Improved Smoothness priors using Bilinear Transform

Arman Kheirati Roonizi, Christian Jutten

Improved Smoothness priors using Bilinear Transform

Arman Kheirati Roonizi[1] and Christian Jutten[2]

[1] *Department of Computer Science, Faculty of Science, Fasa University, Fasa, Iran*
Email: ebad.kheirati.roonizi@gmail.com

[2] *Gipsa-Lab, CNRS, Univ. Grenoble Alpes, France*

Improved Smoothness priors using Bilinear Transform

Arman Kheirati Roonizi[1] and Christian Jutten[2]

[1] *Department of Computer Science, Faculty of Science, Fasa University, Fasa, Iran*
Email: ebad.kheirati.roonizi@gmail.com

[2] *Gipsa-Lab, CNRS, Univ. Grenoble Alpes, France*

Abstract

Smoothness priors is a well-known and most commonly used method in the analysis of stochastic processes making it very useful in the field of stochastic signal processing. It is particularly suited for smoothing the noisy data and detrending the time-series signals. The method is based on an optimization problem where the n-th order derivative of the signal enters as a constraint. When the method is designed in discrete time domain, the backward difference rule is used to perform differential-to-difference conversion. Moreover, the solution depends on a smoothness trade-off parameter. An efficient algorithm for the trade-off parameter selection remains an important and challenging issue. In this paper, first, we propose a closed-form expression for the trade-off parameter. The closed-form expression resulted from a frequency domain interpretation of the smoothness priors procedure. The trade-off parameter determines the amount of frequency components that the procedure allows to pass. We show that the trade-off parameter is related to the arbitrary choice of cutoff frequency. Second, we introduce a new way to the design and implementation of smoothness priors using bilinear transformation method. Frequency analysis and experiments on both synthetic and real world signals with different levels of noise demonstrate that bilinear transform is indeed more effective for smoothness priors implementation when compared with the traditional ones, i.e., the backward difference rule.

Keywords: Smoothness priors, Hodrick-Prescott filter, Signal smoothing, trade-off parameter, Backward difference rule, Bilinear transform, Quadratic variation

1. Introduction

Smoothness priors has a broad range of applications in time series analysis [1, 2, 3, 4], global stereo reconstruction [5], edge-preserving image smoothing [6], image restoration [7], transfer function estimation [8], smoothing noisy data and signal detrending [9, 10, 11, 12], smoothing of discontinuous signals [13], spectral estimation [14], parametric time warping [15] and spline smoothing [16, 17, 18, 19, 20, 21, 22, 23]. It is closely linked to the ill-posed inverse problems and problems of statistical Tikhonov regularization [10, 24, 25, 26]. Although the notation of "smoothness priors" was first introduced by Shiller [27], its conceptual predecessor can be seen in the problem of estimating a smooth trend embedded in white noise addressed by Whittaker in 1923 [28] which was known as the method of graduating data [29, 30, 31, 32, 33]. The problem is described by a constrained convex optimization problem where the output smoothness (the n-th order difference of the signal) enters as a constraint. There are a lot of connections in the literature to the smoothness priors (Whittaker problem), as briefly presented below.

In the field of economics, Robert J. Hodrick and Edward C. Prescott popularized Hodrick-Prescott filtering (also known as Hodrick-Prescott decomposition) to remove the cyclical component from time-series data and estimate the trend signal [34]. It appears that they were unaware of Whittaker work. Hodrick-Prescott filter is one of the standard methods for data detrending. It has been applied to both real data and artificial data where a model is compared with the actual data [35, 36, 37, 38, 39, 40]. The l_1 trend filter is similar to Hodrick-Prescott filter which produces a piecewise linear estimate of the trend makes it suitable for analyzing the time series with an underlying piecewise linear trend [41]. The relation between the Hodrick-Prescott filter and the l_1 trend filter corresponds to the relation between the ridge regression [42] and the popular lasso regression [43]. Hodrick-Prescott filter is a special case of Whittaker work, where the second-order difference of the signal enters as a constraint.

In mathematics, the quadratic variation is a well-known property used in the analysis of stochastic processes such as Brownian motion and Wiener process [44]. The quadratic variation of a Brownian motion is bounded (finite). This property is the cornerstone of Itô calculus [45] which has important applications in mathematical finance and stochastic differential equations.

In biomedical signal processing, the quadratic variation has recently been used as a penalty in a convex optimization problem to remove the electrocardiogram (ECG) baseline wander [46]. The authors used the notation of quadratic variation reduction for their procedure [46]. Concurrently the smoothness

[1]This is the author's version of the manuscript accepted for publication in Elsevier Signal Processing, November 2019. The paper can be found in its final format at: https://doi.org/10.1016/j.sigpro.2019.107381

priors was used in [9] for electromyogram (EMG) and ECG signal detrending. More recently Sameni has used smoothness priors to design a fixed-interval smoother for denoising smooth signals contaminated by wide-band noise and he studied the denoising of ECG signals as a case study [47]. The "order" of a quadratic variation and smoothness prior is given by the order of the derivative of the signal. According to their definition (see [46] and [47]), it can be easily found that the quadratic variation reduction and smoothness priors are the same.

In chemical engineering, the Savitzky-Golay filter [48] is a method of data smoothing based on local least-squares polynomial approximation. The Savitzky-Golay filter, based on established Whittaker procedure was popularized by Savitzky and Golay to smooth the noisy data obtained from chemical spectrum analyzers. They demonstrated that least-squares smoothing increases signal to noise ratio while maintaining the shape and height of waveform peaks (in their case, Gaussian shaped spectral peaks).

The well-known wavelet shrinkage scheme using signal smoothness priors developed by Donoho et al. is another example [49]. Smoothness priors and other related methods (e.g., Hodrick-Prescott filter and quadratic variation reduction) are based on a constrained least squares estimation (LSE) problem which depend on a single trade-off parameter that needs to be adjusted properly for obtaining good performance. Neither Whittaker nor Shiller offered an objective method for choosing the smoothness trade-off at the time they introduced the smoothness priors problem. They left the choice of the smoothness trade-off to the investigator. Akaike applied the concept of Bayesian likelihood and used a maximum likelihood estimation (MLE) algorithm for determining the smoothness trade-off parameter [50, 51]. In [52, 53] a generalized cross-validation computation was used to select the smoothness trade-off. Stein's unbiased risk estimate (SURE) regularization is another approach which similar to generalized cross-validation needs the Jacobian matrix of the nonlinear reconstruction operator with respect to the data. In [46], Fasano and Villani argued that the value of regularization parameter is not crucial while in [47], Sameni proposed to choose the value of trade-off parameter based on the availability of the upper bound of signal roughness. If the upper bound is known, the optimal value is calculated using the upper bound of signal roughness otherwise methods such as the L-curve [54, 55, 56] are used for finding the trade-off parameter.

The difference approximation of the derivative operator for the algorithm is very important in smoothness priors implementation. When the smoothness priors is designed in discrete time domain, the backward difference rule, which is obtained from the backward rectangular integration rule is used to perform differential-to-difference conversion. The basis of the backward difference rule is to approximate the exponential function $z = e^{sT}$ by truncating its series expansion to two terms, which leads to $z = 1 + Ts$. When we analyze the smoothness priors in the frequency domain, we face that the transition band of the smoothing filter is not sufficiently narrow to effectively eliminate the undesired frequency components. This is due to the backward difference rule that is used in the literature for converting the differential operator to the difference operator. In order to improve the smoothness priors performance and increasing the transition-band characteristic of the smoothing filter, in this paper, we propose to use bilinear transform (also known as Tustin's method [57]) to perform differential-to-difference conversion. The bilinear transformation is obtained from the trapezoidal integration rule. It is commonly used to perform analog-to-digital conversion in the design of infinite impulse response (IIR) digital filters. The basis of the bilinear transformation is to represent the exponential function $z = e^{sT}$ by $z = \frac{e^{sT/2}}{e^{-sT/2}}$ and then approximate both exponentials with the first two terms of their Taylor expansions, which leads to $z = \frac{1+sT/2}{1-sT/2}$. In this paper, we show that in smoothing filters design, when the bilinear transform is used to convert the derivative operator to the difference operator, the transition band of the resulting smoothing filters will become narrower and sharper. The contribution of this paper can be summarized as follows. Firstly, to propose a closed-form expression for the smoothness trade-off parameter. Secondly to propose a new way based on the bilinear transform to perform differential-to-difference operator for the constraint part of the smoothness priors procedure. It is notable that the whole procedure is still in the form of fixed-interval smoother and may only be used in the form of a non-causal smoother. It can also be stated in terms of a zero-phase, forward-backward LTI smoothing filter. Comparing to the previous implementations which are all-pole with conjugate symmetric pole pairs [47], the new smoothing filter has a zero-pole impulse response with conjugate symmetric zero and pole pairs.

The rest of the paper is organized as follows. In Section 2, the relevant background on smoothness priors and quadratic variation is reviewed. Section 3 presents a closed-form expression for the trade-off parameter which is derived from a frequency domain interpretation of the smoothness priors procedure. We show that trade-off parameter determines the amount of frequency components that the procedure allows to pass. Section 4 presents a new method to the design and implementation of smoothness priors and quadratic variation. The bilinear transform method is proposed to implement the smoothness priors, Hodrick-Prescott filter, and quadratic variation. Performance analysis is presented in Section 6. General remarks and a discussion are given in the final section.

Throughout this paper, boldface uppercase letters are used to denote matrices, e.g., $\boldsymbol{A}$; boldface lowercase letters for vectors, e.g., $\boldsymbol{a}$; lowercase letters for scalars, e.g., a. 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., $D^n = \frac{d^n}{dt^n}$, $||\cdot||^2$ is the Euclidean norm, ∇ is the backward difference operator and $*$ denotes the convolution.

2. Background

2.1. The Smoothing Problem

In this section, the problem of smoothness priors is revisited in continuous time (CT) domain and discrete time (DT) domain, respectively.

2.1.1. Continuous time (CT)

The original problem addressed by Whittaker in [28] is to estimate an unknown smooth trend $x(t)$ from the observation $y(t)$ in the model:

$$y(t) = x(t) + v(t), \quad t \in [a, b] \tag{1}$$

where $v(t)$ is an additive noise assumed to be uncorrelated with $x(t)$. Whittaker suggested that a solution can be found by appropriate balancing a trade-off between the signal smoothness and minimum mean square error. By adopting the n-th order differential equation constraint on $x(\cdot)$, the penalty for the solution can be expressed as

$$\int_a^b [D^n x(\tau)]^2 d\tau. \tag{2}$$

The objective of smoothness priors is to estimate the unknown signal $x(t)$ from observation $y(t)$ and filter out the observation noise. The best solution is to estimate the desired signal, solving the following constrained LSE problem:

$$\hat{x}(t) = \underset{x(t)}{\operatorname{argmin}} \int_a^b [y(\tau) - x(\tau)]^2 \, d\tau + \lambda \int_a^b [D^n x(\tau)]^2 \, d\tau \tag{3}$$

where λ denotes the trade-off parameter which controls the competition between signal smoothness and minimum mean square error. The role of trade-off parameter is to avoid overfitting. In the following section, the same problem is tackled in the DT domain by following the approach similar to that used in the CT domain.

2.1.2. Discrete time (DT)

Let us consider $y_k = y(kT_s)$, the discrete-time samples of $y(t)$, x_k the sampled desired signal and v_k the sampled observation noise:

$$y_k = x_k + v_k, \quad k = 1, \cdots, L. \tag{4}$$

Now, the objective is to estimate x_k ($1 \le k \le L$) from its observation, y_k. By adopting the n-th order difference equation constraint on $x(\cdot)$, using the backward difference rule, the penalty for the solution can be expressed as

$$\sum_{k=1}^{L} [\nabla^n x_k]^2 \tag{5}$$

where $\nabla x_k = x_k - x_{k-1}$ is the first order difference and $\nabla^n x_k = \nabla(\nabla^{n-1})x_k$ is the n-th order difference. Therefore, in DT domain, (3) can be written as

$$\hat{x}_k = \underset{x_k}{\operatorname{argmin}} \sum_{j=1}^{L} \left[y_j - x_j\right]^2 + \lambda \sum_{j=1}^{L} \left[\nabla^n x_j\right]^2, \quad \text{for } k = 1, \cdots L. \tag{6}$$

2.2. Signal smoothing based on quadratic variation

The quadratic variation of $x(t)$ is defined by

$$QV_1(x) = \int_a^b [Dx(\tau)]^2 \, d\tau \tag{7}$$

where $Dx(t)$ as before, denotes the first derivative of $x(t)$ with respect to t. Given a vector $\boldsymbol{x} = [x_1, x_2, \cdots, x_L]^T \in \mathcal{R}^L$, the quadratic variation of $\boldsymbol{x}$ with first order difference approximation of the derivative is defined as [46]

$$QV_1(\boldsymbol{x}) = \sum_{i=1}^{L-1} \left(x_i - x_{i+1}\right)^2. \tag{8}$$

Note that similar to the smoothness priors, the difference approximation for quadratic variation is obtained by the backward difference rule. (8) can be expressed in the following matrix notation

$$QV_1(\boldsymbol{x}) = \|\boldsymbol{B}_1 \boldsymbol{x}\|^2 \tag{9}$$

where $\boldsymbol{B}_1$ is the matrix of size $(L-1) \times L$:

$$\boldsymbol{B}_1 = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -1 \end{pmatrix}. \tag{10}$$

It is more interesting to define the quadratic variation with high order difference approximation of the derivative operator. The quadratic variation of $\boldsymbol{x}$ with high order difference can be defined by

$$QV_n(\boldsymbol{x}) = \|\boldsymbol{B}_n \boldsymbol{x}\|^2 \tag{11}$$

where $\boldsymbol{B}_n$ is defined as the Toeplitz matrix form of $\boldsymbol{b}_n$ and $\boldsymbol{b}_n$ is defined by the following recursion:

$$\begin{cases} \boldsymbol{b}_1 \triangleq (+1 \ -1) & n = 1 \\ \boldsymbol{b}_n = \boldsymbol{b}_{n-1} * \boldsymbol{b}_1 & n > 1 \end{cases} \tag{12}$$

where $*$ denotes the convolution operator. Following the procedure presented in [46] (i.e., the quadratic variation reduction), $\boldsymbol{x}$ can be estimated searching for a signal component that has reduced variability, with respect to the measured signal. Mathematically, it means that the unknown signal $\boldsymbol{x}$ can be estimated by solving the following LSE problem with an inequality constraint:

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{minimize}} && \|\boldsymbol{y} - \boldsymbol{x}\|^2 \\ &\text{subject to} && \|\boldsymbol{B}_n \boldsymbol{x}\|^2 \le \epsilon \end{aligned} \tag{13}$$

where $\boldsymbol{y} = [y_1, y_2, \cdots, y_L]^T \in \mathcal{R}^L$ is the observation vector and ϵ is the upper bound for the signal's "steepness" which controls the quadratic variation of the estimated signal. The Lagrangian form of (13) is as follows

$$J = \|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda \|\boldsymbol{B}_n \boldsymbol{x}\|^2 \tag{14}$$

which is the matrix expression of the cost function used in (6). It can be concluded that the quadratic variation reduction is a simple instance of the smoothness priors. Their definitions are based on the difference approximation of the derivative operator. Greville [58] showed that there is an unique solution to

(14). The optimal solution that follows from the minimization of J with respect to $\boldsymbol{x}$ is

$$\hat{\boldsymbol{x}} = (I + \lambda \boldsymbol{B}_n^T \boldsymbol{B}_n)^{-1} \boldsymbol{y} \tag{15}$$

The only unknown parameter in (15) is the trade-off parameter, λ. In the following section, we propose a closed-form expression for λ which is a function of the cutoff frequency.

3. Estimation of the trade-off parameter based on the cutoff frequency

In this section, we derive a closed-form formula for estimating the trade-off Parameter. We show that the design parameter can be calculated in terms of the cutoff frequency.

Equation (15) can be expressed as

$$\boldsymbol{y} = (I + \lambda \boldsymbol{B}_n^T \boldsymbol{B}_n) \hat{\boldsymbol{x}} \tag{16}$$

Any component y_k of the vector $\boldsymbol{y}$ can be written in the following convolution form [47]:

$$y_k = (\delta_k + \lambda \boldsymbol{b}_{n,-k} * \boldsymbol{b}_{n,k}) * \hat{x}_k. \tag{17}$$

where δ_k is the Kronecker delta function and $\boldsymbol{b}_{n,-k} = \boldsymbol{b}_{n,L-k+1}$. Let $(\delta_k + \lambda \boldsymbol{b}_{n,-k} * \boldsymbol{b}_{n,k})^{\boxdot}$ denotes the mathematical inverse for convolution product of $(\delta_k + \lambda \boldsymbol{b}_{n,-k} * \boldsymbol{b}_{n,k})$. Because $(\delta_k + \lambda \boldsymbol{b}_{n,-k} * \boldsymbol{b}_{n,k})^{\boxdot} * (\delta_k + \lambda \boldsymbol{b}_{n,-k} * \boldsymbol{b}_{n,k})$ equals a delta function, (17) can be written in the following form [47]:

$$\hat{x}_k = (\delta_k + \lambda \boldsymbol{b}_{n,-k} * \boldsymbol{b}_{n,k})^{\boxdot} * y_k. \tag{18}$$

The impulse response of the smoothing filter is found by substituting y_k with δ_k in (18):

$$g_{n,k} = (\delta_k + \lambda \boldsymbol{b}_{n,-k} * \boldsymbol{b}_{n,k})^{\boxdot}. \tag{19}$$

Taking the $\mathcal{Z}$-transform of Eq. (19), the frequency response of the smoothing filter writes:

$$G_n(z) = \frac{1}{1 + \lambda (1 - z^{-1})^n (1 - z)^n}. \tag{20}$$

Eq. (20) becomes in the Fourier domain

$$G_n(e^{j\omega}) = \frac{1}{1 + \lambda [2 - 2\cos\omega]^n} = \frac{1}{1 + \lambda (2 \sin \frac{\omega}{2})^{2n}}. \tag{21}$$

The above derivations can be found in [47]. Suppose that a smoothing filter with a −6 dB cutoff frequency f_c is desired, then the optimal trade-off parameter, denoted by λ_o, is numerically calculated by solving the following equation for λ_o:

$$\frac{1}{1 + \lambda_o (2 \sin \frac{\omega_c}{2})^{2n}} = \frac{1}{2}, \tag{22}$$

where $\omega_c = 2\pi \frac{f_c}{fs}$. This leads to the following optimal trade-off parameter:

$$\lambda_o = \frac{1}{(2 \sin \frac{\omega_c}{2})^{2n}}. \tag{23}$$

Substituting the value of λ in (20) the following smoothing filter (forward-backward zero-phase low-pass filter) with cutoff frequency ω_c is obtained:

$$G_n^{BD}(z) = \frac{1}{1 + \frac{1}{(2 \sin \frac{\omega_c}{2})^{2n}} (1 - z^{-1})^n (1 - z)^n} \tag{24}$$

where n denotes the degree of the smoothness priors. The "BD " superscript was used to show that the "backward difference" is used to perform the differential-to-difference conversion. In the following section, we propose a new approach based on the bilinear transform to perform the differential-to-difference conversion.

4. Smoothness priors based on bilinear transform

In this section, bilinear transform [59, 60] is employed for implementing smoothness priors. As mentioned before, in the literature, the n-th order difference of the signal is obtained by approximating the n-th order derivative of x using backward difference rule. So, the n-th order derivative of x, $D^n x(t)$, is approximated by

$$\nabla^n x_k = \nabla^{n-1} \nabla x_k. \tag{25}$$

Let us denote the first order derivative of $x(t)$ with $f(t)$, i.e., $f(t) = Dx(t)$. Taking the Laplace transform of f yields to

$$F(s) = sX(s) \tag{26}$$

A discrete version of $f(t)$ can be obtained by taking the inverse transform of its $\mathcal{Z}$-transform. To this purpose, we need to transform (26) into $\mathcal{Z}$ domain. There are several methods that can be used.

Backward difference rule is a method that substitutes s with $1 - z^{-1}$ [59]. Using backward difference rule, we obtain

$$F(z) = (1 - z^{-1}) X(z) \tag{27}$$

Taking the inverse $\mathcal{Z}$-transform of (27), we find

$$f_k = \mathcal{Z}^{-1}\{(1 - z^{-1}) X(z)\} = (1 - E^{-1}) x_k = x_k - x_{k-1} \tag{28}$$

where $E^{-n} x_k = x_{k-n}$. In this case, the high order difference approximation of the derivative operator is described by:

$$QV_n(\boldsymbol{x}) = \sum_{j=n+1}^{L} \left[(1 - E^{-1})^n x_j \right]^2 \tag{29}$$

The traditional way for implementing the smoothness priors is to use (29) as a constraint [46, 47]:

$$\hat{x}_k = \underset{x_k}{\operatorname{argmin}} \sum_{j=1}^{L} (y_j - x_j)^2 + \lambda \sum_{j=n+1}^{L} \left[(1 - E^{-1})^n x_j \right]^2 \tag{30}$$

The optimal solution that follows from the optimization of (30) is equal to (15).

In this paper, we propose to use bilinear transform to discretize $f(t)$. Employing bilinear transform, s is substituted with $\frac{1-z^{-1}}{1+z^{-1}}$ [59]. In this case

$$F(z) = \frac{1 - z^{-1}}{1 + z^{-1}} X(z) \tag{31}$$

Finally taking the inverse $\mathcal{Z}$-transform of (31) yields

$$(1 + E^{-1}) f_k = (1 - E^{-1}) x_k \tag{32}$$

The new definition that we propose for the quadratic variation (or the penalty term in smoothness priors) is as follows

$$QV_n(\boldsymbol{x}) = \sum_{j=n+1}^{L} \left[\left(\frac{1 - E^{-1}}{1 + E^{-1}} \right)^n x_j \right]^2 \tag{33}$$

For instance, the first and second order quadratic variations are defined by

$$\begin{aligned} QV_1(\boldsymbol{x}) &= \sum_{j=2}^{L} \left[\frac{x_j - x_{j-1}}{x_j + x_{j-1}} \right]^2 \\ QV_2(\boldsymbol{x}) &= \sum_{j=3}^{L} \left[\frac{x_j - 2x_{j-1} + x_{i-2}}{x_j + 2x_{j-1} + x_{j-2}} \right]^2 \end{aligned} \tag{34}$$

Using (33) in the penalty term of the smoothness priors problem, the following optimization problem is obtained

$$\hat{x}_k = \underset{x_k}{\operatorname{argmin}} \sum_{j=1}^{L} (y_j - x_j)^2 + \lambda \sum_{j=n+1}^{L} \left[\left(\frac{1 - E^{-1}}{1 + E^{-1}} \right)^n x_j \right]^2 \tag{35}$$

which can equivalently be represented as the following optimization problem:

$$\hat{x}_k = \underset{x_k}{\operatorname{argmin}} \sum_{j=1}^{L} \left[(1 + E^{-1})^n (y_j - x_j) \right]^2 + \lambda \sum_{j=n+1}^{L} \left[(1 - E^{-1})^n x_j \right]^2 \tag{36}$$

For all k, we finally can write (36) in the following matrix notation:

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \, \|\boldsymbol{H}_n (\boldsymbol{y} - \boldsymbol{x})\|^2 + \lambda \, \|\boldsymbol{B}_n \boldsymbol{x}\|^2 \tag{37}$$

where $\boldsymbol{B}_n$ is defined in (10) and $\boldsymbol{H}_n$ is defined as the Toeplitz matrix form of $\boldsymbol{h}_n$ and $\boldsymbol{h}_n$ is defined by the following recursion:

$$\begin{cases} \boldsymbol{h}_1 \triangleq (+1 \;\; +1) & n = 1 \\ \boldsymbol{h}_n = \boldsymbol{h}_{n-1} * \boldsymbol{h}_1 & n > 1 \end{cases} \tag{38}$$

For instance, $\boldsymbol{H}_1$ is defined by

$$\boldsymbol{H}_1 = \begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & 1 \end{pmatrix} \tag{39}$$

It is straightforward to show that the optimal solution that follows from the optimization of (37) is

$$\hat{\boldsymbol{x}} = (\boldsymbol{H}_n^T \boldsymbol{H}_n + \lambda \boldsymbol{B}_n^T \boldsymbol{B}_n)^{-1} \boldsymbol{H}_n^T \boldsymbol{H}_n \boldsymbol{y} \tag{40}$$

Any component $\hat{x}_k$ of $\boldsymbol{x}$ given by (40) can be written in the following form:

$$\hat{x}_k = (\boldsymbol{h}_{n,-k} * \boldsymbol{h}_{n,k} + \lambda \boldsymbol{b}_{n,-k} * \boldsymbol{b}_{n,k})^{\boxdot} * \boldsymbol{h}_{n,-k} * \boldsymbol{h}_{n,k} * y_k. \tag{41}$$

The impulse response and frequency response of the proposed smoothing filter are respectively equal to

$$\begin{aligned} g_{n,k} &= (\boldsymbol{h}_{n,-k} * \boldsymbol{h}_{n,k} + \lambda \boldsymbol{b}_{n,-k} * \boldsymbol{b}_{n,k})^{\boxdot} * \boldsymbol{h}_{n,-k} * \boldsymbol{h}_{n,k} \\ G_n^{BT}(z) &= \frac{\left[(1 + z^{-1})(1 + z)\right]^n}{\left[(1 + z^{-1})(1 + z)\right]^n + \lambda \left[(1 - z^{-1})(1 - z)\right]^n}. \end{aligned} \tag{42}$$

where the "BT " superscript denotes "Bilinear Transformation". Eq. (42) becomes in the Fourier domain

$$G_n^{BT}(e^{j\omega}) = \frac{1}{1 + \lambda \left(\frac{1 - \cos\omega}{1 + \cos\omega} \right)^n} \tag{43}$$

The optimal value of the trade-off parameter corresponds to the -6 dB cutoff frequency f_c

$$\lambda_o = \left(\frac{1 + \cos\omega_c}{1 - \cos\omega_c} \right)^n \tag{44}$$

The frequency response of the smoothing filter using backward difference rule and bilinear transform for different values of cutoff frequencies is depicted in Fig. 1. It acts as a forward-backward low-pass filter which is suited for estimating the smooth components. It is seen that the steepness of the roll-off of the smoother using bilinear transform is sharper than that obtained with backward difference rule.

5. Application to high-pass and band-pass Smoothing filter Design

The smoothness priors can also be used to estimate the high frequency components. To this purpose, one can remove the estimated low frequency components from the measured signal by subtraction. When the smoothness priors is designed by backward difference rule, the high frequency components are estimated by

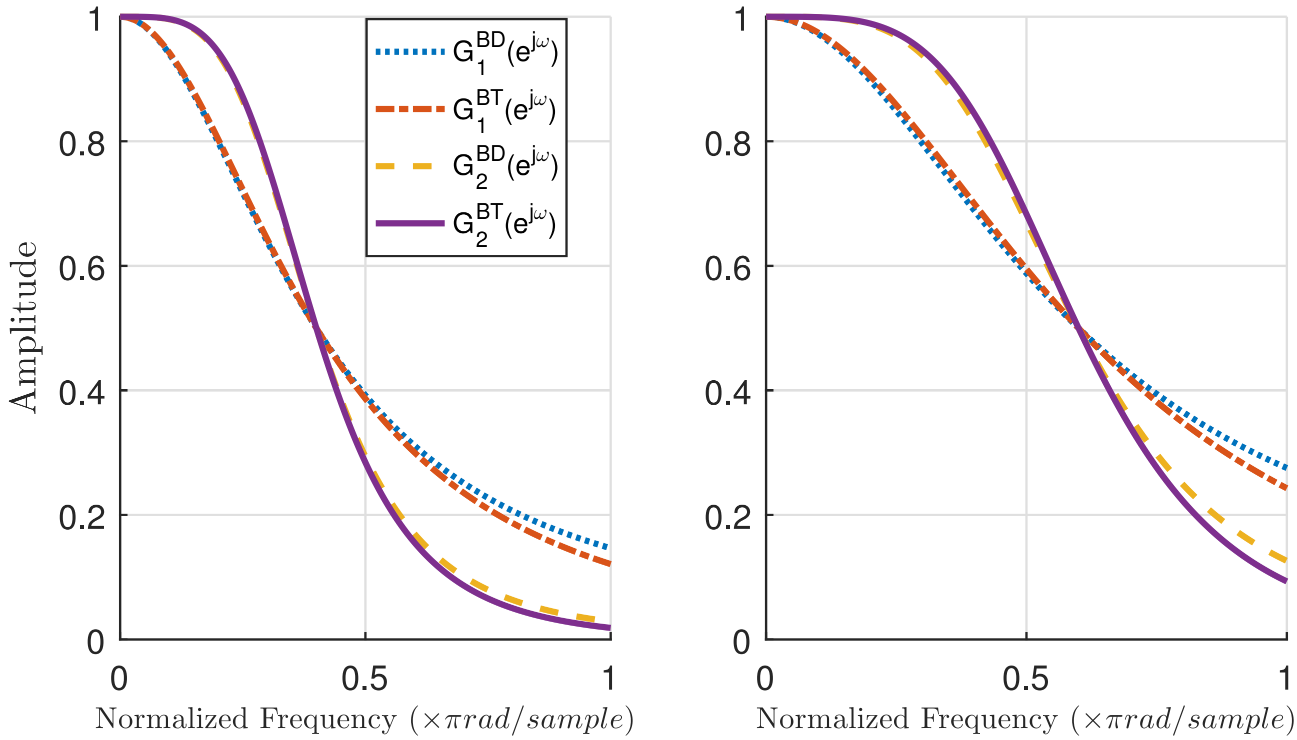
$$\tilde{\boldsymbol{x}} = \boldsymbol{y} - \hat{\boldsymbol{x}} = (I - (I + \frac{1}{(2\sin\frac{\omega_c}{2})^{2n}} B_n^T B_n)^{-1}) \boldsymbol{y} \tag{45}$$

The frequency response of the high-pass smoothing filter using backward difference rule is

$$G_n^{BD}(z) = \frac{\left[(1 - z^{-1})(1 - z)\right]^n}{(2\sin\frac{\omega_c}{2})^{2n} + \left[(1 - z^{-1})(1 - z)\right]^n} \tag{46}$$

When the smoothness priors is designed by bilinear transform, the high frequency components are estimated by

$$\tilde{\boldsymbol{x}} = \boldsymbol{y} - \hat{\boldsymbol{x}} = (I - (H_n^T H_n + \frac{1}{\tan^{2n}\frac{\omega_c}{2}} B_n^T B_n)^{-1} H_n^T H_n) \boldsymbol{y} \tag{47}$$



(a) $\omega_c = 0.4$.

(b) $\omega_c = 0.6$.

Figure 1: Amplitude response of the low-pass smoother, $G_n^I(e^{j\omega})$ and $G_n^{BT}(e^{j\omega})$, for different values of $n = 1, 2$.

The smoothing filter has the following frequency response:

$$G_n^{BT}(z) = \frac{\left[(1-z^{-1})(1-z)\right]^n}{\tan^{2n}\frac{\omega_c}{2}\left[(1+z^{-1})(1+z)\right]^n + \left[(1-z^{-1})(1-z)\right]^n} \tag{48}$$

The amplitude response of the high-pass smoothing filter for different values of n is shown in Fig. 2. Compared to the smoothing filter designed by backward difference rule, the smoothing filter designed by bilinear transform achieves a sharper transition between the passband and the stopband at the same order. Band-pass smoothing filters are another types of smoothing filters which have many applications in a wide range of economic and signal processing contexts [61, 62]. The task of a band-pass smoothing filter is to pass frequencies within a certain range and reject frequencies outside this range. A bandpass smoothing filter can be obtained by cascading a low-pass smoothing filter and a high-pass smoothing filter. The frequency response of the bandpass smoothing filter using backward difference rule and bilinear transform are respectively:

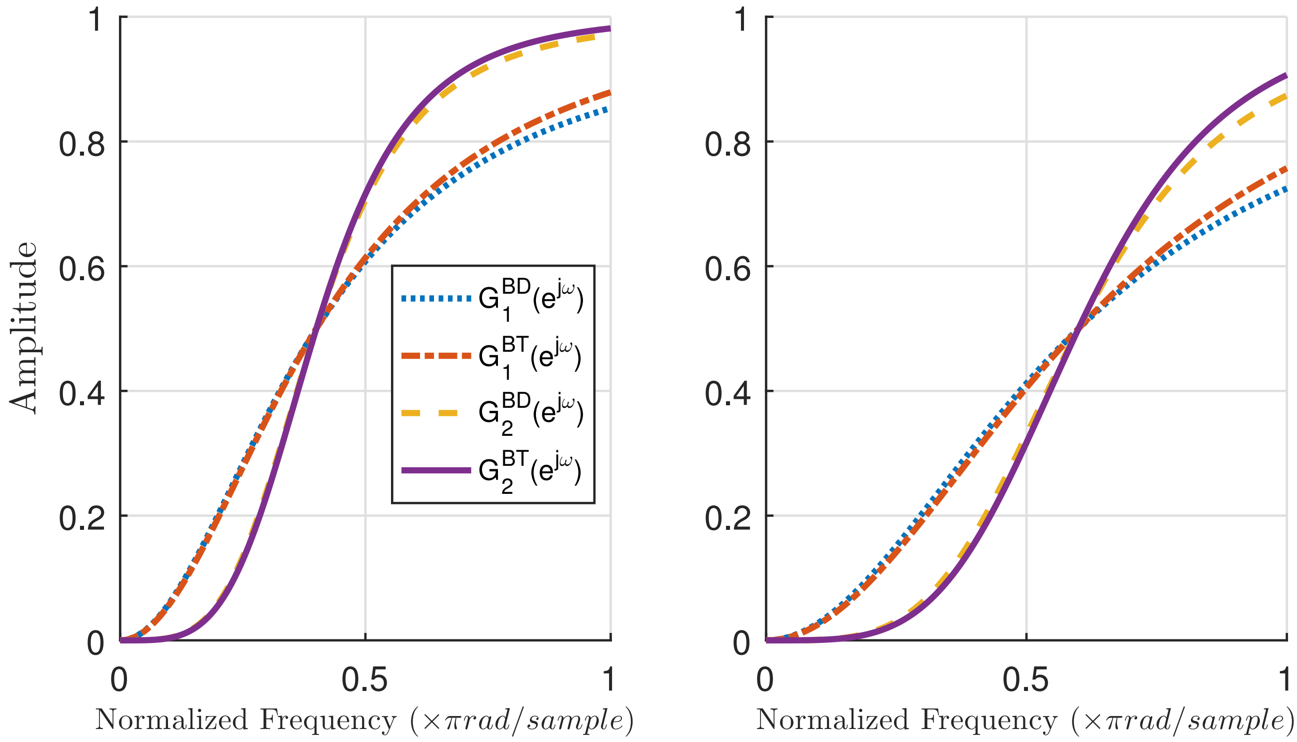
$$G_n^{BD}(z) = \frac{1}{1 + \frac{1}{(2\sin\frac{\omega_{c,1}}{2})^{2n}}(1-z^{-1})^n(1-z)^n} \times \frac{\frac{1}{(2\sin\frac{\omega_{c,2}}{2})^{2n}}(1-z^{-1})^n(1-z)^n}{1 + \frac{1}{(2\sin\frac{\omega_{c,2}}{2})^{2n}}(1-z^{-1})^n(1-z)^n}$$

$$G_n^{BT}(z) = \frac{(1+z^{-1})^n(1+z)^n}{(1+z^{-1})^n(1+z)^n + \frac{1}{\tan^{2n}\frac{\omega_{c,1}}{2}}(1-z^{-1})^n(1-z)^n} \times \frac{\frac{1}{\tan^{2n}\frac{\omega_{c,2}}{2}}(1-z^{-1})^n(1-z)^n}{(1+z^{-1})^n(1+z)^n + \frac{1}{\tan^{2n}\frac{\omega_{c,2}}{2}}(1-z^{-1})^n(1-z)^n} \tag{49}$$

where $\omega_{c,1}$, $\omega_{c,2}$ are the cutoff frequencies of the bandpass smoothing filter. The amplitude response of the band-pass smoothing filter for different values of n is shown in Fig. 3. Compared to the bandpass smoothing filter designed by backward difference rule, the bandpass smoothing filter designed by bilinear transform achieves narrower bandwidth.

6. Performance Analysis

The performance of the proposed approach was investigated using synthetic and real data. We compare the performance of the improved smoothness priors (the bilinear transform based approach) with previous smoothness priors (the backward difference rule based approach). We omitted comparisons with other denoising approaches since extensive comparisons between those approaches and the previous smoothness priors (backward difference rule based approach) were performed in the previous papers [46, 47]. We show the merits of the use of bilinear transform for smoothing filter design comparing to the backward difference rule for the implementation of smoothing filter.



(a) $\omega_c = 0.4$. (b) $\omega_c = 0.6$.

Figure 2: Amplitude response of the high-pass smoother, $G_n^I(e^{j\omega})$ and $G_n^{BT}(e^{j\omega})$, for different values of $n = 1, 2$.

6.1. Synthetic Data Simulation

To verify the performance of the proposed approach, we applied it on simulated data, which allows to quantify the estimation error directly. Application to ECG denoising will be discussed in the next section. To this purpose, synthetic data were obtained using sinusoidal (trigonometric) functions. Any signal can be approximated as a truncated Fourier series:

$$\hat{x}(t) = \sum_{i=0}^{N-1} c_i \cos(2\pi f_i t + \psi_i) \tag{50}$$

where c_i, f_i and ψ_i are respectively, the amplitude, frequency and phase of the i-th sinusoid.

We generate 5000 synthetic series 5s long using (50). The sampling rate was set to 200 Hz, and the frequencies were selected between 0 and $N-1$. That is the frequency ranges of the generated data is in $[0, N)$ Hz. The amplitudes, c_i were randomly selected. We considered different values of $N = 5, 10, 20, 30, 40$. The next step is to contaminate the signals with noise. The noise can be any other signals that its frequency range is out of $[0, N)$ Hz. To this purpose, we used the same approach using (50) to generate synthetic noise but the frequencies f_i were selected between N Hz and $f_s/2$ Hz and their amplitudes were randomly selected. In this experiment, the noise is a broad-band signal with varying power from 0 to 25 dB.

A low-pass smoothing filter with $f_c = N$ Hz can be used to remove the noise and reconstruct the signal. In any signal denoising based on a filter $g(t)$, the output is

$$\hat{x}(t) = g(t) * y(t) = g(t) * [x(t) + v(t)] = g(t) * x(t) + g(t) * v(t)$$

The observation (1) is a mixed of signal and noise. The input signal-to-noise ratio is defined by

$$SNR_{in} = \frac{P_x}{P_v} \tag{51}$$

where P_x denotes the power of x and P_v denotes the power of v. Now, consider the output of the smoothing filter as a new observation:

$$\hat{x}(t) = x_f(t) + v_f(t) \tag{52}$$

If the filter is all-pass, then

$$\hat{x}(t) = y(t) \quad \Longrightarrow \quad x_f(t) = x(t) \text{ and } v_f(t) = v(t) \tag{53}$$

In this special case

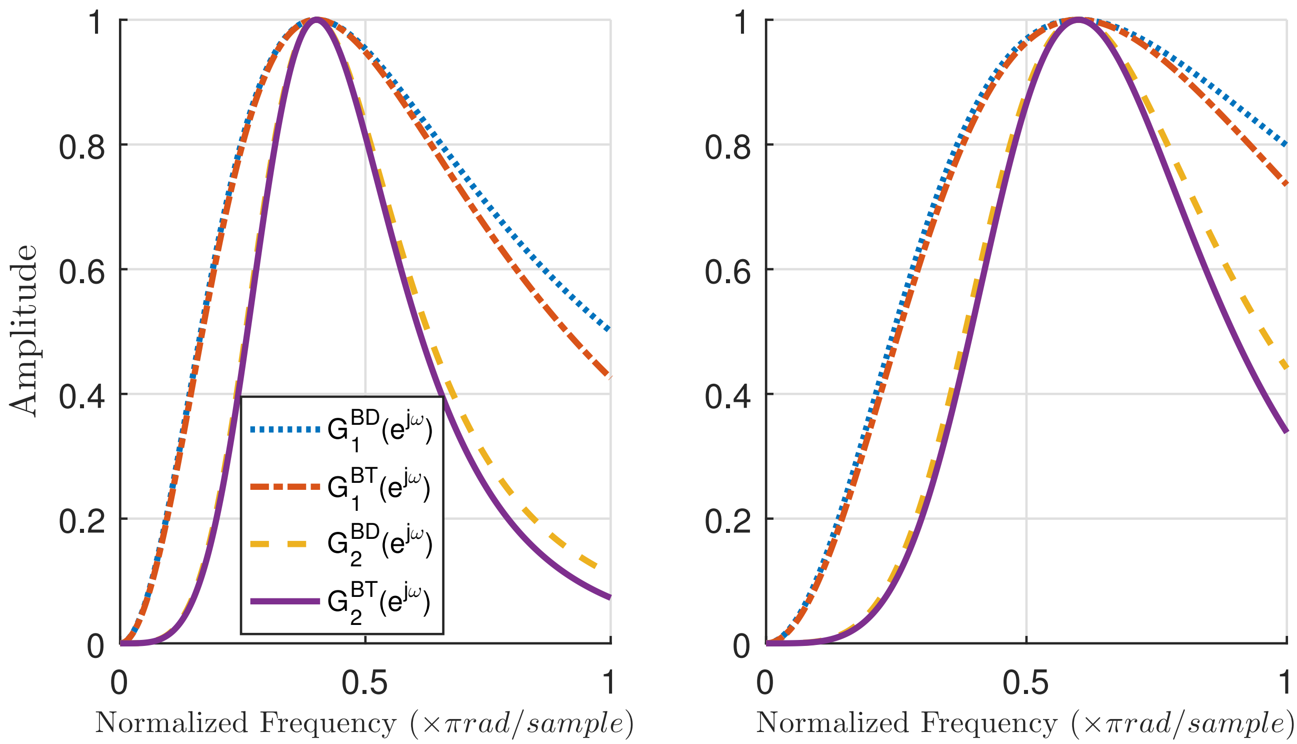
$$SNR_{out} = \frac{P_{x_f}}{P_{v_f}} = \frac{P_x}{P_v} = SNR_{in} \tag{54}$$

That means there is no SNR improvement. If the smoothing filter is ideal, then

$$\hat{x}(t) = x(t) \quad \Longrightarrow \quad x_f(t) = x(t) \text{ and } v_f(t) = 0 \tag{55}$$

In this special case

$$SNR_{out} = \frac{P_{x_f}}{P_{v_f}} = \frac{P_x}{0} = \infty \tag{56}$$



(a) $\omega_c = 0.4$.

(b) $\omega_c = 0.6$.

Figure 3: Amplitude response of the band-pass smoother, $G_n^I(e^{j\omega})$ and $G_n^{BT}(e^{j\omega})$, for different values of $n = 1, 2$.

That means we have the maximum SNR improvement.

In other cases, if the smoothing filter is of high quality, very close to the ideal one ($x_f \approx x$ and $v_f \approx 0$), then

$$\hat{x}(t) = \underbrace{x_f(t)}_{\approx x(t)} + \underbrace{v_f(t)}_{\approx 0} \tag{57}$$

Since the smoothing filter is not ideal and passes some parts of noise and attenuates some parts of signal of interest, we consider the approximation[2]

$$x_f(t) = \alpha x(t)$$
$$v_f(t) = \beta v(t)$$

Therefore, we have

$$\hat{x}(t) \approx \alpha x(t) + \beta v(t) \tag{58}$$

Our results show that α is close to 1 (whatever the input SNR, $\hat{\alpha} \approx 0.9$ or 0.95, for 1st and 2nd order smoothing filters based on bilinear transform approach), which means that the smoothing filter is of high quality, very close to the ideal one. So our method of computation (even if (58) is a simple approximation) provides SNR which is more realistic (taking into account that $g(t)$ is not the ideal filter) and close to measurement. The MMSE estimate of α and β are

$$\hat{\alpha} = \frac{\int_a^b \hat{x}(t)x(t)dt}{\int_a^b x^2(t)dt},$$
$$\hat{\beta} = \frac{\int_a^b \hat{x}(t)v(t)dt}{\int_a^b v^2(t)dt}.$$

and can be used for computing the performance of the denoising. As mentioned before, in the best signal recovery (ideal signal smoothing), the estimated signal, $\hat{x}(t)$, only shares the information about the desired signal, x, but not the noise, v. Consequently, in the ideal case, we should have $\beta = 0$ and $\alpha = 1$. Since the smoothing filter is not ideal, it attenuates some parts of signal and allows some parts of noise to pass, i.e., $\hat{\beta} \neq 0$ and $\hat{\alpha} \neq 1$. Therefore, the output signal-to-noise ratio is [63, 64]

$$SNR_{out} = \frac{P_{x_f}}{P_{v_f}} = \frac{\alpha^2 P_x}{\beta^2 P_v} = \frac{\alpha^2}{\beta^2} SNR_{in}. \tag{59}$$

We further define the SNR improvement as the difference $\mathrm{SNR_{dif}} = \mathrm{SNR_{out}} - \mathrm{SNR_{in}} = \frac{\hat{\alpha}^2 - \hat{\beta}^2}{\hat{\beta}^2}\mathrm{SNR_{in}}$. In Fig. 4, we report the results of the smoothing procedures on synthetic data. The results show that the smoothness priors designed by bilinear transform outperform that of designed by backward difference rule. The results also show that the signal reconstruction improves as n increases. We also compared the methods when

[2] In the ideal case, α is very close to 1 and β is very close to 0.

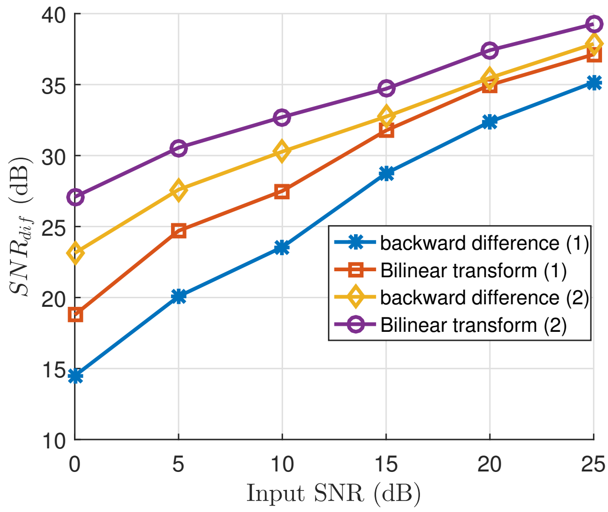


Figure 4: Mean values of SNR_{diff} for signal reconstruction by quadratic variation reduction using backward difference rule and bilinear transform, as a function of the input SNR (the signal and the noise are on disjoint frequency ranges).

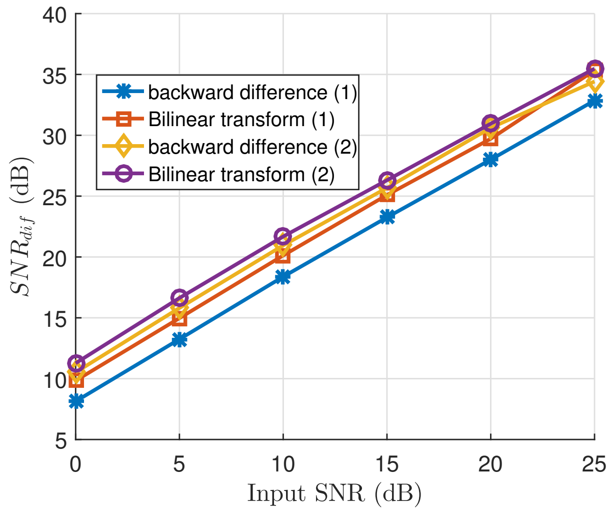


Figure 5: Mean values of SNR_{diff} for signal reconstruction by quadratic variation reduction using backward difference rule and bilinear transform, as a function of the input SNR (noise is white Gaussian noise).

the noise frequency components fall within the signal frequency components (i.e., white Gaussian noise). The results are shown in Fig. 5. Also, in this case the smoothness priors obtained with the filter designed by bilinear transform outperforms that of designed by backward difference rule.

6.2. *Real ECG Data*

The ECG signals used in this study were provided by PhysioNet [65]. The first dataset contains 80 records, originally provided for the PhysioNet/Computers in Cardiology Challenge 2004 [66]. Each record, extracted from a two-lead, sampling frequency 128 Hz Holter ECG recording, 1 min in length. The second dataset is the PTB Diagnostic ECG Database [67] (sampling frequency: 1000 Hz; resolution: 16-bit). 561 ECG segments, each with a duration of 5s, were selected from the first lead. The −6 dB cutoff frequency was swept linearly between 5 Hz to $f_s/2$ Hz, in two smoothness orders, 1 and 2. The cutoff frequency that minimizes the minimum mean square estimation was selected to compute the optimal value of trade-off parameter. As a preliminary example, the result of applying smoothness priors using both transforms (backward difference rule and bilinear transform) for a specific case (record t10m from PhysioNet/Computers in Cardiology Challenge 2004 [66]) is reported in Fig. 6. Figures 6(c)-6(f) show the results of applying both approaches for denoising the noisy ECG with $SNR = 5$ dB. The results of both approaches for denoising the noisy ECG with $SNR = 0$ dB are reported in Figures 6(h)-6(k). The denoised ECG using the smoothing filter designed by bilinear transform is closer to the ideal ECG, compared to the denoised ECG estimated by the backward difference rule.

The smoothness priors procedures were tested in presence of input noise with variable SNR ranging from 0 dB to 25 dB. The average and standard deviation of SNR_{diff} obtained after smoothing was used as a measure of performance and shown in Fig. 7(a). The results, reported in Fig. 7(a), show that the smoothness priors designed by bilinear transform improves the ECG reconstruction. For evaluating the performance of the proposed method, we also used two other measures of improvement. The first one, given by

$$\text{imp} = -10\log_{10}\frac{\sum_k (x_k - \hat{x}_k)^2}{\sum_k (y_k - x_k)^2}\ \text{(dB)}, \tag{60}$$

which considers the ratio between the power of the reconstruction error and the power of the noise in the original signal. The results shown in Fig. 7(b) confirmed the results obtained with SNR_{diff}. A similar result is conveyed in Fig. 7(c) by the third metrics

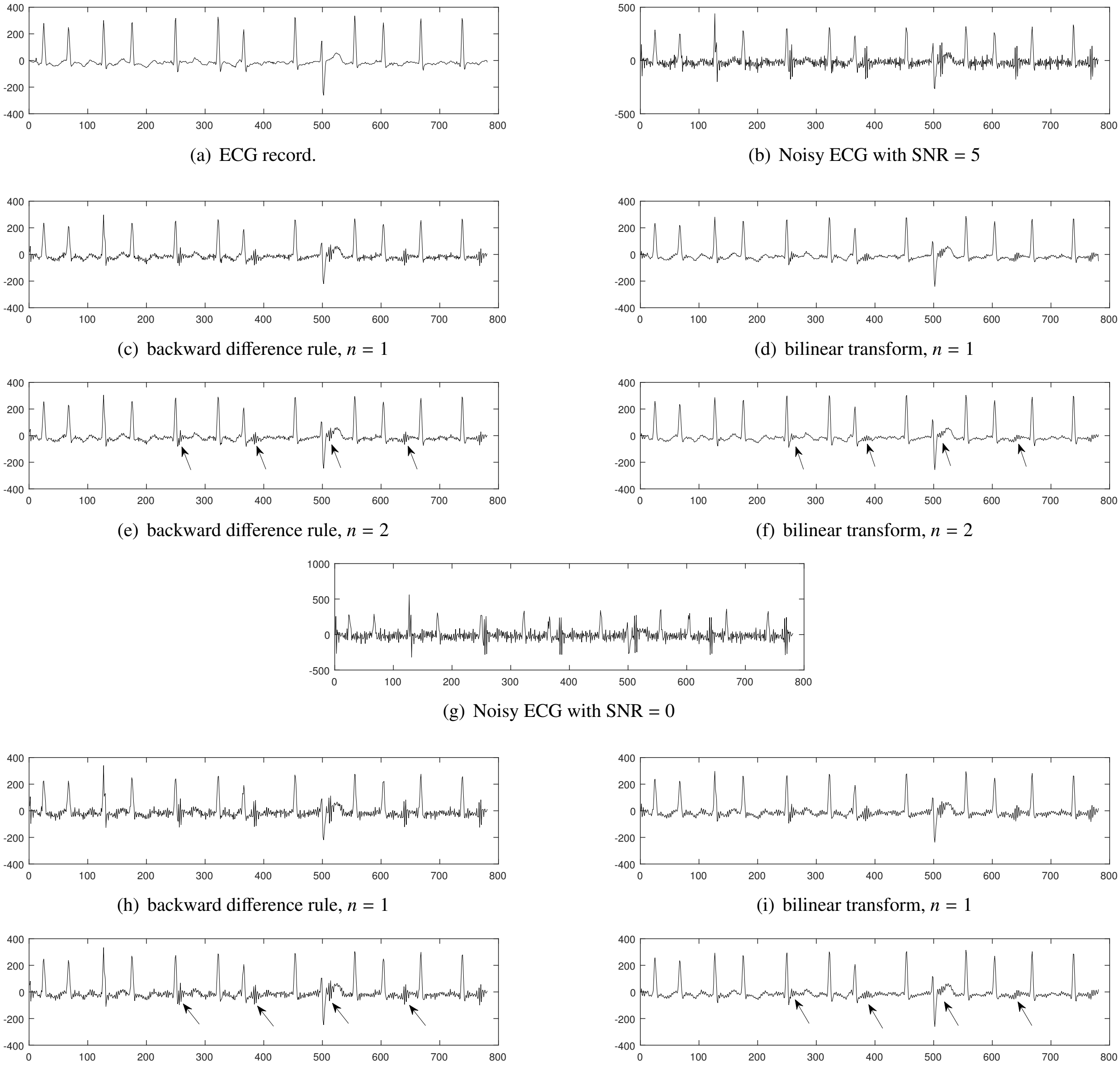
$$\text{NSR} = \sqrt{\frac{\sum_k (x_k - \hat{x}_k)^2}{\sum_k x_k^2}},$$

which is a classical ratio between the power of the reconstruction error and the power of the signal.

7. Sensitivity analysis and computational complexity

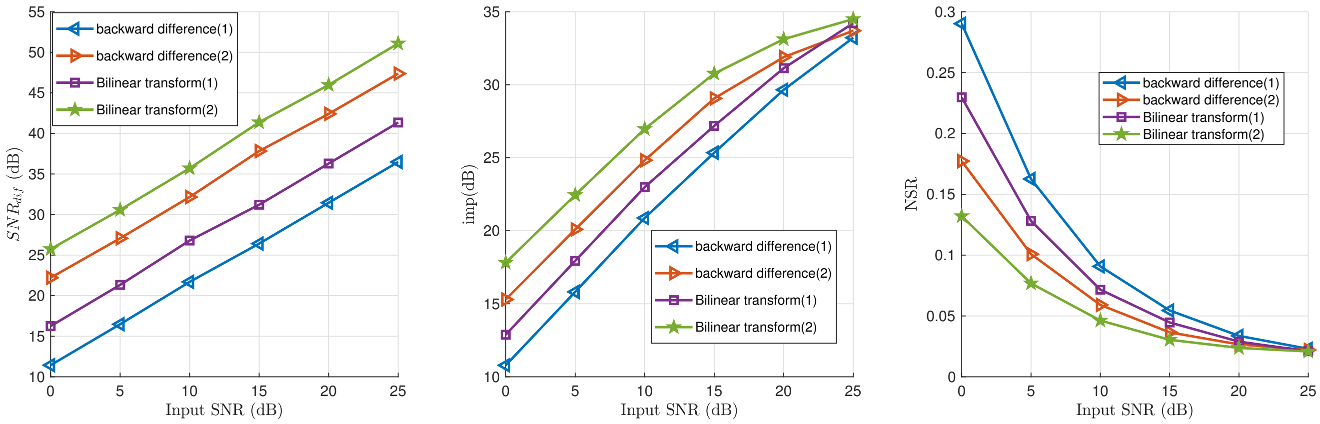
As noted above, the desired signal is estimated with (15) when backward difference rule is used to implement the smoothness priors or it can be estimated using (40) when bilinear transform is employed. We have shown that the value of λ determines the amount of frequency components that is passed by the smoothing filter. So its value depends on the class of signals to detrend. In this section, we analyze how robust the proposed algorithm is to variations of λ. A measure of robustness is the sensitivity of the cutoff frequency to variations of λ. Measuring sensitivity in terms of the derivative of f_c with respect to λ, the following result holds. According to (23) and (44), for large values of λ the cutoff frequency is computed as

$$\begin{aligned} f_c &= \tfrac{1}{\pi}\sin^{-1}\frac{1}{2\sqrt[2n]{\lambda}} \approx \frac{1}{2\pi\sqrt[2n]{\lambda_{BD}}} \quad \textit{or} \\ f_c &= \tfrac{1}{\pi}\tan^{-1}\frac{1}{\sqrt[2n]{\lambda}} \approx \frac{1}{\pi\sqrt[2n]{\lambda_{BT}}} \end{aligned} \tag{61}$$



(a) ECG record.
(b) Noisy ECG with SNR = 5
(c) backward difference rule, n = 1
(d) bilinear transform, n = 1
(e) backward difference rule, n = 2
(f) bilinear transform, n = 2
(g) Noisy ECG with SNR = 0
(h) backward difference rule, n = 1
(i) bilinear transform, n = 1
(j) backward difference rule, n = 2
(k) bilinear transform, n = 2

Figure 6: ECG smoothing provided by smoothness priors using backward difference rule and bilinear transform, for different values of n, for record t10m from PhysioNet/Computers in Cardiology Challenge 2004 (SNR = 5, SNR = 0).



(a) Mean $\mathrm{SNR_{diff}}$ (b) Mean imp (c) Mean NSR

Figure 7: Mean values of SNR_{diff}, NSR and imp for ECG reconstruction by quadratic variation reduction using backward difference rule and bilinear transform, as a function of the input SNR (noise is broadband).

The first equation is for backward difference rule and the second one is for bilinear transform, hence the index added on λ. The sensitivity of cutoff frequency with respect to variations of λ satisfies

$$\frac{\partial}{\partial \lambda} f_c = O\left(\frac{1}{\lambda^{\frac{2n+1}{2n}}}\right) \tag{62}$$

where $O(\cdot)$ denotes the Landau symbol.

Regarding the computational complexity of the methods, both solutions (15) and (40) involve matrix inversion, which have complexity $O(n^3)$. However, since the matrix $\Gamma^T\Gamma + \lambda D^T D$ or $I + \lambda D^T D$ are symmetric, positive-definite, tridiagonal, the system (15) and (40) can be solved efficiently with complexity $O(n)$ [68].

8. Conclusion

In this paper, we proposed a closed-form expression for the smoothness trade-off parameter. The closed-form expression resulting from a frequency domain interpretation of the smoothing procedure. We have shown that the smoothness trade-off is related to the amount of frequency components that the method allows to pass. We introduced a new way to design and implement the smoothness priors, Hodrick-Prescott filter, and quadratic variation. The bilinear transformation method was proposed to design the smoothness priors. Experiments on both synthetic and real world signals with different levels of noise demonstrated that the proposed technique (bilinear transform) is indeed more effective in smoothness priors design when compared to the traditional ones (backward difference rule).

References

[1] G. Kitagawa and W. Gersch, *Smoothness Priors Analysis of Time Series*. New York: Springer, 1996.

[2] W. Gersch, "Smoothness priors," in *New Directions in Time Series Analysis*, D. Brillinger, P. Caines, J. Geweke, E. Parzen, M. Rosenblatt, and M. S. Taqqu, Eds. New York, NY: Springer New York, 1993, pp. 113–146.

[3] 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.

[4] 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.

[5] 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.

[6] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via l0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 174:1–174:12, Dec. 2011.

[7] 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.

[8] W. Gersch and G. Kitagawa, "Smoothness priors transfer function estimation," *Automatica*, vol. 25, no. 4, pp. 603 – 608, 1989.

[9] 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.

[10] 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.

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

[12] Y. J. Wang, J. Li, H. Xu, H. O. Yan, J. Yuan, J. H. Li, X. H. Liu, Q. Zhou, and N. Li, "Smoothness prior approach to removing nonlinear trends from signals in identification of low frequency oscillation mode," in *Renewable Energy and Power Technology II*, ser. Applied Mechanics and Materials, vol. 672. Trans Tech Publications, 11 2014, pp. 1070–1074.

[13] M. Niedzwiecki and W. A. Sethares, "Smoothing of discontinuous signals: the competitive approach," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 1–13, 1995.

[14] 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.

[15] P. H. C. Eilers, "Parametric time warping." *Analytical chemistry*, vol. 76 2, pp. 404–11, 2004.

[16] P. Green and B. Silverman, *Nonparametric regression and generalized linear models: a roughness penalty approach*. United Kingdom: Chapman and Hall, 1994.

[17] T. Hastie and R. Tibshirani, *Generalized additive models*. Wiley Online Library, 1990.

[18] R. J. Shiller, "Smoothness priors and nonlinear regression," *Journal of the

American Statistical Association, vol. 79, no. 387, pp. 609–615, 1984.
[19] G. Wahba, *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics, 1990.
[20] Y. R. Yue, P. L. Speckman, and D. Sun, "Priors for bayesian adaptive spline smoothing," *Ann Inst Stat Math*, vol. 64, pp. 577–613, 2012.
[21] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with b-splines and penalties," *Statistical Science*, vol. 11, pp. 89 – 121, 1996.
[22] B. W. Silverman, "Some aspects of the spline smoothing approach to non-parametric regression curve fitting," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 47, no. 1, pp. 1–52, 1985.
[23] R. L. Eubank, *Spline Smoothing and Nonparametric Regression*. Dekker, New York, 1988.
[24] A. N. Tikhonov, "On the solution of ill-posed problems and the method of regularization," *Dokl. Akad. Nauk SSSR*, vol. 11, pp. 501 – 504, 1963.
[25] G. H. Golub, P. C. Hansen, and D. P. O'Leary, "Tikhonov regularization and total least squares," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, pp. 185 – 194, 1999.
[26] F. Zhang, S. Chen, H. Zhang, X. Zhang, and G. Li, "Bioelectric signal detrending using smoothness prior approach," *Medical engineering and physics*, vol. 36, 05 2014.
[27] R. J. Shiller, "A distributed lag estimator derived from smoothness priors," *Econometrica*, vol. 41, no. 4, pp. 775–788, 1973.
[28] E. T. Whittaker, "On a new method of graduation," *Proc. Edinburgh Math. Soc*, vol. 41, pp. 63–75, 1923.
[29] A. Aitken, "On the theory of graduation," *Proceedings of the Royal Society of Edinburgh*, vol. 47, pp. 36–45, 1926.
[30] R. Henderson, "A new method of graduation," *Transactions of the Actuarial Society of America*, vol. 25, pp. 29–40, 1924.
[31] ——, "Further remarks on graduation," *Transactions of the Actuarial Society of America*, vol. 26, pp. 52–57, 1924.
[32] ——, "Mathematical theory of graduation," *Actuarial Studies 4, Actuarial Society of America, New York*, 1938.
[33] A. Aitken, "The smoothing of data," *DSc dissertation, University of Edinburgh*, 1925.
[34] R. Hodrick and E. Prescott, "Postwar u.s. business cycles: An empirical investigation," *Journal of Money, Credit and Banking*, vol. 29, no. 1, pp. 1–16, 1997.
[35] F. E. Kydland and E. C. Prescott, "Time to build and aggregate fluctuations," *Econometrica*, vol. 50, no. 6, pp. 1345–1370, 1982.
[36] D. K. Backus and P. J. Kehoe, "International evidence on the historical properties of business cycles," *The American Economic Review*, vol. 82, no. 4, pp. 864–888, 1992.
[37] J. P. Danthine and J. B. Donaldson, "Methodological and empirical issues in real business cycle theory," *European Economic Review*, vol. 37, no. 1, pp. 1 – 35, 1993.
[38] R. Fiorito and T. Kollintzas, "Stylized facts of business cycles in the g7 from a real business cycles perspective," *European Economic Review*, vol. 38, no. 2, pp. 235 – 269, 1994.
[39] F. E. Kydland and E. C. Prescott, "Business cycles: real facts and a monetary myth," *Quarterly Review*, no. Spr, pp. 3–18, 1990.
[40] T. F. Cooley and G. D. Hansen, "The inflation tax in a real business cycle model," *The American Economic Review*, vol. 79, no. 4, pp. 733–748, 1989.
[41] S.-J. Kim, K. Koh, S. P. Boyd, and D. M. Gorinevsky, "l_1 trend filtering," *SIAM Review*, vol. 51, pp. 339–360, 2009.
[42] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, 1970.
[43] R. Tibshirani, "Regression shrinkage and selection via the lasso," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 58, pp. 267–288, 1994.
[44] S. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer Science Business Media,Inc., 2004.
[45] L. C. G. Rogers and D. Williams, *Diffusions, Markov processes, and martingales. Vol. 1*, ser. Cambridge Mathematical Library. Cambridge: Cambridge University Press, 2000, foundations, Reprint of the second (1994) edition.
[46] A. Fasano and V. Villani, "Baselinewander removal for bioelectrical signals by quadratic variation reduction," *Signal Processing*, vol. 99, pp. 48–57, 2014.
[47] R. Sameni, "Online filtering using piecewise smoothness priors: Application to normal and abnormal electrocardiogram denoising," *Signal Processing*, vol. 133, pp. 52 – 63, 2017.
[48] 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.
[49] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.
[50] H. Akaike, "Likelihood and the bayes procedure," *Trabajos de Estadistica Y de Investigacion Operativa*, vol. 31, no. 1, pp. 143–166, 1980.
[51] ——, "Factor analysis and AIC," *Psychometrika*, vol. 52, no. 3, pp. 317–332, Sep 1987.
[52] G. Wahba, "A first survey of smoothing problems via the method of generalized cross-validation," *In Application of Statistics, P. R. Krishnaiah, ed. North-Holland, Amsterdam, The Netherlands*, vol. 40, p. 507–524, 1978.
[53] ——, "The relationship between spline smoothing and stochastic processes," *J. R. Statist. Soc. Ser. B*, vol. 40, p. 364–372, 1978.
[54] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge university press, 2004.
[55] P. C. Hansen, *The L-curve and its use in the numerical treatment of inverse problems*. IMM, Department of Mathematical Modelling, Technical University of Denmark, 1999.
[56] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, "Tikhonov regularization and the lcurve for large discrete ill-posed problems," *J. Comput. Appl. Math.*, vol. 3, no. 23, pp. 6160–6173, 2000.
[57] A. Tustin, "A method of analysing the behaviour of linear systems in terms of time series," *Journal of the Institution of Electrical Engineers - Part IIA: Automatic Regulators and Servo Mechanisms*, vol. 94, no. 1, pp. 130–142, May 1947.
[58] T. N. T. Greville, "On smoothing a finite table: A matrix approach," *SIAM Journal of Applied Math.*, vol. 50, pp. 137–154, 1957.
[59] A. Oppenheim and R. Schafer, *Discrete-Time Signal Processing*. Pearson Education, 2011.
[60] S. A. Dyer and J. S. Dyer, "The bilinear transformation," *IEEE Instrumentation Measurement Magazine*, vol. 3, no. 1, pp. 30–34, 2000.
[61] M. Baxter and R. G. King, "Measuring business cycles: Approximate band-pass filters for economic time series," *The Review of Economics and Statistics*, vol. 81, no. 4, pp. 575–593, 1999.
[62] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, 1997, available at www.dspguide.com.
[63] A. Kheirati Roonizi, "A New Approach to ARMAX Signals Smoothing: Application to Variable-Q ARMA Filter Design," *IEEE Transactions on Signal Processing*, vol. 67, no. 17, pp. 4535–4544, Sep. 2019.
[64] 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, Feb 2016.
[65] "The physionet challenge," 2004. [Online]. Available: http://www.physionet.org/challenge/2004/
[66] G. E. Moody, "Spontaneous termination of atrial fibrillation: a challenge from physionet and computers in cardiology 2004," in *Computers in Cardiology, 2004*, 2004, pp. 101–104.
[67] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. e215–e220, 2000.
[68] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.