values of $\rho \in (-1, 1)$ is shown in Figure 6. It shows that the transition band of the BSSF increases as ρ becomes close to 1. We find that at $\rho = 1 - \epsilon \approx 1$ (ϵ is a small value), the BSSF has narrower transition band and its gain at $\omega = \omega_c$ is zero. Therefore, by setting the value of ρ close to 1, a near-ideal BSSF is obtained. In Figure 6, we set ϵ to 10^{-3}. The amplitude response of the BSSF, for $\rho = 0.999$, $\omega_1 = 0.2$, $\omega_2 = 0.8$ and $n = 1, 2$ is shown in Figure 7. Repeating our last example, for second order NSF with $\omega_c = 0.5$, $\omega_1 = 0.47$, $\omega_2 = 0.53$ and different values of ρ, we find that the frequency response of the improved NSF, shown in Figure

8 is nearly ideal when the value of ρ is sufficiently close to 1. The results of Figure 8 also show that the transition band of the NSF is related to ρ. In [40], a general approach for ARMA with exogenous inputs (ARMAX) signals smoothing with application to variable-Q ARMA smoothing filter design was proposed. Specially, a narrowband notch smoothing filter with variable-Q was designed. The NSF proposed in [40] employs an ARMA model. In Figure 9, we plot the amplitude response of the NSF proposed in [40] and the first order NSF in the present paper with cutoff frequencies $\omega_1 = 0.07$ and $\omega_2 = 0.13$. We can see from the amplitude responses that the proposed NSF acts better in the pass band. The NSF proposed in [40] does not have unit gain in the pass band $\omega \in [0, \omega_1]$ while the proposed NSF has approximately unit gain in both pass bands, $\omega \in [0, \omega_1]$ and $\omega \in [\omega_2, \pi]$. We also plot the amplitude response of the first order Butterworth filter and Kaiser with order 40. Kaiser method has better amplitude response in its pass-band and stop-band, but compared to the proposed BSSF, its order is dramatically higher. It is notable that the design of notch filters is a classic problem in recursive (IIR) filter design. Especially, a second-order transfer system:

$$H_n(z) = \frac{1 - 2\cos(\omega_0)z^{-1} + z^{-2}}{1 - 2\rho\cos(\omega_0)z^{-1} + \rho^2 z^{-2}} \quad (37)$$

with a pair of zeros on the unit circle and a pair of poles place inside the unit circle makes a simple and classical notch filter [41], [43]. However, the classical notch filter has several drawbacks that has been discussed in [40]. For avoiding duplication of results presented in [40], we do not give theoretical derivations or algorithm descriptions that are easily found in [40]. It has been shown that the approach presented in [40] is more effective than the classical notch filter described in [41], [43]. That is why we do not compare our proposed filter with filters designed using more classical methods.

Finally, a BPSF with cutoff frequencies ω_1 and ω_2 can be obtained by subtracting the BSSF output (28) from the observed signal (i.e., $\hat{x}_{bp} = y - \hat{x}_{bs}$).

D. A Causal filter

The designed band-stop smoothing filter is a non-causal filter as it takes all the measurements into account [see (24) and (25)]. So it can only be used for off-line applications. In order to implement it for online applications, in the following, we propose a causal filter.

From (28), it can be seen that the matrix $\Upsilon = [M_{11}M_{22} - M_{12}M_{21}]^{-1} [M_{11} + M_{22} - (M_{12} + M_{21})]$ is a symmetric positive-definite matrix. Therefore, it has a unique Cholesky decomposition of the form $\Upsilon = L^T L$, where L is a lower triangular matrix with real and positive diagonal entries.
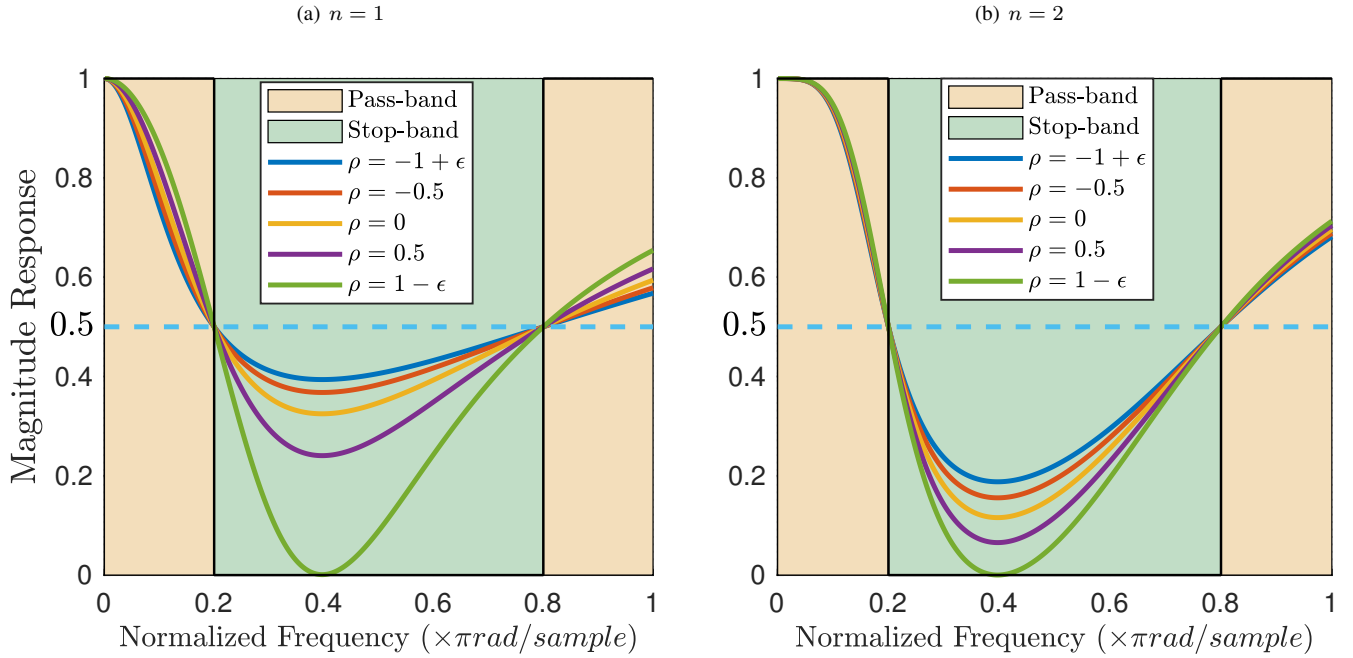
(a) $n = 1$ (b) $n = 2$

Fig. 6. The amplitude response of the proposed BSSF with cutoff frequencies $\omega_1 = 0.2$ and $\omega_2 = 0.8$, for different values of ρ a) the first order b) the second order. In this experiment, we set ϵ to 10^{-3}.
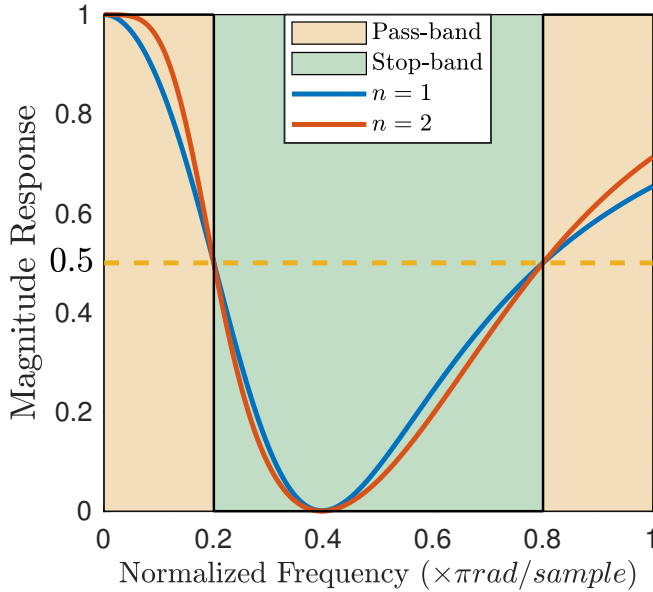

Fig. 7. The amplitude response of the proposed BSSF with cutoff frequencies $\omega_1 = 0.2$ and $\omega_2 = 0.8$, for $\rho = 0.999$ and $n = 1, 2$.
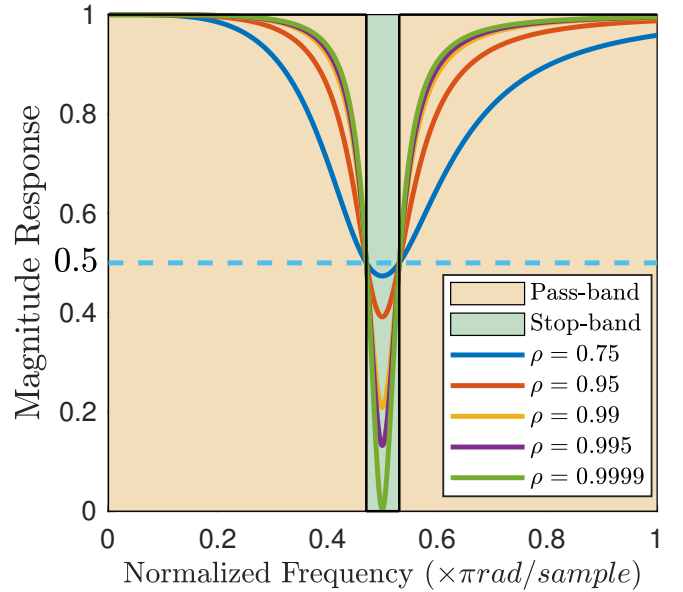
Fig. 8. The amplitude response of the second order NSF with cutoff frequencies $\omega_1 = 0.47$ and $\omega_2 = 0.53$, for different values of ρ. The transition band of the proposed NSF increases as the value of ρ increases.

Letting J denote the exchange matrix (square matrix with 1's on the anti-diagonal), L^T is equal to JLJ. Therefore, (28) can be expressed as

$$\hat{\boldsymbol{x}}_{bs} = \text{JLJL}\boldsymbol{y},$$

As a result, we can implement it using forward-backward filtering scheme with an FIR filter ($\boldsymbol{x}_{bs}^f = \text{L}\boldsymbol{y}$, where the superscript f stands for forward filter). For more details, see [44].

V. APPLICATION TO ECG DENOISING

In order to evaluate the performance of the proposed method, we employed it for ECG signal denoising. To this purpose, we applied the approach on ECG data from PhysioNet. The dataset contains 80 records, originally provided for the PhysioNet/Computers in Cardiology Challenge 2004 [45]. Each record, extracted from a two-lead, sampling frequency 128 Hz Holter ECG recording, 1 minute in length. In many
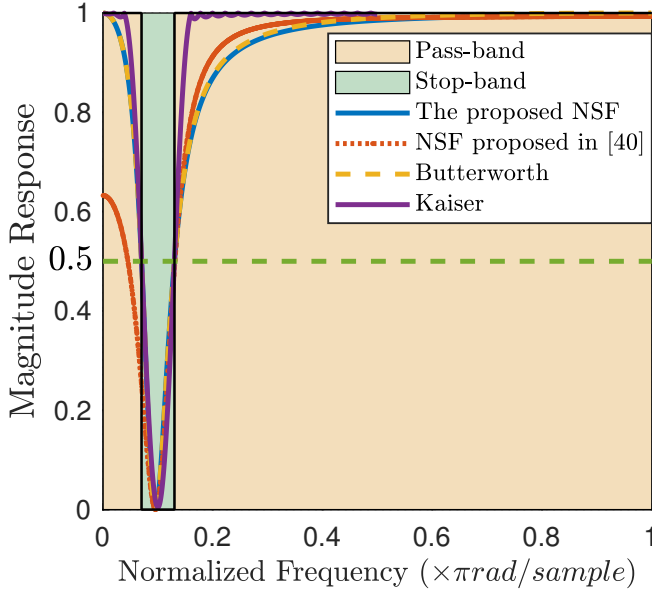
Fig. 9. The amplitude response of the NSF proposed in [40] and the first order NSF proposed in the present paper with cutoff frequencies $\omega_1 = 0.07$ and $\omega_2 = 0.13$. The proposed NSF has approximately unit gain in either pass band. We also plot the amplitude response of the first order Butterworth filter and Kaiser with order 40.

applications of signal processing such as ECG denoising, linear time invariant (LTI) filters (e.g., Butterworth filter) are still the standard choice for ECG system front-ends [46]. Therefore, we also employed a band-stop Butterworth filter for extracting the ECG signals and compared it with our proposed method. In the following case studies, the BSSF order is set to $n = 2$, and the operators $\boldsymbol{D}_2$ and $\boldsymbol{B}_2$ are as defined in (7) and (18), respectively. The Butterworth filters were applied to the data using the *filtfilt* function in Matlab to have zero-phase lag.

As a preliminary example, we plot a specific case (record b01m from PhysioNet/Computers in Cardiology Challenge 2004 [45]) at the top of Figure 10(a). At the bottom of that figure, we show its spectral representation.

At the top of Figure 10(b), we plot the same ECG, but corrupted with a synthetic band-limited noise ($SNR = -15$ dB), generated by the following equation:

$$v_{bs}(t) = \sum_{i=0}^{N-1} c_i \cos(2\pi f_i t + \psi_i) \qquad (38)$$

where c_i, f_i and ψ_i are respectively, the amplitude, frequency and phase of the i-th sinusoid. In this example, the frequencies of the noise were selected in range of $f_i \in (48, \ 52)$ Hz in (38). Therefore, a BSSF with cutoff frequencies 48 and 52 Hz could be used to remove the noise and extract the original ECG. The results of applying the second order simple BSSF (the naive BSSF presented in Section III), the second order zero-phase Butterworth filter and the proposed BSSF ($n = 2$) on the noisy ECG as well as their spectral representation are reported in Figures 10(c)-10(e). The denoised ECG using the proposed BSSF is visually closer to the true ECG, compared

to the denoised ECG by simple BSSF and zero-phase Butterworth filter. The spectral representation of the signals are also reported which confirms the superiority of the proposed BSSF visually (see spectrograms in Figure 10). For more details, we also report the values of NSR defined in (39) obtained by each method. The NSR obtained by our proposed method is smaller than other two methods which indicates the superiority of our proposed BSSF.

For evaluating quantitatively the performance of the different BSSFs, we tested them with variable SNR in the range -20 dB to 20 dB, by controlling the power of additive noise $\boldsymbol{v}_{bs}$. The frequency range of the additive noise were selected such that $f_i \in [46 \pm 4, \ 54 \pm 4]$ Hz in (38).

The average and standard deviation of the SNR improvement obtained after filtering, is denoted $SNR_{diff} = SNR_{out} - SNR_{in}$ and is used as a measure of performance. It is shown in Figure 11(a). The results, reported in Figure 11(a), show that the proposed BSSF outperforms the simple BSSF and zero-phase Butterworth filter. In Figure 11, we also report the result of the third order zero-phase Butterworth filter. In this case, also our proposed BSSF (with $n = 2$) outperforms the third order Butterworth filter. Other metrics that was used for evaluating the performance of the proposed method, are a classical ratio between the power of the reconstruction error and the power of the original signal (a noise-to-signal ratio denoted NSR), given by [36], [47]

$$\text{NSR} = \sqrt{\frac{\sum_k \left(x[k] - \hat{x}[k]\right)^2}{\sum_k x^2[k]}}, \qquad (39)$$

and the ratio between the power of the reconstruction error and the power of noise [36], [47]

$$\text{imp} = -10 \log_{10} \frac{\sum_k \left(\hat{x}_k - x_k\right)^2}{\sum_k \left(y_k - x_k\right)^2} \ \text{(dB)}. \qquad (40)$$

The results of the reconstruction procedures using these metrics are reported in Figures 11(b) and 11(c). These metrics also show that the proposed BSSF outperformed simple BSSF and zero-phase Butterworth filter. Figure 11(b) shows that the error of reconstruction for all methods decreases as the SNR increases (i.e., the reconstruction improvement continues for all methods). The NSR obtained by the proposed BSSF is smaller than the NSR obtained by Butterworth BSSF and simple BSSF which indicates the superiority of the proposed method. Furthermore, Figure 11(c) shows that the SNR improvement obtained by the proposed BSSF is larger than other two methods for all SNR, which means the proposed method outperforms Butterworth and simple BSSF in signal reconstruction.

VI. DISCUSSION AND CONCLUSION

A. Discussion

In this paper, we used the backward difference rule for implementing the derivative and integral due to its simplicity. Actually, the implementation of the unit sequence $u[k]$ is more simple and straightforward than trapezoidal integrator and other methods. In fact, we have already employed the Tustin's or trapezoidal method for improving the smoothness

(a) ECG record.

(b) Noisy ECG with SNR = -15

(c) Naive BSSF of Section III, $n = 2$

(d) Zero-phase Band-stop Butterworth filter, $n = 2$

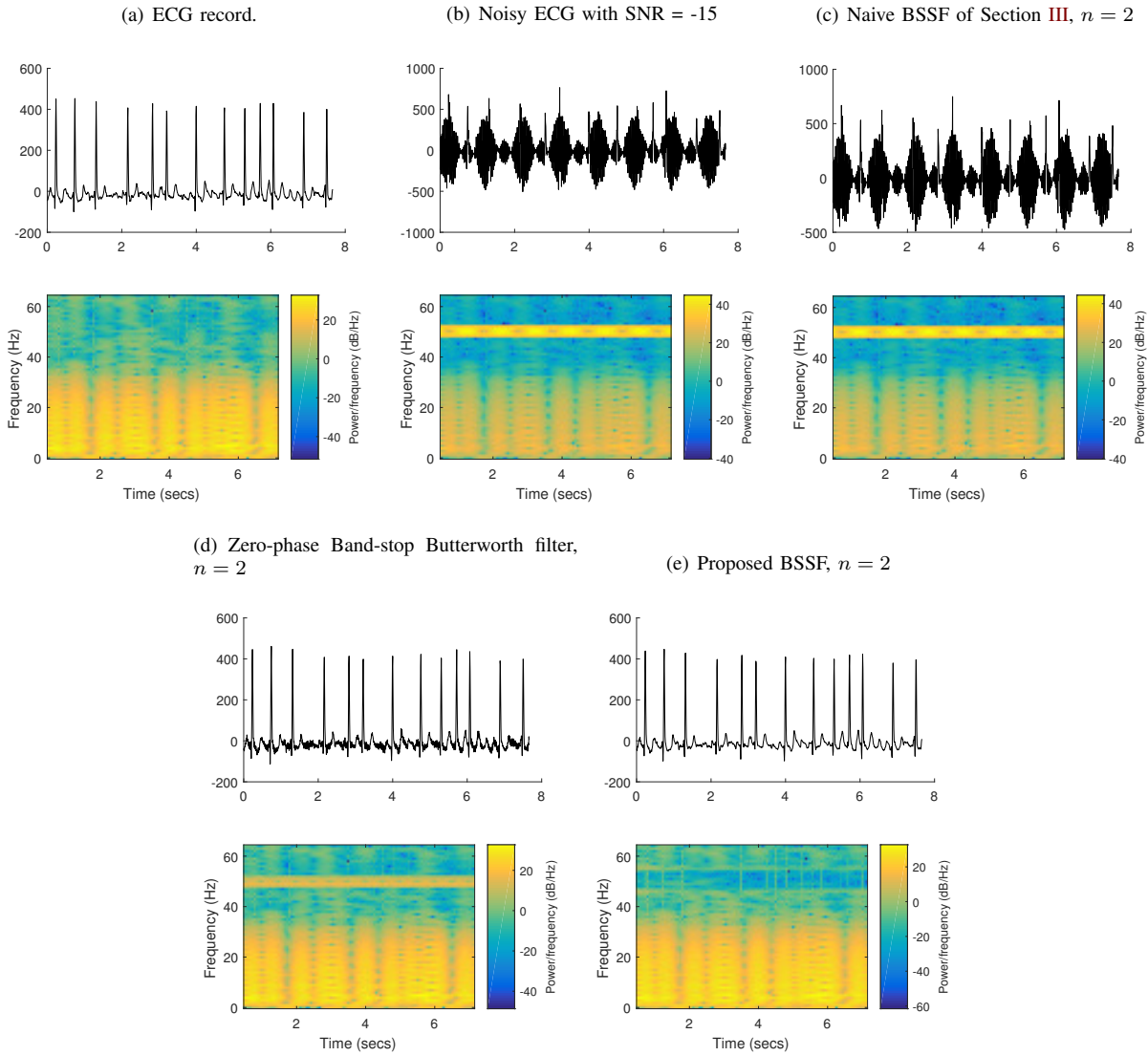(e) Proposed BSSF, $n = 2$

Fig. 10. ECG denoising for record b01m from PhysioNet/Computers in Cardiology Challenge 2004. a) ECG record b) noisy ECG (SNR = -15 dB) c) provided by second order simple BSSF designed in Section III (NSR = 2.1) d) second order zero-phase band-stop Butterworth filter (NSR = 0.1) e) proposed BSSF for $n = 2$ (NSR = 0.02). The spectral representation of the signals is also reported.

(a) Mean SNR_{diff}
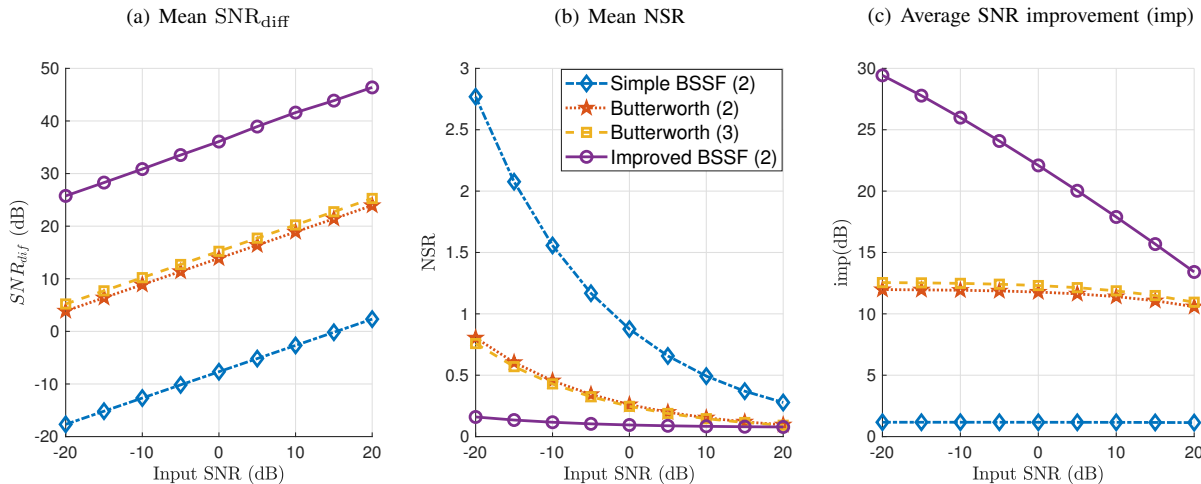
(b) Mean NSR

(c) Average SNR improvement (imp)

Fig. 11. Mean values of SNR_{diff}, imp and NSR for ECG reconstruction by simple BSSF, proposed BSSF and zero-phase Butterworth filter, as a function of the input SNR. The number between braces is the order of the filter.
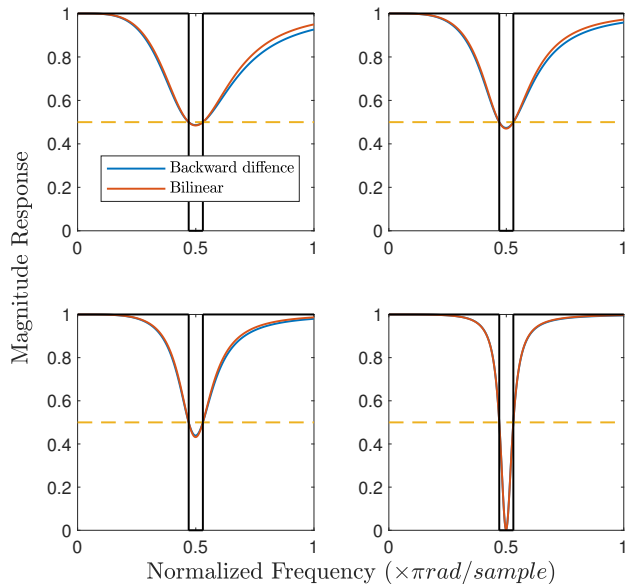
Fig. 12. The comparison of the amplitude response of the proposed BSSF using Tustin and backward difference rule with cutoff frequencies $\omega_1 = 0.47$ and $\omega_2 = 0.53$ (top left $\rho = 0.5$, top right panel $\rho = .75$, bottom left $\rho = 0.9$ and bottom right $\rho = 0.999$). For small values of ρ, the proposed BSSF using Tustin method has sharper transition band while for large values of ρ they are close to each other.

priors in our recent paper [38]. In [38], we have shown that implementing the derivative operator using bilinear transform (Tustin's method) improves the smoothness priors method. In the current paper, we mainly focused on proposing a new idea to band-stop smoothing filter design. However, we have included the Tustin's method and compared it with backward difference rule. Figure 12 shows the amplitude response of the proposed BSSF using Tustin and backward difference rule with cutoff frequencies $\omega_1 = 0.47$ and $\omega_2 = 0.53$ (top left $\rho = 0.5$, top right panel $\rho = 0.75$, bottom left $\rho = 0.9$ and bottom right $\rho = 0.999$). The results show that for small values of ρ, the proposed BSSF using Tustin's method has sharper transition band while for large values of ρ they are close to each other. It means that for higher values of ρ that the BSSF has the sharpest frequency transition band, both methods are close to each other. That is why we chose the backward difference rule in our paper to implement the BSSF. The first and second order derivatives are two main basic operations and blocks in signal filtering/smoothing. In fact, the integer order differential and integral operators are defined uniquely. They are local as they consider the values of the neighboring points to the point of interest. The problem formulation and algorithms described in this paper could be implemented using other methods such as fractional order derivative/integral. The fractional calculus operators are non-local and larger neighborhoods must be considered in the computation. Thus, it results in long-term memory effect. In the literature, a plenty of definitions of the fractional order derivatives [48] can be found: forward Grunwald-Letnikov (GL), Riemann-Liouville (RL) fractional derivatives, Fourier domain fractional derivatives.

In a preliminary experiment, the proposed BSSF has been implemented using fractional calculus and the Fourier domain fractional derivatives was used to implement the frequency response of the proposed BSSF. To demonstrate the impact of fractional order on the shape of the smoothing filter, we plot the amplitude response of the BSSF at various values of n in Figure 13. Figure 13 shows that for lower values of n, the BSSF has a shallow frequency transition band. We will consider further investigations in future work.

In this paper, we compared the proposed BSSF with maximally-flat Butterworth BSSF. The reason for choosing the maximally flat Butterworth filter for comparison is that our proposed least squares optimization based BSSF design has also a maximally flat frequency response. But the difference between them is that our proposed BSSF is directly designed in time domain while the zero-phase Butterworth is first designed in frequency domain, following by polynomial root finding, and finally applying forward filtering backward smoothing which could be quite involved. Another method that has been used for digital filter design is based on optimization technique method [49]. Numerous optimization methods have been developed based on modern heuristics optimization algorithms: Simulated Annealing (SA) algorithm, genetic algorithm (GA), particle swarm optimization (PSO), artificial bee colony and etc.. However, for optimal filter, designed in the frequency domain by the optimization methods, the error function and hence the optimization problem are generally nonconvex with multiple local optimal points. As a result, while it demands a large amount of computations, the designed filters need to be converted into time domain [49]. In the literature, the deep learning (DL) method has been also used for signal denoising, e.g. for ECG denoising [50]–[52]. However, in these papers, even if results seem interesting at first glance, comparison with the best more classical methods are still missing. Moreover, one of the advantages of the method we proposed in this paper is to obtain a closed-form equation relating each cut-off frequency to one parameter used in the optimization algorithm: At first glance, such a cut-off frequency control seems impossible in DL. Anyway, recent theoretical works in explainable DL introduce the idea of algorithm unrolling [53] which relates each iteration of an optimization algorithm to one layer in a deep neural networks. In further work, following this idea, it seems interesting to investigate how algorithms we proposed could be implemented in deep neural networks and compare the performance (accuracy, computational load, etc.) with respect to more classical implementations. Finally, the integration of the proposed optimization approach into advanced control methods, such as those in [54]–[56] and its sensitivity on the parameters and faults will be investigated in the future.

B. Conclusion

Quadratic variation (QV) regularization and smoothness priors are the most commonly used signal denoising techniques. In this paper, an extension of such algorithms to band-stop smoothing filters was investigated. We showed that with the optimization approaches (QV regularization or smoothness
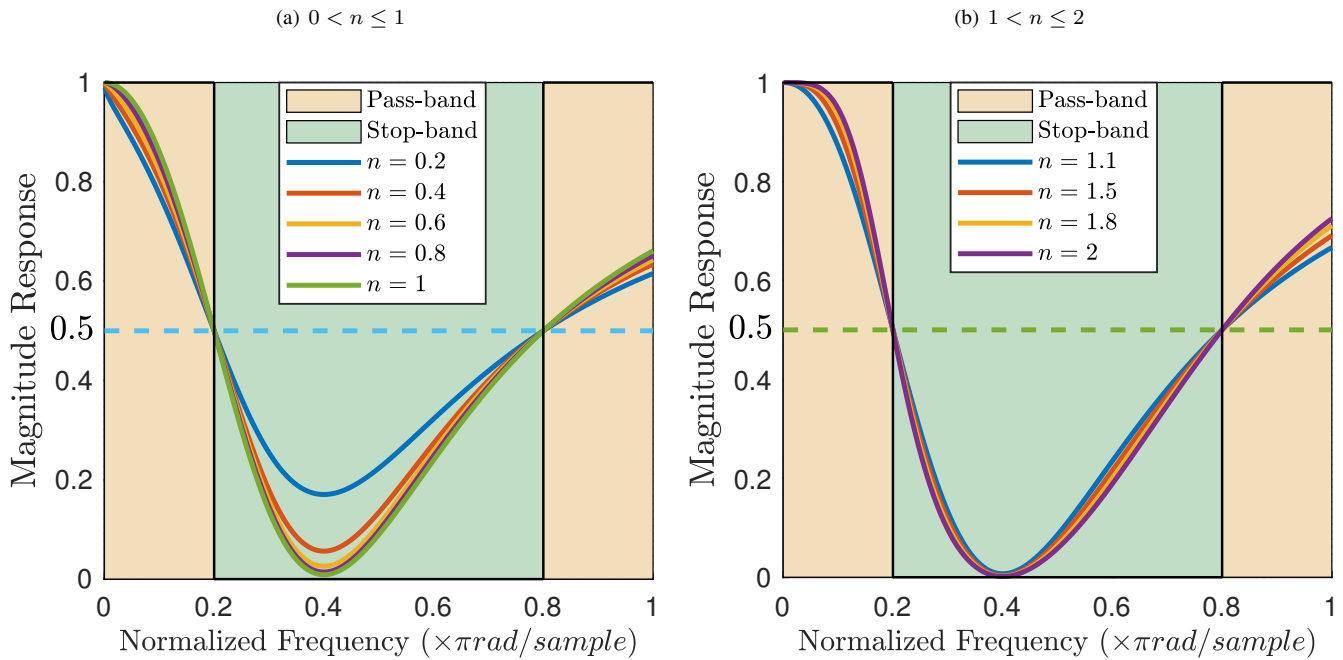
(a) $0 < n \leq 1$ (b) $1 < n \leq 2$

Fig. 13. The comparison of the amplitude response of the proposed BSSF using fractional calculus for $\rho = 0.99$. For large values of n, the proposed BSSF has sharper transition band.

priors), the cutoff frequencies are related to the regularized parameters and the order can be directly (and easily) controlled with the number of derivatives. First, a simple (naive) approach was presented. The simple band-stop smoothing filter can be made out of a low-pass smoothing filter and a high-pass smoothing filter by connecting the two smoothing filter sections in parallel with each other. The simple band-stop smoothing filter suffers from some shortcomings. Its cutoff frequencies are not exactly the planned-for the desired cutoff frequencies, the amplitude at the center frequency is not exactly zero and its transition band is not enough narrow to eliminate the unwanted frequency components and pass desired frequency components. In order to overcome these problems and further improve the naive approach, we presented a new approach to band-stop smoothing filter design. In the proposed approach, we considered a modified model and investigated the role of noise correlation which represents the link between two smoothing filter sections. Our results suggest that the noise correlation increases the steepness of the role-off and boosts the performance. Moreover, positive noise correlation is the key to always get the best band-stop smoothing filter. Specifically, $\rho \approx 1$, gives band-stop smoothing filter with the steepest frequency transition band. The relationship between the cutoff frequencies of the filter and the parameters used in the optimization algorithm was shown in the paper. More precisely, choosing the desired cutoff frequencies lead to choosing the hyperparameters, i.e., the weights of the regularization factors.

REFERENCES

[1] G. Bohlmann, "Ein ausgleichungsproblem," *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pp. 260–271, 1899.

[2] E. T. Whittaker, "On a new method of graduation," *Proc. Edinburgh Math. Soc*, vol. 41, pp. 63–75, 1923.

[3] A. Aitken, "On the theory of graduation," *Proceedings of the Royal Society of Edinburgh*, vol. 47, pp. 36–45, 1926.

[4] R. Henderson, "A new method of graduation," *Transactions of the Actuarial Society of America*, vol. 25, pp. 29–40, 1924.

[5] ——, "Mathematical theory of graduation," *Actuarial Studies 4, Actuarial Society of America, New York*, 1938.

[6] R. J. Shiller, "A distributed lag estimator derived from smoothness priors," *Econometrica*, vol. 41, pp. 775–788, 1973.

[7] G. Kitagawa and W. Gersch, "A smoothness priors-state space modeling of time series with trend and seasonality," *Journal of the American Statistical Association*, vol. 79, no. 386, pp. 378–389, 1984.

[8] ——, "A smoothness priors time-varying ar coefficient modeling of nonstationary covariance time series," *IEEE Transactions on Automatic Control*, vol. 30, no. 1, pp. 48–56, 1985.

[9] R. J. Shiller, "Smoothness priors and nonlinear regression," *Journal of the American Statistical Association*, vol. 79, pp. 609–615, 1984.

[10] R. L. Eubank, "A note on smoothness priors and nonlinear regression," *J Am Stat Assoc*, vol. 81, pp. 514–517, 1986.

[11] G. Kitagawa, T. Higuchi, and K. Fumiyo, "Smoothness prior approach to explore mean structure in large-scale time series," *Theor. Comput. Sci.*, vol. 292, pp. 431–446, 01 2003.

[12] G. Kitagawa and W. Gersch, *The Smoothness Priors Concept*. New York, NY: Springer New York, 1996, pp. 27–32.

[13] D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, and M. Taqqu, *New Directions in Time Series Analysis*. Springer New York, 2012.

[14] M. P. Tarvainen, P. O. Ranta-aho, and P. A. Karjalainen, "An advanced detrending method with application to hrv analysis," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 2, pp. 172–175, 2002.

[15] F. Zhang, S. Chen, H. Zhang, X. Zhang, and G. Li, "Bioelectric signal detrending using smoothness prior approach," *Med. Eng. Phys.*, vol. 36, no. 8, pp. 1007–1013, 2014.

[16] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6160–6173, 2016.

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

[18] X. Mao, K. Qiu, T. Li, and Y. Gu, "Spatio-temporal signal recovery based on low rank and differential smoothness," *IEEE Trans. Signal Process.*, vol. 66, no. 23, pp. 6281–6296, 2018.

[19] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l0 gradient minimization," *ACM Trans. Graph.*, vol. 30, pp. 174:1–174:12, 2011.

[20] A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, "A study of methods of choosing the smoothing parameter in image restoration by regularization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 326–339, 1991.

[21] G. Kitagawa and W. Gersch, "A smoothness priors long ar model method for spectral estimation," *IEEE Transactions on Automatic Control*, vol. 30, no. 1, pp. 57–65, 1985.

[22] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with b-splines and penalties," *Statistical Science*, vol. 11, pp. 89 – 121, 1996.

[23] D. Terzopoulos, "Multilevel computational processes for visual surface reconstruction," *Computer Vision, Graphics, and Image Processing*, vol. 24, no. 1, pp. 52–96, 1983.

[24] W. Grimson, *From Images to Surfaces: A Computational Study of the Human Early Visual System*. MIT Press, 1981.

[25] B. Horn, *Robot Vision*. MIT Press, 1986.

[26] O. Woodford, P. Torr, I. Reid, and A. Fitzgibbon, "Global stereo reconstruction under second-order smoothness priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2115–2128, Dec. 2009.

[27] A. Fasano and V. Villani, "Baselinewander removal for bioelectrical signals by quadratic variation reduction," *Signal Processing*, vol. 99, pp. 48–57, 2014.

[28] R. Hodrick and E. Prescott, "Postwar u.s. business cycles: An empirical investigation," *JMCB*, vol. 29, pp. 1–16, 1997.

[29] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures." *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[30] G. Golub, P. Hansen, and D. O'Leary, "Tikhonov regularization and total least squares," *SIAM J. Matrix Anal. Appl.*, vol. 21, pp. 185–194, 1999.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[32] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[33] H. W. Engl, "Discrepancy principles for tikhonov regularization of ill-posed problems leading to optimal convergence rates," *Journal of Optimization Theory and Applications*, vol. 52, pp. 209–215, 1987.

[34] H. Akaike, "Factor analysis and AIC," *Psychometrika*, vol. 52, no. 3, pp. 317–332, Sep 1987.

[35] Y. C. Eldar, "Generalized sure for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, pp. 471–481, 2009.

[36] A. Kheirati Roonizi and C. Jutten, "Improved smoothness priors using bilinear transform," *Signal Processing*, vol. 169, p. 107381, 2020.

[37] A. Oppenheim and R. Schafer, *Discrete-Time Signal Processing*. Pearson Education, 2011.

[38] A. Kheirati Roonizi and C. Jutten, "Forward-backward filtering and penalized least-squares optimization: A unified framework," *Signal Process.*, vol. 178, p. 107796, 2021.

[39] M. Unser, A. Aldroubi, and M. Eden, "Recursive regularization filters: design, properties, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 3, pp. 272–277, 1991.

[40] 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, 2019.

[41] S. Pei, B. Guo, and W. Lu, "Narrowband notch filter using feedback structure," *IEEE Signal Process. Mag.*, vol. 33, pp. 115–118, 2016.

[42] J. Smith, *Introduction to Digital Filters: with Audio Applications*. U.K., Exeter:W3K Publishing, 2007.

[43] Chien-Cheng Tseng and Soo-Chang Pei, "Stable IIR notch filter design with optimal pole placement," *IEEE Transactions on Signal Processing*, vol. 49, no. 11, pp. 2673–2681, 2001.

[44] A. Kheirati Roonizi, "An efficient algorithm for manuvering target tracking," *IEEE Signal Process. Mag.*, vol. 37, pp. 122–130, 2021.

[45] G. Moody, "Spontaneous termination of atrial fibrillation: A challenge from physionet and computers in cardiology 2004," *Computers in Cardiology*, vol. 31, pp. 101–104, 2004.

[46] A. Widmann, E. Schröger, and B. Maess, "Digital filter design for electrophysiological data – a practical approach," *Journal of Neuroscience Methods*, vol. 250, pp. 34 – 46, 2015.

[47] E. Kheirati Roonizi and R. Sassi, "A Signal Decomposition Model-Based Bayesian Framework for ECG Components Separation," *IEEE Transactions on Signal Processing*, vol. 64, no. 3, pp. 665–674, 2016.

[48] M. D. Ortigueira, "An introduction to the fractional continuous-time linear systems: the 21st century systems," *IEEE Circuits and Systems Magazine*, vol. 8, no. 3, pp. 19–26, 2008.

[49] "Design of digital IIR filter: A research survey," *Applied Acoustics*, vol. 172, p. 107669, 2021.

[50] C. T. C. Arsene, R. Hankins, and H. Yin, "Deep learning models for denoising ecg signals," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[51] V. Ravichandran, B. Murugesan, S. M. Shankaranarayana, K. Ram, P. S. P, J. Joseph, and M. Sivaprakasam, "Deep network for capacitive ECG denoising," *CoRR*, vol. abs/1903.12536, 2019.

[52] H. Chiang, Y. Hsieh, S. Fu, K. Hung, Y. Tsao, and S. Chien, "Noise reduction in ecg signals using fully convolutional denoising autoencoders," *IEEE Access*, vol. 7, pp. 60 806–60 813, 2019.

[53] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *To be appeared in IEEE Signal Processing Magazine*, 2021.

[54] K. Sun, L. Liu, J. Qiu, and G. Feng, "Fuzzy adaptive finite-time fault-tolerant control for strict-feedback nonlinear systems," *IEEE Transactions on Fuzzy Systems*, pp. 1–1, 2020.

[55] S. K., Q. Jianbin, H. R. Karimi, and Y. Fu, "Event-triggered robust fuzzy adaptive finite-time control of nonlinear systems with prescribed performance," *IEEE Transactions on Fuzzy Systems*, pp. 1–1, 2020.

[56] K. Sun, J. Qiu, H. R. Karimi, and H. Gao, "A novel finite-time control for nonstrict feedback saturated nonlinear systems with tracking error constraint," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–12, 2019.

Arman Kheirati Roonizi received the Ph.D. degree in computer science from University of Milan, Italy, in 2017. After graduating, he worked as Postdoctoral Researcher with GIPSA-lab at the Universitè Grenoble Alpes, France. Since February 2018, he is an Assistant Professor at Fasa University, Iran. His research interests include signal modeling/decomposition, filtering and processing, cardiac modeling and simulation, multimodal signal processing and functional data analysis.

Christian Jutten (AM'92-M'03-SM'06-F'08) received Ph.D. (1981) and Doctor es Sciences (1987) degrees from Grenoble Institute of Technology, France. He was Associate Professor (1982-1989), Professor (1989-2019), and is Professor Emeritus since September 2019 at University Grenoble Alpes. Since 1980's, his research interests are in machine learning and source separation, including theory and applications (brain and hyperspectral imaging, chemical sensing, speech). He is author/co-author of four books, 110+ papers in international journals and 250+ publications in international conferences. He is a fellow of the IEEE and EURASIP. He was a visiting professor at EPFL (Switzerland, 1989), RIKEN labs (Japan, 1999) and University of Campinas (Brazil, 2010). He served as the head of the signal/image processing laboratory, scientific advisor for signal/image processing at the French Ministry of Research (1996–1998) and CNRS (2003–2006 and 2012-2019). He was organizer or program chair of many international conferences, including the first Independent Component Analysis Conference in 1999 and IEEE MLSP 2009. He was the technical program co-chair ICASSP 2020. He was a member of the IEEE MLSP and SPTM Technical Committees. He was associate editor in Signal Processing and IEEE Trans. on Circuits and Systems, and guest co-editor for IEEE Signal Processing Magazine (2014) and the Proceedings of the IEEE (2015). He has been a Senior Member of Institut Universitaire de France since 2008, and the recipient of the 2012 ERC Advanced Grant for the project Challenges in Extraction and Separation of Sources (CHESS). He received many awards, e.g. Best Paper Awards of EURASIP in 1992 and IEEE GRSS in 2012, Medal Blondel in 1997 from the French Electrical Engineering Society, and one Grand Prix of the French Académie des Sciences in 2016. Since January 2021, he is the editor-in-chief of IEEE Signal Processing Magazine.