On the Selection of Optimum Savitzky-Golay Filters

Sunder Ram Krishnan and Chandra Sekhar Seelamantula, *Senior Member, IEEE*

***Abstract*—Savitzky-Golay (S-G) filters are finite impulse response lowpass filters obtained while smoothing data using a local least-squares (LS) polynomial approximation. Savitzky and Golay proved in their hallmark paper that local LS fitting of polynomials and their evaluation at the mid-point of the approximation interval is equivalent to filtering with a *fixed* impulse response. The problem that we address here is, "how to choose a pointwise minimum mean squared error (MMSE) S-G filter length or order for smoothing, while preserving the temporal structure of a time-varying signal." We solve the bias-variance tradeoff involved in the MMSE optimization using Stein's unbiased risk estimator (SURE). We observe that the 3-dB cutoff frequency of the SURE-optimal S-G filter is higher where the signal varies fast locally, and vice versa, essentially enabling us to suitably trade off the bias and variance, thereby resulting in near-MMSE performance. At low signal-to-noise ratios (SNRs), it is seen that the adaptive filter length algorithm performance improves by incorporating a regularization term in the SURE objective function. We consider the algorithm performance on real-world electrocardiogram (ECG) signals. The results exhibit considerable SNR improvement. Noise performance analysis shows that the proposed algorithms are comparable, and in some cases, better than some standard denoising techniques available in the literature.**

***Index Terms*—Bias, local polynomial regression, MMSE, Savitzky-Golay filters, Stein's lemma, SURE, variance.**

I. Introduction

REAL-WORLD signals such as speech, electrocardiogram (ECG) and geophysical signals are time-varying in one or more properties. For signals such as speech, approximate generation models have been developed [1], whereas some signals are not amenable to such modeling. In the present work, we are interested in a local signal model than one based on the signal production. This would help us to apply the technique to a wider class of natural signals. In particular, we will use the well-known idea of local polynomial regression (LPR), which offers a method of data smoothing based on a LS criterion. The signal model under consideration is:

$$x[nT] = s[nT] + w[nT]; \quad n = 1, 2, 3, \ldots, N, \qquad (1)$$

where T denotes the sampling period, and N denotes the total number of observations. For simplicity, throughout the paper we will denote the nth sample of a signal a by a_n; that is, $a_n \stackrel{\triangle}{=} a[nT]$. Now, s_n is the nth sample of a deterministic and time-varying signal $s(t)$, and w_n is a sample of additive white Gaussian noise of mean zero and variance σ^2. Note that the index n here need not necessarily refer to time. The signal s is assumed to be sufficiently smooth so as to admit a local polynomial approximation, whereas the exact functional form is unknown. We proceed to fit a polynomial of fixed order to the observations locally, so that the regression is optimal in the LS sense. Once the polynomial coefficients are computed, one can obtain the smoothed output value by evaluating the polynomial at the central index of the approximation window. The fitting and evaluation process is then repeated at each reconstruction instant. This is the essence of LPR. For a detailed review of the technique, we refer the reader to [2], [3]. Although it might appear at first look, that such a local fitting and evaluation process cannot be interpreted as shift-invariant filtering, Savitzky and Golay had shown in their original work [4] that the procedure is equivalent to discrete-time convolution with a *fixed* impulse response—the resulting filters being referred to as S-G filters. LPR is of widespread use in domains such as kernel density estimation, spectral estimation, non-linear time-series modeling. Recently, it has been put to use in some applications in biomedical signal processing [5]. In particular, S-G filters were used for ECG signal processing by Hargittai [6]. The local modeling using polynomials uses only a few parameters, at the same time providing a smooth representation. We would like the resulting model to have two important properties, the first one being adaptability to signal variations, and the second one being robustness to noise. The second requirement stems from the fact that real-world measurements suffer from noise.

In all the applications using LPR, the important question to be answered is, "What is the optimal data length to be selected for the regression?". This is the classical *bandwidth selection* problem in the realm of statistics. We intend to solve this problem by optimizing the SURE objective, which is an unbiased estimator of the mean squared error (MSE). Note that we have a pointwise MMSE objective here. Stein, in his seminal work [7], derived an unbiased estimator for the average squared error loss function. His derivation relies on an identity concerning expectations of functions of a Gaussian random variable. The elegance of SURE comes about in the fact that an estimate of the MSE can be computed reliably based on the noisy observations without requiring a prior on the underlying signal [8]. SURE has been used by signal processing engineers in denoising applications for determination of thresholds and regularization parameters minimizing an estimate of the MSE, both in the domains of image and speech processing [9]–[16]. A technique for determining the smoothing parameter value

Manuscript received February 09, 2012; revised June 10, 2012 and September 18, 2012; accepted October 01, 2012. Date of publication October 16, 2012; date of current version December 21, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Petr Tichavsky. This work is supported by the Government of India, Department of Science and Technology-Intensive Research in High Priority Areas (DST-IRHPA) grant (DSTO-943).

The authors are with the Department of Electrical Engineering, Indian Institute of Science, Bangalore-560012, Karnataka, India (e-mail: sunder@ee.iisc.ernet.in; chandra.sekhar@ieee.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2012.2225055

for smoothing spline regression using SURE is discussed in [17]. We show that the LPR *bandwidth selection* problem can be solved elegantly using SURE. We also show that addition of a regularization term in the SURE cost [16] improves the performance of our technique.

The motivation for the present work lies in a recent article by Schafer [18]. We noted that it is possible to interpret the bandwidth selection problem as one of selecting an optimum S-G filter, which is easily designed. A simple design technique is also presented in [18]. The useful observation to be made here is that data smoothing based on a local LS polynomial regression is equivalent to a discrete convolution with a fixed impulse response, as long as we maintain the condition that data samples are symmetric about the origin. The derivation of this result is presented again briefly in this paper in a later section in order to make the exposition self-contained. The filtering equivalence enables ready deployment of fast filtering algorithms. Schafer has given a nice frequency-domain interpretation of the S-G filters giving an empirical relationship of the 3-dB cutoff frequency in terms of the polynomial order and the impulse response half-width, which shows that they are inversely related. Also, it is worth noting that Savitzky and Golay have shown experimentally that LS smoothing guarantees both the requirements we had of the model—those of reducing noise, and maintaining the shape and height of the waveform peaks. Consideration of S-G filters shows that we are indeed addressing the issue of "bandwidth selection"; that is, "What is the 3-dB cutoff frequency of the optimal S-G filter at each instant?". We would like to emphasize here that though, having a fixed window length (implying a fixed S-G filter) is reasonable in the absence of noise, the noisy case requires a solution to a tradeoff between capturing the fine signal variations and noise-smoothing (implying a different S-G filter at each instant). This is essentially the *bias-variance tradeoff* [19] put in simple terms. We show that SURE helps us in trading off bias and variance, thus helping us in choosing the optimal-length S-G filter at each time instant. Once we obtain the "correct" length of the filter to be used at each instant, we can readily compute the smoothed output sequence by a filtering operation. After addressing the bandwidth selection issue for a fixed order in some detail, we consider the other equally interesting problem in LPR—one of choosing the appropriate polynomial order, for a given bandwidth. We seek the MMSE solution in this case, and demonstrate experimentally that a SURE solution to this problem is feasible, with the problem formulation and approach being identical to the bandwidth selection problem. Thus, in this case, we fix a filter length and seek the MMSE filter order (which fixes the optimum cutoff frequency) to be used at each instant. We shall refer to this as the adaptive filter order selection algorithm, with the former being the adaptive filter length/bandwidth selection algorithm.

The organization of the paper is as follows. A brief overview of S-G filters is given in Section II-A followed by basic concepts of SURE in Section II-B. The bias-variance tradeoff problem encountered on optimizing the MSE is explained in Section III-A, and we detail the SURE solution to the problem subsequently in Section III-B. A discussion on the choice of order of the regression is given in Section III-C. A regularized version of the SURE cost is introduced in Section III-D. Simulation results on both synthesized and real data are presented in Section IV. We draw the concluding remarks in Section V.

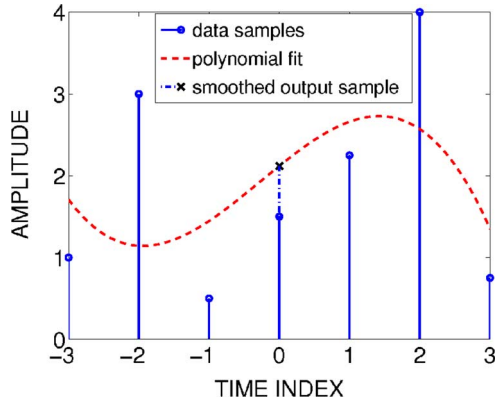

Fig. 1. (Color in electronic version) The key idea of local LS polynomial regression.

II. Background Theory and Problem Formulation

A. A Brief Overview of S-G Filters

Consider a symmetric window about the point of reconstruction, which is presently taken to be $n_0T = 0$. Our objective is to fit a polynomial $q_n = \sum_{k=0}^{p} a_k(nT)^k$ in a LS sense to the data samples, that is, we seek to minimize the cost function:

$$\mathcal{C}_p = \sum_{n=-M}^{M} (q_n - x_n)^2 = \sum_{n=-M}^{M} \left(\sum_{k=0}^{p} a_k(nT)^k - x_n \right)^2 .$$

The basic idea of LS local polynomial regression is shown in Fig. 1, wherein we have locally fitted a third degree polynomial to a set of 7 samples centered at time index 0. The smoothed output value at the central index is also indicated. The analysis that follows holds true for any approximation interval "half-length," MT. Once we have fitted a polynomial, we can proceed to compute the smoothed output value at the central index 0. The result is the zeroth polynomial coefficient; that is, $y_0 = q_0 = a_0$. When computing the output sequence at any other instant n_0T, we will shift the samples so that once again, the reconstruction is to be performed at the instant zero. This operation is called *data centering*.

Presently, we briefly outline the proof given in [18] showing that LS polynomial smoothing for different shifts of the $2M + 1$-sample interval is equivalent to filtering with a *fixed* impulse response. Towards achieving this goal, we must first obtain the optimal polynomial coefficients by differentiating $\mathcal{C}_p$ with respect to the unknown coefficients and setting the derivatives equal to zero. We obtain the following set of $p+1$ *normal* equations in $p + 1$ unknowns:

$$\sum_{k=0}^{p} \left(\sum_{n=-M}^{M} (nT)^{j+k} a_k \right) = \sum_{n=-M}^{M} (nT)^j x_n, \qquad j = 0, 1, 2, \ldots, p.$$

That we are now closer towards achieving our goal is made clear if we write the above set of equations in matrix form. We define a $(2M+1) \times (p+1)$ matrix $\mathbf{A}$ whose (n, j)th element is

given as $u_{n,j} = (nT)^j; -M \leq n \leq M, j = 0, 1, 2, \ldots, p$. The next step is to take a look at the $(p+1) \times (p+1)$ symmetric matrix $\mathbf{B} = \mathbf{A}^T\mathbf{A}$, which has its individual elements $v_{j,k} = \sum_{n=-M}^{M} (nT)^{j+k}$; $j, k = 0, 1, 2, \ldots, p$. Defining the vectors $\mathbf{x}$ and $\mathbf{a}$ as $\mathbf{x} = [x_{-M}, \ldots, x_{-1}, x_0, x_1, \ldots, x_M]^T$, and $\mathbf{a} = [a_0, a_1, \ldots, a_p]^T$, we arrive at the desired matrix form of the normal equations $\mathbf{Ba} = \mathbf{A}^T\mathbf{Aa} = \mathbf{A}^T\mathbf{x}$. From the preceding calculations, we obtain the coefficient vector as $\mathbf{a} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{x} = \mathbf{Hx}$. Recall that now we need to only compute a_0, which is the smoothed output value. Consequently, we need only find the zeroth row of the $\mathbf{H}$ matrix, which gives the desired a_0 as a linear combination of the observations. Further, the $\mathbf{H}$ matrix is independent of the input samples (it depends only on M and p) and thus, the same weighting coefficients are obtained at each approximation interval covering $2M + 1$ samples. Considering the convolution equation $y_n = \sum_{m=-M}^{M} h_{n-m}x_m$, the output sample at the index 0 is given as: $y_0 = \sum_{m=-M}^{M} h_{-m}x_m = a_0$. We observe that the elements of the zeroth row of the $\mathbf{H}$ matrix are the flipped impulse response coefficients of the S-G filter. The first row alone is of interest, and could be computed using efficient matrix inversion techniques.

In the presence of noise, it becomes important to adapt the filter length at each reconstruction point to minimize the pointwise MSE. Next, we present some introductory material on the SURE formalism in a manner relevant to our problem.

B. SURE Theory

In our model as given in (1), we seek to obtain an MMSE estimate of s_i based upon observing the noisy x_i within the observation window, that is, we would like to minimize the following risk:

$$\mathcal{R} = \mathcal{E}\left\{\frac{1}{N_0}\sum_{i=1}^{N_0}(f_i(\mathbf{x}) - s_i)^2\right\}, \tag{2}$$

where $\mathcal{E}$ denotes the expectation operator, $f_i(\mathbf{x})$ denotes our estimate of s_i based on the observation $\mathbf{x}$. In the expansion of $\mathcal{R}$, we will obtain terms containing the product of the unknown s_i and $f_i(\mathbf{x})$ thus rendering the direct minimization infeasible. This problem is circumvented using an unbiased estimator of the risk $\mathcal{R}$. First, we present a result that is useful in deriving the unbiased estimator, the proof of which can be found in [7]. A detailed presentation and proof that is more accessible to the signal processing audience is given in [13] and [16].

Lemma 1: (*Stein* [7], 1981) Let $\mathbf{x}$ be a $\mathcal{N}(\mathbf{s}, \sigma^2\mathbf{I})$ real random vector and let $f_i : \mathbb{R}^{N_0} \to \mathbb{R}$ be weakly differentiable. Suppose also that $\mathcal{E}\{|\frac{\partial f_i(\mathbf{x})}{\partial x_i}|\} < \infty$. Then,

$$\mathcal{E}\{s_i f_i(\mathbf{x})\} = \mathcal{E}\{x_i f_i(\mathbf{x})\} - \sigma^2\mathcal{E}\left\{\frac{\partial f_i(\mathbf{x})}{\partial x_i}\right\}.$$

Let us presently see how it assists us in deriving the random variable that is an unbiased estimator of the risk in (2). In the derivation we need consider only one term in the summation given in (2), as expectation is a linear operator. We seek to minimize the cost:

$$\begin{aligned}\mathcal{R}_i &\triangleq \mathcal{E}\left\{(s_i - f_i(\mathbf{x}))^2\right\} \\ &= \mathcal{E}\left\{f_i(\mathbf{x})^2 + s_i^2 - 2f_i(\mathbf{x})s_i\right\}. \end{aligned} \tag{3}$$

Using Lemma 1, we have that (3) reduces to

$$\mathcal{R}_i = \mathcal{E}\left\{f_i(\mathbf{x})^2 - 2f_i(\mathbf{x})x_i + 2\sigma^2\frac{\partial f_i(\mathbf{x})}{\partial x_i}\right\} + s_i^2.$$

Hence, it can be concluded that

$$\epsilon_i = \underbrace{f_i(\mathbf{x})^2 - 2f_i(\mathbf{x})x_i + 2\sigma^2\frac{\partial f_i(\mathbf{x})}{\partial x_i}}_{\tilde{\epsilon}_i} + s_i^2,$$

is an unbiased estimator for the squared error risk defined in (3). This is referred to as SURE of $\mathcal{R}_i$. Although ϵ_i contains the signal term s_i^2, it does not affect the process of minimization since we are minimizing with respect to f_i. Therefore, it is sufficient to minimize $\tilde{\epsilon}_i$. The unbiased estimator for $\mathcal{R}$ is now easily seen to be:

$$\begin{aligned}\epsilon = &\underbrace{\frac{1}{N_0}\sum_{i=1}^{N_0} f_i(\mathbf{x})^2 - \frac{2}{N_0}\sum_{i=1}^{N_0} f_i(\mathbf{x})x_i + \frac{2\sigma^2}{N_0}\sum_{i=1}^{N_0}\frac{\partial f_i(\mathbf{x})}{\partial x_i}}_{\tilde{\epsilon}} \\ &+ \frac{1}{N_0}\sum_{i=1}^{N_0} s_i^2, \end{aligned} \tag{4}$$

where again, we need to only minimize $\tilde{\epsilon}$. Note that though we have assumed white noise in our model in (1), the application of SURE formalism does not require this assumption.

III. SURE Approach to the MMSE Problem

In this section, we first consider the bias-variance tradeoff in MMSE optimization, and then detail the SURE approach to solving this problem.

A. Bias-Variance Tradeoff

In the signal model (1), we make an assumption that $s(t)$ is smooth enough to permit a local polynomial approximation of a particular order p in the interval $[n_0T - MT, n_0T + MT]$. The assumption is justified by the Weierstrass's theorem [20], which states that a function, continuous in a finite closed interval can be approximated to any desired accuracy using polynomials. Thus, we can express the signal as $s_n = \sum_{k=0}^{p} a_k(nT)^k + \mathcal{O}\{(nT)^{p+1}\}$, for $nT \in [n_0T - MT, n_0T + MT]$, and where $\mathcal{O}$ denotes the Landau symbol. For convenience, we denote the length of the approximation interval as L. We note that in our problem, L is the number of symmetric impulse response samples of the corresponding S-G filter (except for the central sample), that is, $L = 2M$. The coefficients of the polynomial are obtained by employing a LS-fit to the observations over the approximation interval, and the values so obtained suffer from

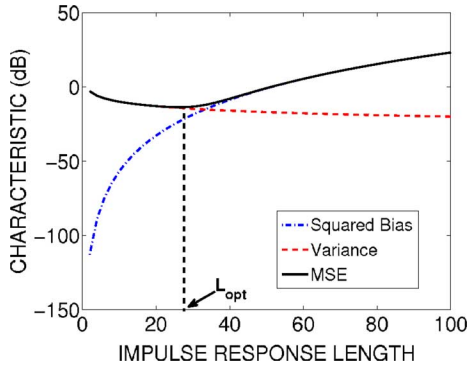

Fig. 2. (Color in electronic version) Illustration of bias-variance tradeoff as a function of impulse response length.

estimation errors as the observations are noisy. In order to compute the smoothed output sample, we evaluate the polynomial at the central instant n_0T. Thus, we have obtained $\hat{s}_{n_0}$, which is our estimate of s_{n_0}.

Using a rectangular window and considering the asymptotic case wherein the number of samples in the observation window tends to infinity (or equivalently as $T \to 0$), bias and variance expressions for the estimator have been given in [2], which we reproduce here for the reader's convenience.

$$\text{Bias}\{\hat{s}(t)\} \approx \begin{cases} \dfrac{s^{(p+1)}(t)}{(p+2)!}\left(\dfrac{L}{2}\right)^{p+1}, & \text{if } p \text{ is odd,} \\ \dfrac{s^{(p+2)}(t)}{(p+3)!}\left(\dfrac{L}{2}\right)^{p+2}, & \text{if } p \text{ is even.} \end{cases}$$

$$\text{Var}\{\hat{s}(t)\} = \frac{\sigma^2}{L}.$$

The notation $s^{(q)}(t)$ stands for the qth derivative of s with respect to t. Only the dominant term has been retained in the previous equation. The tradeoff between bias and variance versus L is evident from these expressions and is graphically represented in Fig. 2. As L increases, the bias increases, whereas variance decreases, and vice versa. This is intuitive because, if the filter length is increased, we will not be able to capture the finer variations of the signal (implying greater bias), whereas the noise becomes better smoothed out (implying lesser variance). In the figure, we show the squared bias, variance, and MSE of the LPR estimator obtained by using a third-order local LS polynomial fit, versus the impulse response length. The MSE attains the minimum at $L = L_{\text{opt}}$. Bias, variance, and MSE are characteristics of an estimator and hence the y-axis is commonly labelled as *characteristic*. Schafer has shown empirically that the nominal normalized cutoff frequency is inversely proportional to the impulse response half-length, whereas, it varies directly (approximately) as the polynomial order for a fixed impulse response length. From the preceding discussion, we would want our algorithm to select an S-G filter having a shorter impulse response length (translating to a higher cutoff frequency) in those regions where, locally, the signal varies fast and a higher impulse response length (translating to a lower cutoff frequency) S-G filter in those regions where the local signal variations are relatively flat.

B. Selecting the Optimal-Length S-G Filter Using SURE

Now, we are in a position to formulate the optimal filter selection problem. Consider reconstruction at an instant n_0T (once we center the data, reconstruction need be done only at 0). We define a set $B_{n_0,L}$ as those samples that fall in the window centered at n_0T. Formally,

$$B_{n_0,L} = \left\{ x_n : nT \in \left[n_0T - \frac{LT}{2}, n_0T + \frac{LT}{2} \right] \right\}. \quad (5)$$

We can now formulate the cost function in (4) using $N_0 = |B_{n_0,L}|$ (indexing the samples within each window from left to right with $i = 1, 2, 3, \ldots, |B_{n_0,L}|$), where $|\cdot|$ denotes the cardinality of a set and using a form for $f_i(\mathbf{x}); x_n \in B_{n_0,L}$, as follows: $f_i(\mathbf{x}) = \sum_{k=0}^{p} a_{k,n_0}(\mathbf{x})(nT)^k$. This amounts to employing a LS-fit of a polynomial locally and evaluating the same at the sampling instants within the observation interval and this, as mentioned previously, is S-G filtering.

The steps given by the Algorithm 1 are repeated at each reconstruction instant, n_0T.

Algorithm 1: To Calculate L_{opt} at an Instant n_0T

Require: $L_{\min} = (p+1)$
Ensure: Optimality with respect to SURE objective $\tilde{\epsilon}$
  $L \leftarrow L_{\min}$
  if $n_0 \neq 0$ **then**
    $x_n \leftarrow x_{n+n_0}$
  end if
  while $L \leq L_{\max}$ **do**
    Employ LS-fit over $[-\frac{LT}{2}, \frac{LT}{2}]$
    Evaluate $\tilde{\epsilon}$ with this L
    $L \leftarrow L + 2$
  end while
  $L_{\text{opt}} \leftarrow \arg\min \tilde{\epsilon}(L)$

Note that we need a minimum of $p+1$ samples as we have a polynomial of order p. If the number of samples is less than $p+1$, we have an ill-conditioned system of equations in solving the LS problem whereas, the number being p results in an exact fit using the polynomial and no smoothing happens. We choose to increment the window length by 2 samples (minimum number that can be symmetrically taken about the point of reconstruction) in each iteration, as it enables a finer arithmetic search within the space of window lengths, and consequently better performance than a coarse geometric search. Within a particular $L_{\min}$ and $L_{\max}$, the selection space of geometrically progressing window lengths is fully contained within the arithmetic set we propose to use. A geometric progression of filter lengths would lead to a faster and computationally more efficient algorithm than the one used here, but it may be associated with a suboptimal performance. Thus, there is a tradeoff between performance and computational complexity in the filter length selection. Also, we can fix a value for $L_{\max}$ depending upon the amount of smoothing required and the fineness of the signal

features to be captured. An important consideration here is that we should constrain our search space to reduce the complexity of the algorithm. We experimentally observed that a small and reasonable window length space was sufficient for satisfactory performance of our method. A maximum window length of 64 gave satisfactory results in all our experiments. For a fixed-order polynomial regression throughout the observation window, we fixed the order to be $p = 3$ (therefore $L_{\min} = 4$). Indeed, if one selects a higher polynomial order, $L_{\min}$ would have to be increased; otherwise the LS fit would not be unique, and the result would not be accurate. The basis for selecting the global order of regression to be three is explained below. The aspect of locally adapting the order is discussed in Section IV-B2.

C. Order Selection

We performed an experiment with ECG signal 1 (see Section IV-B) to observe how the MSE varies with polynomial order. We simulated a noisy signal with white Gaussian noise samples at two different SNRs. The output using the SURE-based bandwidth selection algorithm is obtained and averaged over 100 Monte Carlo trials to estimate the MSE. The variation of MSE versus polynomial order is shown in Fig. 3. A similar trend was observed for other SNR values as well. We observe that the lowest MSE occurs for $p = 3$. Considering the asymptotic case, Fan and Gijbels proved that for a given bandwidth, the variability in the estimate increases with increase in the polynomial order, even though a reduction in bias is to be expected. However, the variance of the estimator would increase only when passing from an odd p to the next higher even order regression. There is no difference in the variance of the estimator while moving from an even p to the next odd order value, asymptotically. Therefore, in the noisy case, even-order regression is not recommended, as one can potentially achieve bias reduction with the next higher odd-order polynomial regression with there being no increase in variability. Again, since the bandwidth is locally adapted, the lowest odd-order regressions ($p = 1, 3$) need only be considered. This is to take care of overfitting, increased variability, and increased computational cost resulting from fitting higher odd-order polynomials. However, near the signal peaks and valleys, local cubic fits are preferable to constant fits. Cubic polynomial fits are also known to result in minimum curvature in case of signal interpolation [21]. We would also like to mention at this point that a data-driven algorithm, which simultaneously optimizes the order and the bandwidth is complicated to realize because of the inherent correlation between the two critical parameters [22]. However, the order can be locally adapted for a fixed bandwidth—this aspect is addressed in Section IV-B2.

D. Regularized SURE

The performance of the basic SURE technique can be improved in low SNR regimes by introducing a regularization term in the SURE cost function. Eldar proposed the notion of regularized SURE [16], in which one considers a regularized risk function:

$$\mathcal{R}_\lambda = \mathcal{R} + \lambda\,\mathcal{E}\left\{\frac{1}{N_0}\sum_{i=1}^{N_0}(\mathbf{L}f_i(\mathbf{x}))^2\right\}, \qquad (6)$$

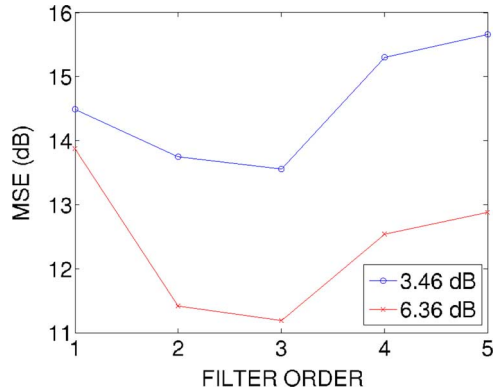

Fig. 3. Variation of MSE with polynomial order for ECG signal 1 for two different input SNRs. The globally optimum order to perform LPR is 3.

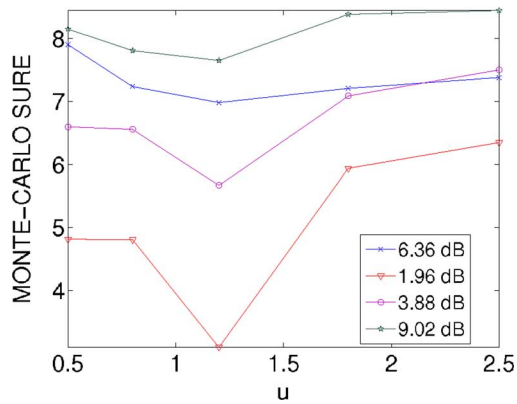

Fig. 4. Monte Carlo SURE versus u; $\lambda = u\sigma^2$.

where $\mathbf{L}$ denotes a regularization operator. Since here, we assume that the clean signal is smooth, we choose to penalize the estimate if it is not smooth by taking $\mathbf{L}$ as the first derivative operator though higher-order derivatives may also be considered. The unbiased estimator for the risk in (6) is

$$\epsilon_\lambda = \epsilon + \frac{\lambda}{N_0}\sum_{i=1}^{N_0}\left(\frac{\partial f_i(\mathbf{x})}{\partial x_i}\right)^2.$$

The regularization parameter λ has an important role in the noise suppression and since we want to have the parameter to be directly related to the noise variance, we express $\lambda = u\sigma^2$, u being a parameter to be determined. With regularization, we note that the estimator of s_n based on x_n is parameterized by both the chosen bandwidth and the regularization parameter. A SURE-based optimization algorithm for the choice of different denoising function parameters is discussed in [23]. It is observed there that the derivative term in the SURE objective (divergence term in higher dimensions) can be computed analytically only when the dependence of the denoising function on the parameters to be optimized is of a certain form. Otherwise, a Monte Carlo approach can be taken to approximate the term, in order to arrive at the estimated MSE value. It is also to be noted that this Monte Carlo SURE approach assists one in selecting the regularization parameter involved in the problem-specific cost, and is more convenient than traditional methods such as generalized cross validation. An important point to be emphasized here is that the algorithm in [23] does not require the explicit

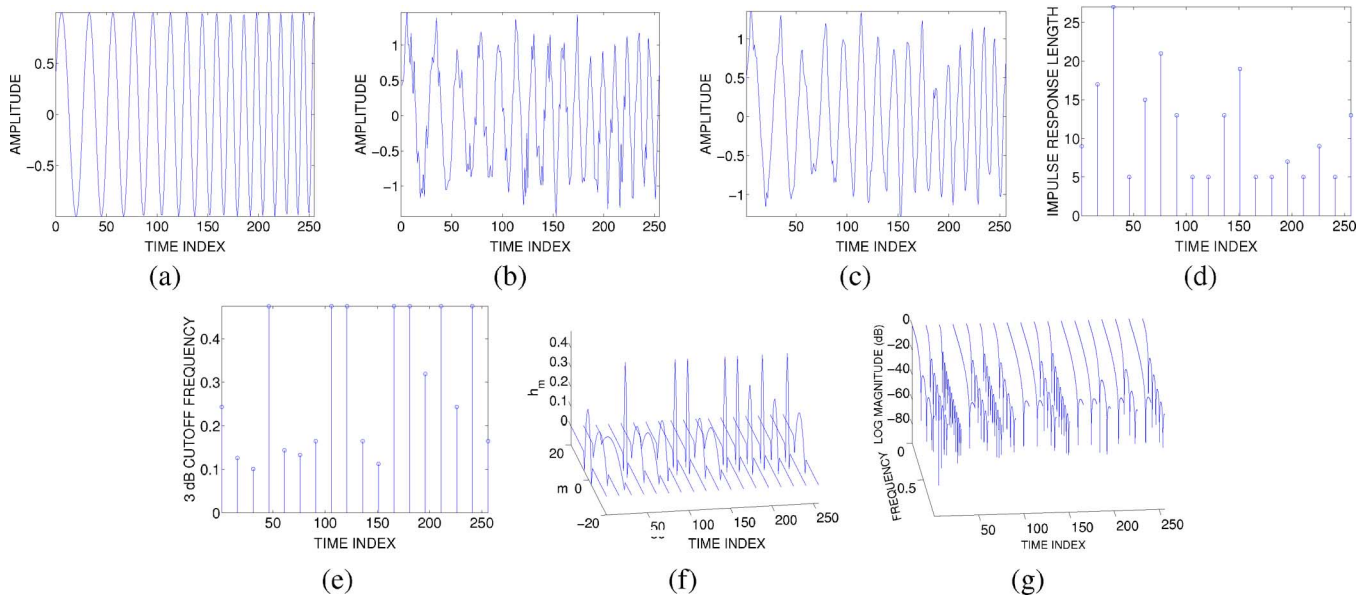

Fig. 5. (a) Linear FM signal, (b) Noisy signal (SNR $= 8.7$ dB), (c) Smoothed output obtained using Algorithm 1 (SNR $= 11.98$ dB), (d) Optimal impulse response lengths versus time, (e) Optimal 3-dB frequencies versus time, (f) Instantaneous filter impulse responses, and (g) corresponding frequency responses.

form of the denoising function to be specified; instead, it only takes into account the response of the algorithm to added noise to compute the MSE estimate. Following their algorithm (see Algorithm 2 in [23]), we compute the divergence term in the SURE cost for different values of λ, obtain the overall SURE objective (cf. (6) in [23]), and average it over multiple noise realizations in order to arrive at a suitable estimate of the MSE. This is repeated for multiple input SNR values. We used additive Gaussian noise of mean zero and standard deviation 0.1 to perturb the observation vector leading to the estimation of the divergence term. The behavior of the Monte Carlo SURE with different λ values for certain input SNRs is given in Fig. 4 (we used ECG signal 1, see Section IV-B). We obtained a value of $\lambda = 1.2\sigma^2$ to be optimum.

IV. Experimental Results

We next test the algorithm detailed in the previous section using both synthesized and real data.

A. Experiment 1

First, we synthesize a linear frequency-modulated chirp as defined below:

$$s_n = \sin\left(0.2\frac{n^2}{256} + 0.2n + 0.4\right), \quad 0 \le n \le 255. \tag{7}$$

Noisy versions of s_n are generated by adding noise samples drawn independently from a Gaussian distribution of mean zero and variance σ^2 and fed as input to our algorithm. Initially, we assume the scenario wherein noise variance is known. The clean signal is shown in Fig. 5(a) and the noisy signal in Fig. 5(b) with the noise standard deviation being $\sigma = 0.25$. The smoothed filter output obtained is shown in the Fig. 5(c). Note that there is more than 3-dB improvement in SNR. The impulse response lengths, the corresponding 3-dB frequencies, the actual impulse response plots, and frequency responses of the optimum S-G filters at a subset of the reconstruction instants are shown in Figs. 5(d)–(g). Note that, in plotting the frequency responses in this paper, we have used a frequency axis normalized with respect to π.

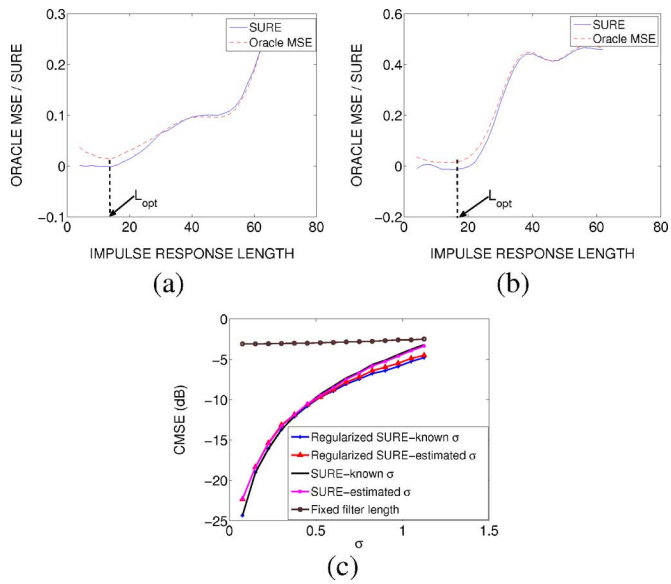

Fig. 6. (Color in electronic version) Plot of SURE and Oracle MSE versus filter length for instant (a) 3, and (b) 7; (c) Plot of CMSE obtained using fixed length filter (impulse response length $= 37$), and estimators using SURE and regularized SURE versus noise standard deviation for the cases of σ known and estimated σ.

As seen from the plots, we obtain considerable noise reduction, with the SURE-optimal S-G filter impulse response lengths being small (implying broader frequency responses) at the instants where the local signal variation is fast and vice versa at other instants.

Since we have the ground truth in this simulation experiment, it is of interest to consider the match between the oracle MSE and SURE. We consider two instants 3 and 7, and plot the oracle MSE and SURE of the MSE averaged over 100 noise realizations versus the filter length in Figs. 6(a) and (b), respectively.

From the figures it is clear that SURE provides a good approximation to the oracle MSE with the variance of the risk estimator being greater at lower filter lengths. More importantly, their minima match.

Next, we proceed to observe the statistical performance of our estimators with the two cost functions—SURE and regularized SURE. That an estimator based on a fixed length S-G filter is inferior to the variable length filter-based technique is also to be shown explicitly. Also, the importance of having exact knowledge of noise variance in the performance of the algorithm needs investigation. For this, we intend to plot the cumulative MSE (CMSE), which is the MSE averaged over all the reconstruction instants and multiple realizations of the noise (we used 100 noise realizations), versus the noise standard deviation σ, with and without exact knowledge of noise variance and the plot so obtained is shown in Fig. 6(c). When the noise standard deviation is not known, we resort to the median estimator [24]:

$$\hat{\sigma} = \frac{\{\text{median}(|x_n - x_{n-1}|;\, n = 2, 3, \ldots, N)\}}{0.6745}.$$

From the CMSE obtained, we see that noise estimation does not deteriorate the estimator performance independent of whether we use a simple SURE cost function or a modified version. However, it is to be noted that using a regularized cost pays dividends towards higher noise levels than the simple SURE cost. Again, one can perceive that the estimator using a fixed length S-G filter (we used a filter length of 37 for illustration) gives poor CMSE performance as compared with the proposed estimators.

Algorithm 2: To Calculate p_{opt} at an Instant $n_0 T$

Require: L to be at least $(p_{\max} + 1)$
Ensure: Optimality with respect to SURE objective $\tilde{\epsilon}$
  $p \leftarrow p_{\min}$
  if $n_0 \neq 0$ **then**
    $x_n \leftarrow x_{n+n_0}$
  end if
  while $p \leq p_{\max}$ **do**
    Employ LS-fit over $[-\frac{LT}{2}, \frac{LT}{2}]$
    Evaluate $\tilde{\epsilon}$ with this p
    $p \leftarrow p + 1$
  end while
  $p_{\text{opt}} \leftarrow \arg\min \tilde{\epsilon}(p)$

B. Experiment 2

In this section, we provide results on ECG signals with two variants of the SURE-optimal S-G filter selection algorithm, the first being that based on adaptive filter length as discussed so far in the paper.

1) Results With Adaptive Filter Length—Algorithm 1: We now test the performance of the algorithm with real-world ECG signals, which have a nice temporal structure important to be preserved by our processing. An earlier work applying S-G filters for ECG signal processing [6], investigated the best choice of the filter order and length for smoothing considering the performance on the smallest relevant waves in ECG. We perform the test with two different ECG signals, both of which are taken from the PhysioBank database [25], where the signals are recorded at 720 Hz. In Fig. 7(a), we show the first measured signal (with inherent measurement noise), 7(b), the simulated noisy signal, 7(c), the enhanced signal using SURE cost (Algorithm 1 with SURE abbreviated as $A_{1,\text{S}}$ henceforth), 7(d) and 7(e), the impulse response lengths and the corresponding cutoff frequencies, 7(f) and (g), the actual impulse and frequency responses respectively, of the optimal S-G filters obtained at a subset of the reconstruction instants using the algorithm. The corresponding figures with regularized SURE cost (Algorithm 1 with regularized SURE abbreviated as $A_{1,\text{RS}}$ henceforth) are also depicted in Fig. 7. We note that the filter lengths chosen are again consistent with the results obtained in the simulation experiment in the sense that a large pass band filter is chosen at the instants where the local variation of the noisy signal is slow so that better smoothing is obtained, resulting in low variance of the estimator, and a filter with a smaller pass band is chosen in other positions where the signal variation locally, is fast, in order not to miss out on the rapid signal changes. Also, one can observe that the SNR improvement obtainable using the regularized risk is considerable, it being close to 7 dB in this case. $A_{1,\text{S}}$ gives more than 4 dB gain. It is evident that the algorithm using the modified risk selects relatively a greater fraction of large impulse response lengths, resulting in more smoothing. The same series of figures are plotted for the second signal using $A_{1,\text{S}}$ and $A_{1,\text{RS}}$ and are shown in Fig. 8. One can conclude that there is considerable improvement in SNR, again, in both the cases. Specifically, the gain is above 8 dB with $A_{1,\text{RS}}$, and $A_{1,\text{S}}$ gives close to 5 dB.

2) Results With Adaptive Filter Order—Algorithm 2: Up to now, we focussed on choosing the appropriate bandwidth parameter keeping the order fixed at three. In this discussion, we address the problem of adaptive MMSE order selection, for a given bandwidth. Our interest is to show that SURE can be used to address the order selection issue as well. The adaptive order problem is addressed in [22], where Fan and Gijbels noted that with an adaptive, data-driven order selection algorithm, a simple rule of thumb can be used for the bandwidth choice. They found that the variable order algorithm is robust to bandwidth choice. That is, if the bandwidth is too large, then the algorithm would select a higher-order polynomial to reduce the bias, and vice versa—a concept referred to as *bandwidth robustification*. It is worth noting that the approach taken in [22] is one of empirical risk estimation, which can therefore, also be solved using SURE. For the sake of illustration, we consider ECG signal 1, add white Gaussian noise to obtain an input SNR of 6.36 dB as in Fig. 7, and use a fixed filter length of 31 samples. Algorithm 2 follows essentially the same steps as in $A_{1,\text{S}}$, with the exception that we now compute the SURE cost over different filter orders (we used $p_{\min} = 1$, $p_{\max} = 5$) (the algorithm being henceforth referred to as $A_{2,\text{S}}$). The results are shown in Fig. 9.

We observe that the result exhibits more than 4 dB improvement over the one obtained with $A_{1,\text{S}}$ (cf. Fig. 7(c)), where the performance was slightly affected by the higher variance of the risk estimator at low filter lengths. From the optimum filter order

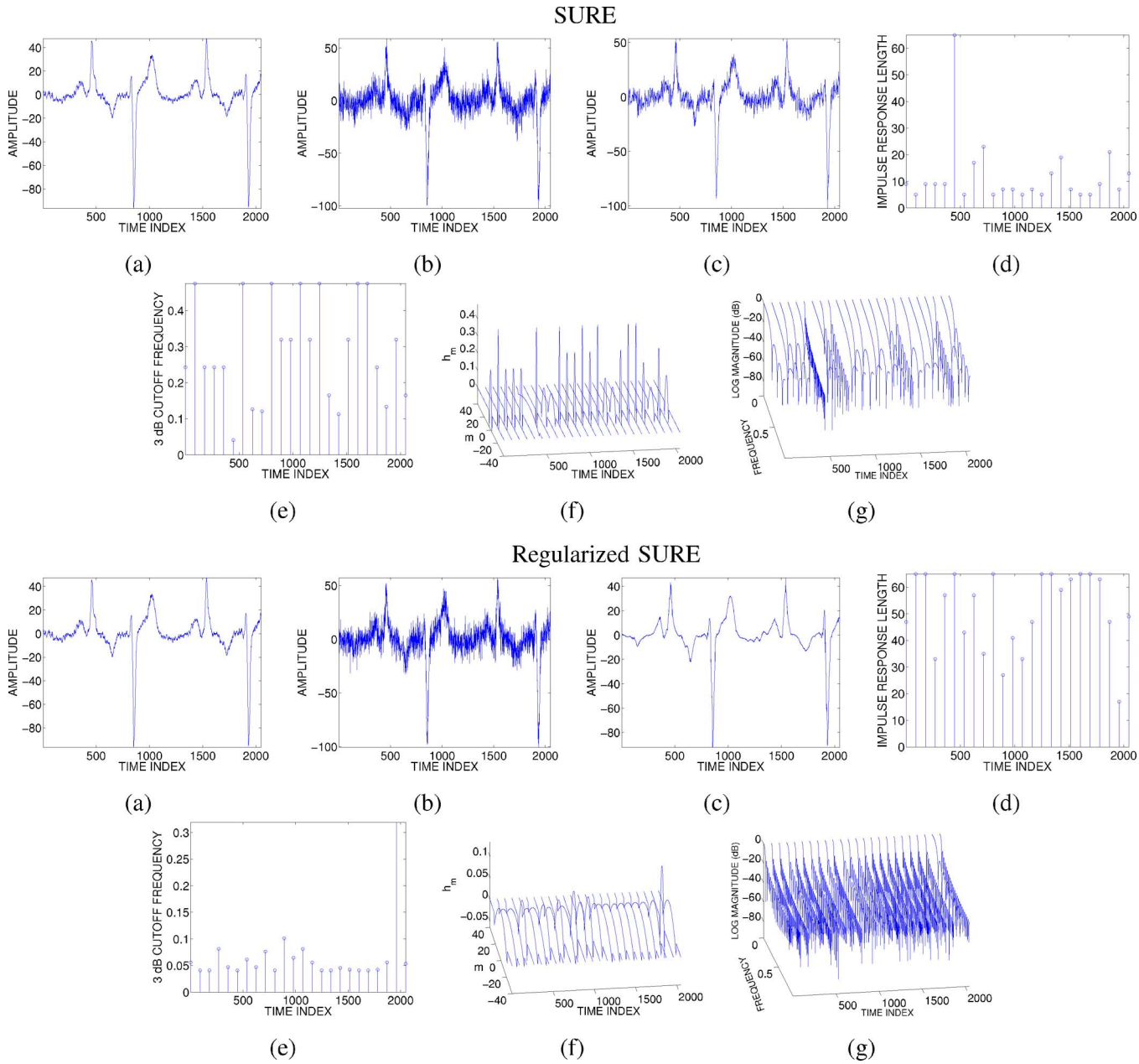

Fig. 7. (a) Measured ECG signal 1, (b) Simulated noisy signal (SNR $= 6.36$ dB), (c) Smoothed output using SURE (SNR $= 10.55$ dB) and regularized SURE (SNR $= 13.19$ dB), (d) Optimal impulse response lengths vs. time, (e) Optimal 3-dB frequencies vs. time, (f) Corresponding filter impulse responses, and (g) frequency responses.

plot (cf. Fig. 9(c)), we conclude that higher-order filters are selected local to the signal peaks and valleys (implying an S-G filter of higher cutoff frequency), and vice versa at the other instants. We adapted the fixed filter length with the input SNR in such a manner that a higher length is chosen with increased noise strengths, and vice versa. This was found to provide better results than fixing the filter length for all input SNRs. This is reasonable as the higher filter length also serves to suppress more noise in the low SNR regime, with the lower length assisting the adaptive order algorithm in preserving the signal features better in the high SNR regime. In the experiments with different SNRs, we changed the value of L from 70 for the lowest SNR (-3.5 dB) to 10 for the highest SNR considered (20 dB).

C. Performance Comparison With Other Denoising Algorithms

Since we now have an algorithm for ECG denoising, it is instructive to compare it against various popular and performant denoising algorithms available in the literature. First, we performed a comparative study of the proposed algorithm and the standard wavelet-based methods making use of soft-thresholding [26], wherein the threshold is selected as $\sigma\sqrt{2\log(N)}$. The performance of our algorithms, $A_{1,\mathrm{S}}$, $A_{1,\mathrm{RS}}$, and $A_{2,\mathrm{S}}$ is compared with that obtained using four wavelets in the wavelet-based denoising algorithm—Daubechies 6,

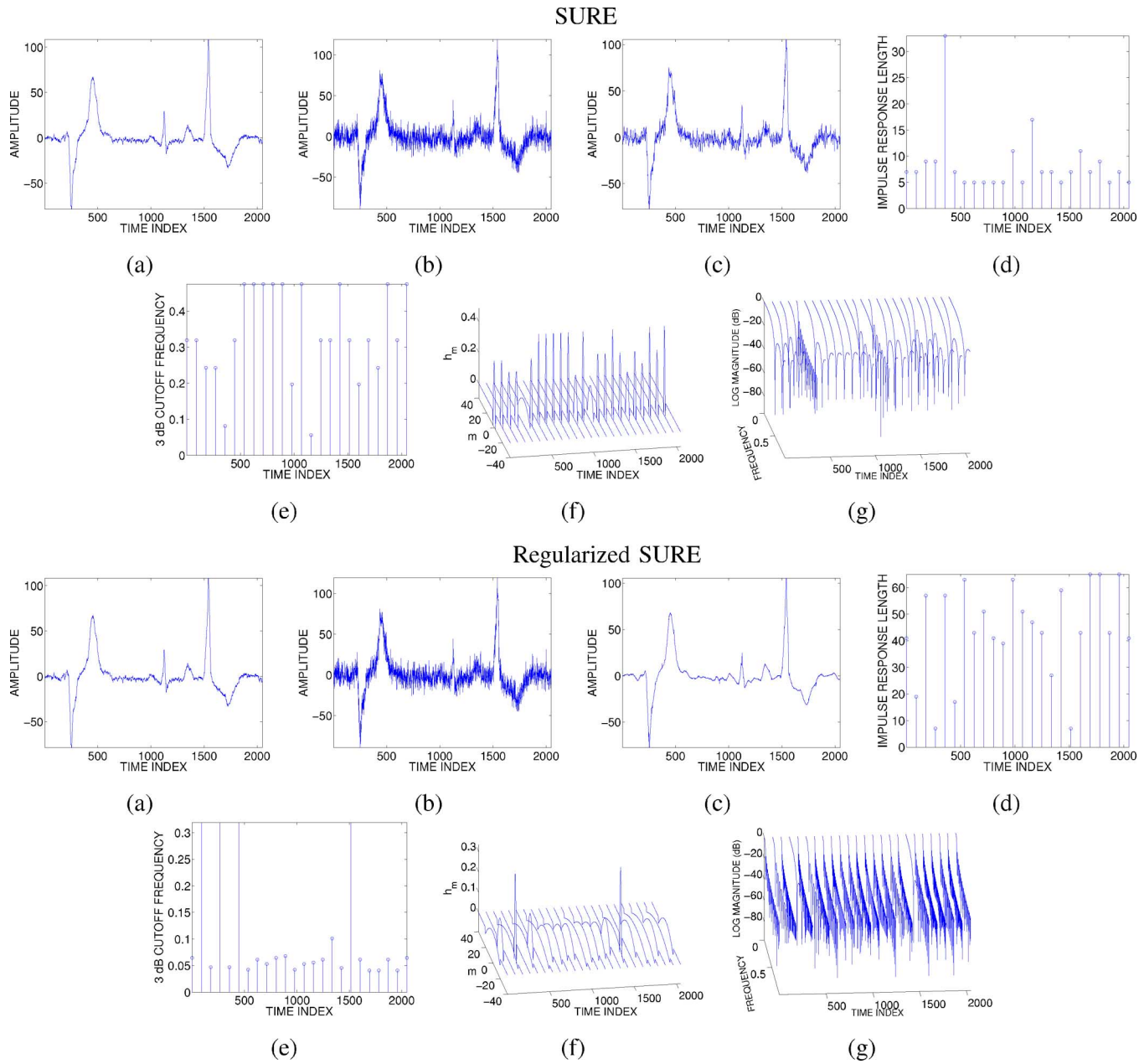

Fig. 8. (a) Measured ECG signal 2, (b) Simulated noisy signal (SNR $= 8.12$ dB), (c) Smoothed output using SURE (SNR $= 13.06$ dB) and regularized SURE (SNR $= 16.92$ dB), (d) Optimal impulse response lengths versus time, (e) Optimal 3-dB frequencies versus time, (f) Corresponding filter impulse responses, and (g) frequency responses.

Daubechies 14, Symmlet 4, and Coiflet[1]. In fact, it was observed experimentally that the output SNR obtained in the wavelet technique did not vary considerably with different wavelets, and hence we limited the presentation of results to these four wavelets.

A three-level decomposition is performed in all the cases. We consider the output SNRs obtained by using $A_{1,\mathrm{S}}$, $A_{1,\mathrm{RS}}$, $A_{2,\mathrm{S}}$, and wavelet denoising over a wide range of input SNRs for the two ECG signals shown in Figs. 7(a) (signal 1) and 8(a) (signal 2). The resulting plots are shown in Figs. 10(a) and 10(b), respectively. From the plots, we observe that at low input SNR values (-3.5 to 5 dB), $A_{1,\mathrm{RS}}$ gives superior performance as compared with the wavelet-based technique by about 6 dB, whereas $A_{1,\mathrm{S}}$ does not result in improvement. The gain obtained by adding the regularization term persists till the input

[1]Implemented using WAVELAB toolbox available at: http://www-stat.stanford.edu/~wavelab/Wavelab_850/index_wavelab850.html.

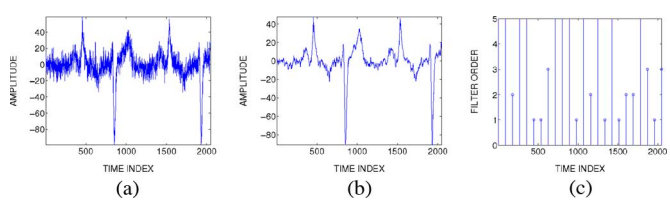

Fig. 9. (a) Simulated noisy ECG signal (SNR $= 6.36$ dB), (b) Smoothed output using SURE-optimal filter orders (SNR $= 14.62$ dB), (c) Optimal filter orders vs. time.

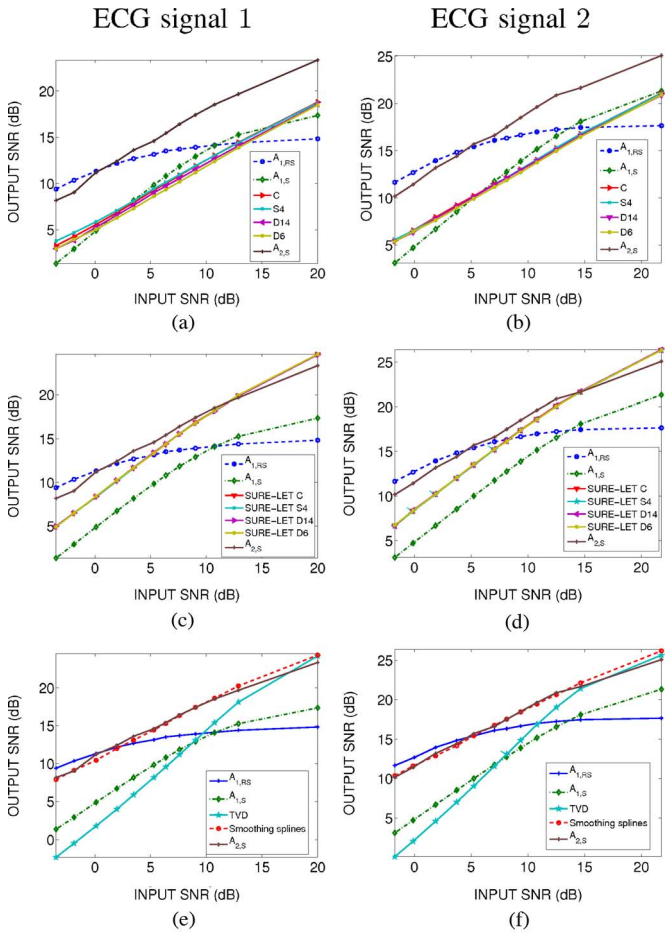

Fig. 10. (Color in electronic version) A comparison of the denoising performance of the proposed algorithms ($A_{1,\mathrm{S}}$ and $A_{1,\mathrm{RS}}$—Adaptive filter length selection algorithms using SURE and regularized SURE, respectively, and $A_{2,\mathrm{S}}$—Adaptive filter order selection algorithm using SURE) with various popular and performant denoising methods reported in the literature. Row 1: wavelet-based denoising techniques using soft-thresholding; Row 2: OWT SURE-LET of [9]; and Row 3: Regularized denoising methods—smoothing splines and TVD. The wavelet abbreviations are as follows: Coiflet: C, Symmlet 4: S4, Daubechies 14 and Daubechies 6: D14 and D6, respectively.

SNR reaches about 13 dB after which the SURE cost takes over. The performance of $A_{1,\mathrm{S}}$ itself is comparable to that obtained using the wavelet technique in the mid-to-high SNRs. It outperforms the wavelet thresholding method over a wide range of input SNRs, resulting in a maximum relative improvement of about 1.3 dB. $A_{2,\mathrm{S}}$ outperforms all considered algorithms from an input SNR of 0 dB, up to which $A_{1,\mathrm{RS}}$ results in better gain.

Next, we consider a novel and sophisticated orthonormal wavelet denoising algorithm proposed in [9], which elegantly combines the SURE formalism with a linear expansion of thresholds (SURE-LET) for the denoising function. Luisier *et al.* observed that their technique outperforms the classical *Bayesshrink* by $+1$ dB on the average. It should be noted that *Bayesshrink* has been seen to have better MSE performance than *Visushrink*, *Sureshrink*, etc. Therefore, we consider SURE-LET based denoising algorithm[2] with the same four wavelets as in the standard wavelet denoising algorithms, and compare the average performance on the two ECG signals. The results obtained in this case for both signals are depicted in Figs. 10(c) and (d), respectively. Though $A_{1,\mathrm{S}}$ performs poorer compared with the SURE-LET-based technique for every value of input SNR, for very low SNRs (we considered values from -3.5 dB) up to an SNR of about 6 to 7 dB, $A_{1,\mathrm{RS}}$ outperforms the SURE-LET algorithm with the SNR gain being close to 5 dB at the lowest SNR value. Also, $A_{2,\mathrm{S}}$ outperforms SURE-LET up to an input SNR value of close to 12 dB with a maximum relative gain of around 3.2 dB. Only at the highest SNR considered, does its performance fall slightly below that of SURE-LET.

In order to fairly compare the proposed technique using a regularized objective, we need to take other denoising algorithms relying on optimizing regularized costs into account. For this, we consider spline smoothing [27], and total-variation denoising (TVD) [28], for which the results are shown in Figs. 10(e) and (f), respectively. From the plots, we observe that better performance is obtained using $A_{1,\mathrm{RS}}$ than a cubic smoothing spline for low SNRs. The SNR improvement using the proposed algorithm persists for an input SNR of 2 dB for the signal 1, and up to 5 dB for the second signal. A maximum gain of 1.5 dB is observed for the lowest SNR value considered. However, TVD gives lower SNR gains than $A_{1,\mathrm{RS}}$ up to an input SNR of close to 10 dB, after which it dominates. We also note that optimizing the SURE objective in our algorithm gives better results than TVD in the low to mid-SNR ranges. Even though $A_{2,\mathrm{S}}$ is not based on a regularized cost, its performance is comparable to or better than spline smoothing up to high SNRs, and is also consistently better than TVD. Only at the last two SNR values considered, does it provide slightly lower SNR gains as compared with smoothing splines.

We observe that the performance of $A_{1,\mathrm{S}}$ suffers from the increased variability of the risk estimator at lower sample sizes (note that we start with a filter length of 5). This in turn results in an optimum filter length of 5 at flat portions of the signal also, as seen from Figs. 7 and 8, implying lesser smoothing. The regularization term penalizes the non-smoothness of the estimate, thus resulting in improved performance at low SNRs. The increase in SURE variance at low filter sizes, combined with the extra regularization term explains the deterioration in performance obtained from the regularized cost minimization at high SNRs. Also the performance of the adaptive, data-driven filter order selection algorithm outperforms, or at least gives comparable SNR gains as almost all algorithms considered. In summary, $A_{1,\mathrm{RS}}$ is robust at low SNRs, with $A_{2,\mathrm{S}}$ being the algorithm of choice at other input SNR values.

D. A Note on Computational Complexity of the Bandwidth and Order Selection Algorithms

Even though LS fitting of polynomials and their evaluation are computationally intensive, we note from the S-G filtering perspective that the result is an $\mathcal{O}(N)$ algorithm, N being the number of observations. The smoothed output at the center of the interval (which is the origin) is a_0, with the corresponding kth derivative at 0 being $k!a_k$; $k = 1, 2, \ldots, p$. A table of the filter coefficients corresponding to different filter lengths and orders is given in [4], but they can also be pre-computed. With

[2]SURE-LET MATLAB software is available at the website: http://bigwww.epfl.ch/demo/suredenoising/.

the filtering operations, we computed the total number of operations performed in our algorithm. The rows of the $\mathbf{H}$ matrix, which are dependent only on M and p (and therefore pre-computed in the implementation), represent the flipped filter impulse response coefficients for obtaining the polynomial coefficients. For computing the derivative term in the SURE cost function, we need to obtain the trace of the square matrix $\mathbf{AH}$, which can also be computed a priori. The results presented for the bandwidth selection algorithm correspond to filter lengths $5, 7, 9, \ldots, 65$, and for order selection, the polynomial orders considered were 1, 2, 3, 4, and 5.

Let ℓ_r denote the filter length chosen at the iteration index r, and I, the total number of iterations at a particular instant. For each index $n \in 1, 2, \ldots, N$, the exact number of operations in the rth iteration when using $A_{1,\mathrm{S}}$ is $[4\mathcal{O}(\ell_r \log_2 \ell_r) + 15.5\ell_r - 1.5]$, and with $A_{1,\mathrm{RS}}$ is $[4\mathcal{O}(\ell_r \log_2 \ell_r) + 17.5\ell_r + 1.5]$. The total number of computations can be obtained by summing up these values over the filter lengths at various iterations, $r = 1, 2, \ldots, I$, and then multiplying with the number of observations. We have made $I = 31$ iterations (since maximum filter length is 65) over $N = 2048$ data samples. Since we are computing the objective function at each instant for different filter lengths, the number of computations is high when compared with the denoising algorithms to which its performance has been compared. For each index n, the number of computations in the rth iteration of $A_{2,\mathrm{S}}$ (with a filter order of p_r) is $[(p_r+1)\mathcal{O}(\ell \log_2 \ell) + 3p_r\ell - 3p_r + 6.5\ell + 1.5]$, with ℓ denoting the fixed filter length used. We have performed $I = 5$ iterations (since, we choose $p_{\max} = 5$) over $N = 2048$ data points. Typically, the number of iterations in this case is much lesser compared with that of the bandwidth selection algorithms, resulting in considerable computational savings. We next present the average execution times in MATLAB on a Macintosh machine having a 2×2.4 GHz quad-core Intel Xeon processor, along with the average times taken by the other benchmarked methods considered (cf. Table I). We have taken 2048 samples of ECG signal 1 at SNR of 6.36 dB for the experiment, and the result is averaged over 10 noise realizations. We observe that $A_{2,\mathrm{S}}$ is fastest among the proposed techniques as the number of iterations in this case is lower (5, as against 31 for the bandwidth selection algorithms), followed by $A_{1,\mathrm{S}}$ and $A_{1,\mathrm{RS}}$, both taking almost equal computation times. The other state-of-the-art methods offer lower computation times, though in terms of performance, our techniques present a competitive alternative. The proposed S-G filter-based algorithms may be further accelerated by optimizing the software using a C language-based implementation.

TABLE I
COMPARISON OF EXECUTION TIMES OF DIFFERENT TECHNIQUES

Method	Computation time (seconds)
$A_{1,\mathrm{S}}$	0.4407
$A_{1,\mathrm{RS}}$	0.4347
$A_{2,\mathrm{S}}$	0.0854
Wavelet soft-thresholding	0.0028
SURE-LET	0.0093
Smoothing splines	0.0163
TVD	0.0270

V. CONCLUSION

We addressed the problem of selecting optimal length S-G filters for data smoothing. Though having a fixed length filter is reasonable in the noise-free case, the presence of noise requires us to choose a filter with a different impulse response length at each reconstruction instant. We formulated the optimal filter selection problem using a pointwise MMSE criterion. From the bias and variance expressions of the local regression-based estimator, it was clear that the dependence of these two estimator characteristics on the filter length was conflicting. This was identified as the bias-variance tradeoff to be solved for the filter selection and subsequently, we took the help of SURE in solving the tradeoff. The addition of a regularization term in the SURE cost function resulted in better performance in the low SNR regime. The resulting algorithm was tested on both synthetic and real data providing satisfactory results and helping us in choosing the "optimal S-G filter length" (or, equivalently, the "optimal S-G filter 3-dB cutoff frequency") at each instant, that is, the filter chosen was found to be of large cutoff frequency when, locally, the signal varied fast thus reducing the error due to bias of the estimator, and of small cutoff frequency at those portions where the local signal variation was slow, thereby reducing the variance of the estimator. Subsequent to addressing the bandwidth selection issue, we demonstrated that a SURE formalism is feasible for tackling the adaptive order problem, with the filter length being fixed according to the noise level. The results with ECG signals using the adaptive order selection algorithm exhibited better performance than other popular denoising algorithms and the adaptive filter length selection algorithms, with lesser computations than the bandwidth selection technique.

We have chosen the ECG denoising problem as an application and proof of concept for SURE-based optimum S-G filter selection. Therefore, even though negative SNRs may not be realistic in ECG signals, we have considered such SNR values to assess the denoising performance under high noise conditions. We note that the two-dimensional counterparts of the proposed algorithms may also be developed. Also, even if the assumption of white Gaussian noise does not hold, we can still determine corresponding SURE-optimal S-G filters based on the generalized SURE versions for exponential families [16]. Again, it is possible to choose the regularization operator $\mathbf{L}$ in (6), to be a linear combination of first and second derivative operators, for instance, and observe the performance of the proposed technique. Our ongoing work includes extension to the case when we have non-uniform data samples.

ACKNOWLEDGMENT

We would like to thank Arun Venkitaraman for leading us to Schafer's article on Savitzky-Golay filters, the anonymous reviewers for their comments, which helped improve the quality of the paper, and the developers of WAVELAB and Luisier *et al.* for making their MATLAB software available online, which facilitated the comparisons.

REFERENCES

[1] J. L. Flanagan, "Speech synthesis," in *Speech Analysis, Synthesis, and Perception*, 2nd ed. New York: Springer-Verlag, 1976, ch. 6, pp. 232–246.

[2] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*, 1st ed. London, U.K.: Chapman & Hall, 1996, pp. 57–303.

[3] W. Hardle, *Applied Nonparametric Regression*, 1st ed. Cambridge, U.K.: Cambridge Univ. Press, 1990.

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

[5] Z. G. Zhang, Y. S. Hung, and S. C. Chan, "Local polynomial modeling of time-varying autoregressive models with application to time-frequency analysis of event-related EEG," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 557–566, Mar. 2011.

[6] S. Hargittai, "Savitzky-Golay least-squares polynomial filters in ECG signal processing," in *Proc. Comput. Cardiol.*, Budapest, Hungary, 2005, pp. 763–766.

[7] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.

[8] M. Raphan and E. P. Simoncelli, "Least squares estimation without priors or supervision," *Neural Comput.*, vol. 23, pp. 374–420, 2011.

[9] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 593–606, Mar. 2007.

[10] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[11] A. Benazza-Benyahia and J.-C. Pesquet, "Building robust wavelet estimators for multicomponent images using Stein's principle," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1814–1830, Nov. 2005.

[12] N. R. Muraka and C. S. Seelamantula, "A risk-estimation-based comparison of mean square error and itakura-saito distortion measures for speech enhancement," in *Proc. INTERSPEECH*, Florence, 2011, pp. 349–352.

[13] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2778–2786, Nov. 2007.

[14] J.-C. Pesquet and D. Leporini, "A new wavelet estimator for image denoising," in *Proc. 6th Int. Conf. Image Process. Its Appl.*, Dublin, Ireland, 1997, vol. 1, pp. 249–253.

[15] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 3, pp. 407–422, May 2011.

[16] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, Feb. 2009.

[17] S. Seifzadeh, M. Rostami, A. Ghodsi, and F. Karray, "Parameter selection for smoothing splines using Stein's unbiased risk estimator," in *Proc. Int. Joint Conf. Neural Netw.*, San Jose, CA, 2011, pp. 2733–2740.

[18] R. W. Schafer, "What is a Savitzky-Golay filter?," *IEEE Signal Process. Mag.*, vol. 28, no. 4, pp. 111–117, July 2011.

[19] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, 1st ed. Englewood Cliffs, NJ: Prentice-Hall, 1993, pp. 15–77.

[20] H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, 3rd ed. Cambridge, U.K.: Cambridge Univ. Press, 1988, pp. 446–448.

[21] P. Brigger, J. Hoeg, and M. Unser, "B-spline snakes: A flexible tool for parametric contour detection," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1484–1496, Sept. 2000.

[22] J. Fan and I. Gijbels, "Adaptive order polynomial fitting: Bandwidth robustification and bias reduction," *J. Comput. Graph. Statist.*, vol. 4, no. 3, pp. 213–227, Sept. 1995.

[23] S. Ramani, T. Blu, and M. Unser, "Monte Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, Sep. 2008.

[24] V. Katkovnik and L. Stankovic, "Instantaneous frequency estimation using the Wigner distribution with varying and data-driven window length," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2315–2325, Sep. 1998.

[25] PhysioBank database [Online]. Available: http://www.physionet.org/physiobank/database/aami-ec13/

[26] D. L. Donoho, "Denoising by soft thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[27] C. H. Reinsh, "Smoothing by spline functions," *Numer. Math.*, vol. 10, pp. 177–183, 1967.

[28] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.

Sunder Ram Krishnan was born in Thiruvananthapuram, Kerala, India, on December 29, 1988. He received the Bachelor's of Technology degree in electronics and communication engineering from the College of Engineering, Thiruvananthapuram, University of Kerala, in 2010.

Currently, he is working toward the Master's of Science degree in signal processing at the Department of Electrical Engineering, Indian Institute of Science, Bangalore, India, under the supervision of Dr. C. Sekhar Seelamantula. His research interests include statistical signal processing, time-frequency methods in signal processing, and sampling theory.

Chandra Sekhar Seelamantula (SM'12) was born in Gollapalem, Andhra Pradesh, India, on August 5, 1976. He received the B.E. degree in electronics and communication engineering, with gold medal, from Osmania University, Hyderabad, India, in 1999 and the Ph.D. degree from the Indian Institute of Science, Bangalore, India, in 2005.

During his research at the Indian Institute of Science, he received the IBM (India Research Laboratory, New Delhi) Research Fellowship for the period 2001 to 2004. From April 2005 to March 2006, he worked as Technology Consultant at M/s. Esqube Communication Solutions Private Limited, where he developed proprietary audio coding solutions. From April 2006 to July 2009, he worked as a Postdoctoral Fellow at the Biomedical Imaging Group with Prof. M. Unser, at the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. On July 20, 2009, he joined as Assistant Professor at the Department of Electrical Engineering, Indian Institute of Science. His main research interests include nonstationary signal processing methods with applications to speech/audio, time-frequency representations, auditory modeling, bioinspired signal processing, sampling theories, and biomedical imaging.