Fully Adaptive Savitzky-Golay Type Smoothers

1st Maciej Niedźwiecki
Department of Automatic Control
Gdańsk University of Technology
Gdańsk, Poland
maciekn@eti.pg.edu.pl

2nd Marcin Ciołek
Department of Automatic Control
Gdańsk University of Technology
Gdańsk, Poland
marcin.ciolek@pg.edu.pl

***Abstract*—The problem of adaptive signal smoothing is considered and solved using the weighted basis function approach. In the special case of polynomial basis and uniform weighting the proposed method reduces down to the celebrated Savitzky-Golay smoother. Data adaptiveness is achieved via parallel estimation. It is shown that for the polynomial and harmonic bases and cosinusoidal weighting sequences, the competing signal estimates can be computed in both time-recursive and order-recursive way.**

***Index Terms*—signal denoising, adaptive selection of estimation bandwidth and model order, Savitzky-Golay smoothers**

I. Introduction

The need for smoothing of noise-corrupted signals arises in many practical applications. When the prior knowledge about the signal – in the form of a stochastic or deterministic model – is available, and when noise characteristics are known, smoothing can be performed in a statistically efficient way. For example, when the signal admits the linear state space description and the additive measurement noise is Gaussian, the optimal (in the mean square sense) smoother can be designed based on the Kalman filter theory [1].

In the absence of prior knowledge, one has to rely on some general purpose smoothing schemes, such as the local estimation techniques, which provide a sequence of "pointwise" signal estimates, or block-oriented methods based on wavelet decomposition and thresholding (shrinkage). The most popular local estimation methods include polynomial approximation [2], [3], kernel regression [4], [5] and order-statistical filtering [6], [7]. The wavelet shrinkage procedures [8], [9], [10], [11] can be operated (with hard or soft thresholding) using different wavelet bases.

Even though wavelet shrinkage denoising does not require any assumptions about the nature of the signal, and has been theoretically proven to be nearly optimal when degree of signal smoothness is unknown, its comparison with multi-scale local estimation techniques is far from conclusive. In particular, as demonstrated in [12], [13], [14] on a number of artificially corrupted 1D and 2D benchmark signals, for moderate and low signal to noise ratios (SNR $\leq$ 30), very simple *medley* smoothers, combining several averaging and median filters with different estimation bandwidths, outperform the state-of-the-art wavelet-based procedures, both in the quantitative sense (according to the mean square error measure) and qualitative sense (according to the perceptual structural similarity measure, developed for image processing applications). Similar results, obtained for adaptive multiresolution kernel smoothers, were reported in [15]. This means that development of new local estimation techniques has not lost its significance and still remains an important research topic.

Our present work is focused on the basis function type smoothers, the well-known example of which, corresponding to the choice of polynomial basis, are Savitzky-Golay (SG) filters [16], [17]. Due to their analytical and computational simplicity, and good smoothing capabilities, Savitzky-Golay filters have been extensively used in such research areas as spectroscopy [18], [19], voltammetry [20] and biomedical signal processing [21], [22], [23], among many others. Prior to using a Savitzky-Golay filter, two important decisions must be taken, about the order of the approximating polynomial and the estimation bandwidth, i.e., the size of the local analysis (fitting) window. When these design parametrs are inappropriately chosen (underfitted or overfitted), the smoothing results deteriorate, both in the quantitative and qualitative sense. To cope with this problem a number of adaptive Savitzky-Golay algorithms were proposed in the literature allowing one to select in a signal-dependent way the filter order [20], its bandwidth [25], [26], or both [2], [19], [23].

The paper presents the first unified treatment of the problem of adaptive joint order and bandwidth tuning of SG-like filters, valid for an *arbitrary* choice of basis and weighting functions.

II. Problem statement

Consider the problem of recovering a signal $s(t)$ from a sequence of noisy measurements $y(t)$

$$y(t) = s(t) + \eta(t) \tag{1}$$

where $t = \ldots, -1, 0, 1, \ldots$ denotes normalized (dimensionless) discrete time, and $\eta(t)$ denotes additive measurement noise. In the absence of a phenomenological model of $s(t)$ – whether deterministic or stochastic – we will pursue a local estimation technique based on "curve fitting": we will assume that at each instant t the signal $s(t)$ can be locally approximated by a linear combination of a certain number of known functions of time, further referred to as basis functions.

This work was partially supported by the National Science Center under the agreement UMO-2018/29/B/ST7/00325. The authors are with the Faculty of Electronics, Telecommunications and Informatics. Calculations were carried out at the Academic Computer Centre in Gdańsk.

However, neither the order of approximation (the number of basis functions), nor its local range of applicability will be regarded as known. More precisely, at each time instant t we will consider a family of hypothetical signal descriptions of the form

$$\mathcal{H}_{n|k}(t): \quad \begin{aligned} & y(t+i) = s(t+i) + \eta(t+i) \\ & s(t+i) = \mathbf{f}^{\mathrm{T}}_{n|k}(i)\boldsymbol{\beta}_n, \quad \mathrm{var}[\eta(t+i)] = \rho(t) \\ & i \in I_k = [-k, k] \end{aligned} \tag{2}$$

where $k \in \mathcal{K} = \{k_1, \ldots, k_K\}$ denotes the approximation range, $n \in \mathcal{N}_k = \{1, \ldots, N_k\}$, $N_k = \min(k, N)$, denotes the order of approximation, and $\mathbf{f}_{n|k}(i) = \left[f_{1|k}(i), \ldots, f_{n|k}(i)\right]^{\mathrm{T}}$, $i \in I_k$, denotes the vector of n linearly independent discrete-time basis functions. We will assume that, for each value of t, $\{\eta(t+i), i \in I_k\}$ is a sequence of independent random variables with constant but unknown variance $\rho(t)$.

The two possible choices of basis functions, which have some computational advantages that will be discussed later, are powers of time (corresponding to Taylor series approximation)

$$f_{l|k}(i) = \left(\frac{i}{k}\right)^{l-1}, \quad l = 1, \ldots, n, \quad i \in I_k \tag{3}$$

and harmonic functions (corresponding to Fourier-type approximation)

$$\begin{gathered} f_{1|k}(i) = 1 \\ f_{2l|k}(i) = \sin\frac{l\pi i}{k}, \quad f_{2l+1|k}(i) = \cos\frac{l\pi i}{k} \\ l = 1, \ldots, n_0, \quad n = 2n_0 + 1, \quad i \in I_k \ . \end{gathered} \tag{4}$$

Assuming validity of the local signal representation (2), the vector of coefficients $\boldsymbol{\beta}_n$ can be estimated using the method of weighted least squares

$$\widehat{\boldsymbol{\beta}}_{n|k}(t) = \arg\min_{\boldsymbol{\beta}_n} \sum_{i \in I_k} w_k(i) \left[y(t+i) - \mathbf{f}^{\mathrm{T}}_{n|k}(i)\boldsymbol{\beta}_n\right]^2$$

where $\{w_k(i), i = -k, \ldots, k\}$, $w_k(0) = 1$, denotes a nonnegative, symmetric bell-shaped window of width $2k+1$ used for localization purposes. Straightforward calculations lead to

$$\widehat{\boldsymbol{\beta}}_{n|k}(t) = \mathbf{P}^{-1}_{n|k}\mathbf{p}_{n|k}(t) \tag{5}$$

where

$$\begin{aligned} \mathbf{P}_{n|k} &= \sum_{i \in I_k} w_k(i)\mathbf{f}_{n|k}(i)\mathbf{f}^{\mathrm{T}}_{n|k}(i), \\ \mathbf{p}_{n|k}(t) &= \sum_{i \in I_k} w_k(i)\mathbf{f}_{n|k}(i)y(t+i) \ . \end{aligned}$$

Based on (5), the local estimate of $s(t)$ can be obtained in the form

$$\widehat{s}_{n|k}(t) = \mathbf{f}^{\mathrm{T}}_{n|k}(0)\widehat{\boldsymbol{\beta}}_{n|k}(t) \ . \tag{6}$$

When needed, in a similar way one can obtain the estimate of the l-th derivative of $s(t)$ (see e.g. [27] – [29])

$$\widehat{s}^{(l)}_{n|k}(t) = [\mathbf{f}^{(l)}_{n|k}(0)]^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{n|k}(t) \tag{7}$$

where

$$\mathbf{f}^{(l)}_{n|k}(i) = \widetilde{\mathbf{f}}^{(l)}_{n|k}(\tau)|_{\tau=i}, \quad \widetilde{\mathbf{f}}^{(l)}_{n|k}(\tau) = \frac{d^l\widetilde{\mathbf{f}}_{n|k}(\tau)}{d\tau^l}$$

and $\widetilde{\mathbf{f}}_{n|k}(\tau), \tau \in R$, denotes the continuous time "prototype" of the basis function vector $\mathbf{f}_{n|k}(i)$.

Our task will be to choose, at each time instant t, the most appropriate values of $n \in \mathcal{N}_k$ and $k \in \mathcal{K}$, leading to the following local approximation of $s(t)$

$$\widehat{s}(t) = \mathbf{f}^{\mathrm{T}}_{\widehat{n}(t)|\widehat{k}(t)}(0)\widehat{\boldsymbol{\beta}}_{\widehat{n}(t)|\widehat{k}(t)}(t) \ . \tag{8}$$

The choice of n and k should depend on local properties of the approximated signal, such as the local compliance with the incorporated basis functions and the local signal-to-noise ratio. In statistical terms the best choice is the one that trades off the bias and variance components of the mean squared signal estimation error (MSE)

$$\mathrm{MSE}(t) = \mathrm{E}\{[s(t) - \widehat{s}(t)]^2\} = \mathrm{MSE}_b(t) + \mathrm{MSE}_v(t) \tag{9}$$

where $\mathrm{MSE}_b(t) = [s(t) - \bar{s}(t)]^2$, $\bar{s}(t) = \mathrm{E}[\widehat{s}(t)]$ and $\mathrm{MSE}_v(t) = \mathrm{var}[\widehat{s}(t)] = \mathrm{E}\{[\widehat{s}(t) - \bar{s}(t)]^2\}$. Since the bias component decreases with growing n and decreasing k, while the variance component shows the opposite tendency, some sort of compromise is needed. In the remaining part of the paper we will present and compare two approaches to adaptive selection of n and k.

III. Prediction based approach

Suppose that the noisy signal $y(t)$ obeys (2) and denote by $\widetilde{\Omega}_k(t) = \{\widetilde{\eta}(t-k), \ldots, \widetilde{\eta}(t+k)\}$ another realization of measurement noise, independent of $\Omega_k(t) = \{\eta(t-k), \ldots, \eta(t+k)\}$. The corresponding measurements will be denoted by $\widetilde{y}(t)$

$$\widetilde{y}(t+i) = s(t+i) + \widetilde{\eta}(t+i), \quad i \in I_k \ .$$

Following Akaike [30], one can adopt as an instantaneous measure of fit the final prediction error (FPE) statistic

$$\delta_{n|k}(t) = \mathrm{E}\{[\widetilde{y}(t) - \mathbf{f}^{\mathrm{T}}_{n|k}(0)\widehat{\boldsymbol{\beta}}_{n|k}(t)]^2\} \tag{10}$$

where averaging is carried out over $\widetilde{\Omega}_k(t)$ *and* $\Omega_k(t)$. According to (10) the quality of the local model is evaluated in terms of the mean squared prediction error (called final prediction error in [30]) observed when the model is used to predict another realization of the noisy signal, different from the one used for identification purposes.

If the hypothetical description $\mathcal{H}_{n|k}(t)$ holds true, one obtains

$$\begin{aligned} \Delta\widehat{\boldsymbol{\beta}}_{n|k}(t) &= \widehat{\boldsymbol{\beta}}_{n|k}(t) - \boldsymbol{\beta}_n = \mathbf{P}^{-1}_{n|k}\mathbf{r}_{n|k}(t) \\ \mathbf{r}_{n|k}(t) &= \sum_{i \in I_k} w_k(i)\mathbf{f}_{n|k}(i)\eta(t+i) \end{aligned}$$

and

$$\mathrm{E}[\mathbf{r}_{n|k}(t)\mathbf{r}^{\mathrm{T}}_{n|k}(t)] = \rho(t)\mathbf{R}_{n|k}. \tag{11}$$

This leads to

$$\text{cov}[\widehat{\boldsymbol{\beta}}_{n|k}(t)] = \text{E}\left[\Delta\widehat{\boldsymbol{\beta}}_{n|k}(t)\Delta\widehat{\boldsymbol{\beta}}^{\text{T}}_{n|k}(t)\right] = \rho(t)\mathbf{Q}_{n|k} \tag{12}$$

where

$$\begin{aligned}\mathbf{Q}_{n|k} &= \mathbf{P}^{-1}_{n|k}\mathbf{R}_{n|k}\mathbf{P}^{-1}_{n|k} \\ \mathbf{R}_{n|k} &= \sum_{i\in I_k} w_k^2(i)\mathbf{f}_{n|k}(i)\mathbf{f}^{\text{T}}_{n|k}(i) \ .\end{aligned}$$

Due to mutual independence of $\widetilde{\eta}(t)$ and $\Delta\widehat{\boldsymbol{\beta}}_{n|k}(t)$, one obtains

$$\begin{aligned}\delta_{n|k}(t) &= \text{E}\{[\widetilde{\eta}(t) - \mathbf{f}^{\text{T}}_{n|k}(0)\Delta\widehat{\boldsymbol{\beta}}_{n|k}(t)]^2\} \\ &= \text{E}[\widetilde{\eta}^2(t)] + \mathbf{f}^{\text{T}}_{n|k}(0)\text{cov}[\widehat{\boldsymbol{\beta}}_{n|k}(t)]\mathbf{f}_{n|k}(0) \\ &= \rho(t)[1 + q_{n|k}]\end{aligned} \tag{13}$$

where

$$q_{n|k} = \mathbf{f}^{\text{T}}_{n|k}(0)\mathbf{Q}_{n|k}\mathbf{f}_{n|k}(0) \ .$$

Consider the following estimate of the instantaneous noise variance

$$\begin{aligned}\widehat{\rho}_{n|k}(t) &= \frac{1}{L_k}\sum_{i\in I_k} w_k(i)\left[y(t+i) - \mathbf{f}^{\text{T}}_{n|k}(i)\widehat{\boldsymbol{\beta}}_{n|k}(t)\right]^2 \\ &= \frac{1}{L_k}\sum_{i\in I_k} w_k(i)\left[\eta(t+i) - \mathbf{f}^{\text{T}}_{n|k}(i)\Delta\widehat{\boldsymbol{\beta}}_{n|k}(t)\right]^2 \\ &= \frac{1}{L_k}\sum_{i\in I_k} w_k(i)\eta^2(t+i) - \frac{1}{L_k}\mathbf{r}^{\text{T}}_{n|k}(t)\mathbf{P}^{-1}_{n|k}\mathbf{r}_{n|k}(t)\end{aligned} \tag{14}$$

where $L_k = \sum_{i\in I_k} w_k(i)$ denotes the effective window width. Note that

$$\begin{aligned}\text{E}\left[\mathbf{r}^{\text{T}}_{n|k}(t)\mathbf{P}^{-1}_{n|k}\mathbf{r}_{n|k}(t)\right] &= \text{tr}\left\{\mathbf{P}^{-1}_{n|k}\text{E}[\mathbf{r}_{n|k}(t)\mathbf{r}^{\text{T}}_{n|k}(t)]\right\} \\ &= \rho(t)\text{tr}\{\mathbf{P}^{-1}_{n|k}\mathbf{R}_{n|k}\} \ .\end{aligned} \tag{15}$$

Combining (14) with (15) and denoting $v_{n|k} = \text{tr}\{\mathbf{P}^{-1}_{n|k}\mathbf{R}_{n|k}\}/L_k$, one obtains

$$\text{E}[\widehat{\rho}_{n|k}(t)] = \rho(t)\left[1 - v_{n|k}\right]$$

which leads to the following unbiased estimate of the final prediction error

$$\widehat{\delta}_{n|k}(t) = \frac{1 + q_{n|k}}{1 - v_{n|k}}\widehat{\rho}_{n|k}(t) \tag{16}$$

and the corresponding selection rule

$$\{\widehat{n}(t), \widehat{k}(t)\} = \arg\min_{\substack{k\in\mathcal{K}\\ n\in\mathcal{N}_k}} \widehat{\delta}_{n|k}(t) \ . \tag{17}$$

IV. Interpolation based approach

As an alternative to the final prediction error based approach to selection of n and k, we will consider a technique known as cross validation, based on evaluation of signal interpolation errors. Denote by

$$\varepsilon^{\circ}_{n|k}(t) = y(t) - \mathbf{f}^{\text{T}}_{n|k}(0)\widehat{\boldsymbol{\beta}}^{\circ}_{n|k}(t) \tag{18}$$

the leave-one-out signal interpolation error arising when the sample $y(t)$ is estimated based exclusively on the preceding $(\{y(i), i < t\})$ and succeeding $(\{y(i), i > t\})$ samples

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{\circ}_{n|k}(t) &= \arg\min_{\boldsymbol{\beta}_n} \sum_{\substack{i\in I_k\\ i\neq 0}} w_k(i)\left[y(t+i) - \mathbf{f}^{\text{T}}_{n|k}(i)\boldsymbol{\beta}_n\right]^2 \\ &= [\mathbf{P}^{\circ}_{n|k}]^{-1}\mathbf{p}^{\circ}_{n|k}(t)\end{aligned} \tag{19}$$

where

$$\begin{aligned}\mathbf{P}^{\circ}_{n|k} &= \sum_{\substack{i\in I_k\\ i\neq 0}} w_k(i)\mathbf{f}_{n|k}(i)\mathbf{f}^{\text{T}}_{n|k}(i) \\ \mathbf{p}^{\circ}_{n|k}(t) &= \sum_{\substack{i\in I_k\\ i\neq 0}} w_k(i)\mathbf{f}_{n|k}(i)y(t+i) \ .\end{aligned} \tag{20}$$

The estimates of n and k can be obtained using the following cross-validation (CV) based rule

$$\{\widehat{n}(t), \widehat{k}(t)\} = \arg\min_{\substack{k\in\mathcal{K}\\ n\in\mathcal{N}_k}} \widehat{\sigma}_{n|k}(t) \tag{21}$$

where $\widehat{\sigma}_{n|k}(t)$ denotes the local estimate of the variance of $\varepsilon^{\circ}_{n|k}(t)$, evaluated in the decision window of width $M = 2m + 1$ centered at t

$$\widehat{\sigma}_{n|k}(t) = \frac{1}{2m+1}\sum_{i=-m}^{m}[\varepsilon^{\circ}_{n|k}(t+i)]^2 \ . \tag{22}$$

To prevent the decision rule from behaving in an erratic way and, at the same time, retain its adaptivity, the recommended range of values for m is [20, 30].

The leave-one-out interpolation errors can be easily evaluated in terms of "regular" interpolation errors

$$\varepsilon_{n|k}(t) = y(t) - \widehat{s}_{n|k}(t) = y(t) - \mathbf{f}^{\text{T}}_{n|k}(0)\widehat{\boldsymbol{\beta}}_{n|k}(t) \ . \tag{23}$$

Actually, combining the formula

$$\begin{aligned}[\mathbf{P}^{\circ}_{n|k}]^{-1} &= \mathbf{P}^{-1}_{n|k} + \frac{\mathbf{P}^{-1}_{n|k}\mathbf{f}_{n|k}(0)\mathbf{f}^{\text{T}}_{n|k}(0)\mathbf{P}^{-1}_{n|k}}{1 - z_{n|k}} \\ z_{n|k} &= \mathbf{f}^{\text{T}}_{n|k}(0)\mathbf{P}^{-1}_{n|k}\mathbf{f}_{n|k}(0)\end{aligned}$$

obtained from (20) using the matrix inversion lemma [31] (note that $w_k(0) = 1$), with the formula

$$\mathbf{p}^{\circ}_{n|k}(t) = \mathbf{P}_{n|k}\widehat{\boldsymbol{\beta}}_{n|k}(t) - \mathbf{f}_{n|k}(0)y(t)$$

which stems from (20), one arrives at

$$\mathbf{f}^{\text{T}}_{n|k}(0)\widehat{\boldsymbol{\beta}}^{\circ}_{n|k}(t) = \frac{\mathbf{f}^{\text{T}}_{n|k}(0)\widehat{\boldsymbol{\beta}}_{n|k}(t) - z_{n|k}y(t)}{1 - z_{n|k}}$$

which leads to

$$\varepsilon^{\circ}_{n|k}(t) = \frac{\varepsilon_{n|k}(t)}{1 - z_{n|k}} \ . \tag{24}$$

Hence, evaluation of $\varepsilon^{\circ}_{n|k}(t)$ does not require implementation of the modified estimation scheme (19), which makes the proposed solution computationally attractive.

V. Computational aspects

First, observe that

$$\mathbf{f}^{\mathrm{T}}_{n|k}(i)\widehat{\boldsymbol{\beta}}_{n|k}(t) = \boldsymbol{\gamma}^{\mathrm{T}}_{n|k}(i)\mathbf{p}_{n|k}(t)$$

where the time-independent vectors $\boldsymbol{\gamma}_{n|k}(i) = \mathbf{P}^{-1}_{n|k}\mathbf{f}_{n|k}(i)$ can be precomputed and memorized for all values of n, k and i. Hence, neither computation of $\widehat{s}_{n|k}(t)$ nor computation of residual errors in (14) requires matrix inversion. Second, since the quantities $\mathbf{p}_{n|k}(t), n = 1, \ldots, N_k - 1$ are subvectors of $\mathbf{p}_{N_k|k}(t)$, the computational scheme is order-recursive. Finally, for the polynomial and harmonic bases and weighting sequences that belong to the cosinusoidal family, computations can be carried out in the time-recursive way. The simplest choice which guarantees this property is a cosinusoidal window of the form

$$w_k(i) = \cos[\pi i/(2k)] \ . \tag{25}$$

To derive the time-recursive computational formula, note that for the basis (3) it holds that

$$\mathbf{f}_{N_k|k}(i-1) = \mathbf{A}_{N_k|k}\mathbf{f}_{N_k|k}(i)$$

where $\mathbf{A}_{N_k|k} = [a_{ij|k}]_{N_k \times N_k}$ denotes the lower triangular matrix with elements

$$a_{ij|k} = \begin{cases} \binom{i-1}{i-j}/(-k)^{i-j} & \text{for} \quad i \geq j \\ 0 & \text{for} \quad i < j \end{cases} \ .$$

Note also that $w_k(i) = \mathrm{Re}\{u_k(i)\}$ where $u_k(i) = e^{j\frac{\pi i}{2k}}$ denotes the recursively computable complex-valued window, namely:

$$u_k(i-1) = \gamma_k u_k(i), \quad \gamma_k = e^{-j\frac{\pi}{2k}}. \tag{26}$$

Exploiting (25) and (26), one can compute the quantity $\mathbf{p}_{N_k|k}(t)$ recursively using the following algorithm

$$\begin{aligned} \mathbf{p}_{N_k|k}(t) &= \mathrm{Re}\{\mathbf{g}_{N_k|k}(t)\} \\ \mathbf{g}_{N_k|k}(t) &= \sum_{i \in I_k} u_k(i)\mathbf{f}_{N_k|k}(i)y(t+i) \\ = \gamma_k \mathbf{A}_{N_k|k} &\left[\mathbf{g}_{N_k|k}(t-1) - u_k(-k)\mathbf{f}_{N_k|k}(-k)y(t-k-1)\right] \\ &+ u_k(k)\mathbf{f}_{N_k|k}(k)y(t+k). \end{aligned}$$

The analogous recursive algorithm can be derived for the harmonic basis (4). Another bell-shaped window which in combination with the polynomial or harmonic basis allows for recursive computation of $\mathbf{p}_{N_k|k}(t)$ is the classical Hann, i.e., raised cosine window $w_k(i) = [1 + \cos(\pi i/k)]/2$.

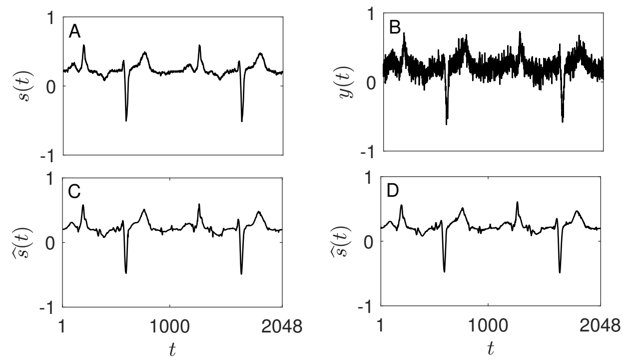

Figure 1: The fragment of measured ECG signal (A), the simulated noisy ECG signal (B) [white noise, SNR = 10 dB] and two results of denoising obtained with the proposed adaptive smoothing schemes with FPE-based (C) and CV-based (D) design parameters selection.

VI. Computer simulations

In our simulation experiment signal denoising was carried out for the ECG signal, sampled at 720 Hz, available from the PhysioBank online database [32]. Fig 1A shows the analyzed ECG signal of length 2048, covering two heartbeat cycles. Noisy ECG signals were generated by adding white Gaussian noise of different variances. Five input signal-to-noise-ratios ($\mathrm{SNR}_{\mathrm{in}}$) were considered: 5 dB, 10 dB, 15 dB, 20 dB, and 25 dB. Fig. 1B shows the simulated noisy ECG signal with $\mathrm{SNR}_{\mathrm{in}}$ = 10 dB.

Table 1 shows comparison of denoising results obtained for 9 nonadaptive Savitzky-Golay type smoothers corresponding to different choices of design parameters k (20, 40, 80) and n (1, 3, 5), and for the proposed adaptive smoothing schemes with FPE-based and CV-based selection of design parameters. In each scenario, corresponding to different $\mathrm{SNR}_{\mathrm{in}}$, the output SNRs were averaged over 100 realizations of the noisy ECG signal. In the CV-based approach the length of the decision window $M = 2m + 1$ was set to 51. The best results obtained for nonadaptive and adaptive approaches were shown in boldface. Because of the space limitations we present only the results obtained for the ECG test signal, the polynomial basis and the cosinusoidal window (25). More simulation results and the corresponding MATLAB codes can be found at the web page [33]. The proposed adaptive approaches yield either better or comparable results to those provided by the best smoothing algorithms with fixed settings. As shown in Figs. 1C and 1D, the results obtained using the FPE and CV criteria are very similar.

The proposed approaches were next compared with the wavelet shrinkage methods *VisuShrink* [10] and *BayesShrink* [11], known of their very good signal denoising capabilities. The corresponding MATLAB codes are available at the Stanford University web page [34]. We used typical settings – Daubechies 6 wavelet with a three-level decomposition. Table 2 shows comparison of denoising results for different $\mathrm{SNR}_{\mathrm{in}}$. Note that for low $\mathrm{SNR}_{\mathrm{in}}$ the FPE and the CV approaches yield better results than the wavelet methods. For high $\mathrm{SNR}_{\mathrm{in}}$ the results are comparable.

Table I: Comparison of denoising results of the noisy ECG signal (for 5 input SNRs) obtained for 9 nonadaptive Savitzky-Golay type smoothers corresponding to different choices of parameters k (20, 40, 80) and n (1, 3, 5), and for the proposed adaptive smoothing schemes with FPE-based and CV-based design parameters selection. The output SNRs were averaged over 100 realizations of noisy ECG signal.

	$\mathrm{SNR_{in}}$														
	5 dB			10 dB			15 dB			20 dB			25 dB		
$k\backslash n$	1	3	5	1	3	5	1	3	5	1	3	5	1	3	5
20	16.5	16.5	15.0	18.1	**21.0**	19.8	18.7	**24.2**	24.0	18.9	26.3	**27.3**	19.0	27.2	**29.1**
40	12.1	15.9	**17.2**	12.4	17.4	20.7	12.5	17.9	22.8	12.5	18.1	23.8	12.5	18.2	24.1
80	9.3	11.3	13.2	9.4	11.5	13.7	9.4	11.6	13.9	9.4	11.6	14.0	9.4	11.6	14.0
FPE/CV	17.2/**17.6**			21.4/**21.7**			24.7/**24.9**			27.1/**27.3**			28.7/**29.0**		

Table II: A comparison of denoising results of the ECG signal obtained for the proposed adaptive smoothing schemes (FPE, CV) and for the wavelet-based methods: *VisuShrink* with soft thresholding (ST) and hard thresholding (HT), and *BayesShrink* (BS).

	$\mathrm{SNR_{in}}$				
	5 dB	10 dB	15 dB	20 dB	25 dB
FPE	17.2	21.4	24.7	27.1	28.7
CV	**17.6**	**21.7**	**24.9**	**27.3**	29.0
ST	10.4	14.1	17.6	21.0	23.9
HT	14.8	19.0	22.3	25.3	27.6
BS	16.0	20.2	24.0	27.0	**29.1**

VII. Conclusion

The paper presents a new solution to design of data-adaptive Savitzky-Golay type smoothers. The proposed bandwidth and order adaptation mechanisms were based on cross-validatory analysis and generalized final prediction error statistic, respectively. The proposed adaptive approaches yield either better results or results that are comparable to those provided by the best smoothing algorithms with fixed settings. They also favorably compare with the state-of-the-art denoising algorithms based on the wavelet signal decomposition.

References

[1] F. Lewis, Optimal Estimation. New York: Wiley, 1986.

[2] J. Fan and I. Gijbels, "Adaptive order polynomial fitting: Bandwidth robustification and bias reduction," J. Comput. Graph. Statist., vol. 4, pp. 213–227, 1995.

[3] J. Fan and I. Gijbels, Local Polynomial Modeling and Its Applications. London: Chapman & Hall, 1996.

[4] W. Härdle, Applied Nonparametric Regression, Cambridge University Press, 1990.

[5] J.S. Simonoff, (1996), Smoothing Methods in Statistics, New York: Springer-Verlag, 1996.

[6] A.C. Bovik, T.S. Huang and D.C. Munson, "A generalization of median filtering using linear combinations of order statistics," IEEE Trans. Acoust., Speech, Signal Processing, vol. 31, pp. 1342–1350, 1983.

[7] Y.H. Lee and A.S. Kassam, "Generalized median filtering and related nonlinear filtering techniques," IEEE Trans. Acoust., Speech, Signal Processing, vol. 33, pp. 672-683, 1985.

[8] D.L. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," Biometrika, vol. 81, pp. 425–455, 1994.

[9] D.L Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," American Statistical Assoc., vol. 90, pp. 1200–1224, 1995.

[10] D.L. Donoho, "Denoising by soft thresholding," IEEE Trans. Inf. Theory, vol. 41, pp. 613–627, 1995.

[11] G. Chang, B. Yu and M. Vetterli, "Adapting wavelet thresholding for image denoising and compression," IEEE Trans. Image. Process., vol. 9, pp. 1532–1546, 2000.

[12] M. Niedźwiecki, "Easy recipes for cooperative smoothing," Automatica, vol. 46, pp. 716-720, 2010.

[13] M. Niedźwiecki and S. Gackowski, "Medley filters – simple tools for efficient signal smoothing," in Proc. 18th European Signal Processing Conference (EUSIPCO 2010), Aalborg, Denmark, 2010, pp. 641–645.

[14] M. Niedźwiecki and S. Gackowski, "On cooperative image denoising," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011), Prague, Czech Republic, 2011, pp. 933–936.

[15] V. Katkovnik, "A new method for varying adaptive bandwidth selection", IEEE Trans. Signal Process.,", vol. 47, pp. 2567–2571, 1999.

[16] A. Savitzky and M.J.E. Golay, "Smoothing and differentiation of data with simplified least squares procedures," Annals of Chemistry, vol. 36, pp. 1627–1639, 1964.

[17] R.W. Schafer, "What is a Savitzky-Golay filter?," IEEE Signal Process. Mag., vol. 28, pp. 111-117, 2011.

[18] B.C.H. Turton, "A novel variant of the Savitzky-Golay filter for spectroscopic applications," *Measurement Science Technology*, vol. 3, pp. 858–863, 1992.

[19] J. Li, H. Deng, P. Li, and B. Yu, "Real-time infrared detection based on adaptive Savitzky-Golay algorithm," Applied Physics B, vol. 120, pp. 207–216, 2015.

[20] M. Jakubowska and W.W. Kubiak, "Adaptive-degree polynomial filter for volumetic signals," Analytica Chimica, vol. 512, pp. 241–250, 2004.

[21] S. Hargittai, "Savitzky-Golay least-squares polynomial filters in ECG signal processing," in *Proc. Computers in Cardiology*, pp. 763–766, 2005.

[22] S. Goel, G. Kaur, and P. Tomar "ECG signal processing by Savitzky Golay least square polynomial filter," *International Journal of Advances in Electronics and Computer Science*, vol. 3, pp. 20–22, 2016

[23] D. Acharya, A. Rani, S. Agarwal, and V. Singh, "Application of adaptive Savitzky-Golay filter for EEG signal processing," Perspectives in Science, vol. 8, pp. 677–679, 2016.

[24] P. Barak, "Smoothing and differentiation by an adaptive-degree polynomial," *Anal. Chem.*, vol. 67, pp. 2758–2762, 1995.

[25] M. Browne, N. Mayer, and T.R.H. Cutmore, "A multiscale polynomial filter for adaptive smoothing," Digital Sign. Process., vol. 17, pp. 69–75, 2007.

[26] S.R. Krishnan and C.S. Seelamantula, "On the selection of optimum Savitzky-Golay filters," IEEE Trans. Signal Process.,", vol. 61, pp. 380–391, 2013.

[27] H. Mark and J. Workman, "Derivatives in spectroscopy: Part III – computing the derivative," *Spectroscopy*, vol. 18, pp. 106–111, 2003.

[28] J. Luo, J. Bai, P. He, and K. Ying, "Axial strain calculation using a low-pass digital differentiator in ultrasound elastography," *IEEE Transactions on Ultrasonics Ferroelectrics, and Frequency Control*, vol. 51, 2004.

[29] J. Karpińska, "Basic principles and analytical application of derivative spectrophotometry," in Macro to Nano Spectrometry, J. Uddin Ed., pp. 253–268, 2012.

[30] H. Akaike, "Fitting autoregressive models for prediction," Ann. Inst. Statist. Math., vol. 21, pp. 243–247, 1969.

[31] T. Söderström and P. Stoica, System Identification. Englewoods Cliffs NJ: Prentice-Hall, 1988.

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

[33] SG toolbox [Online]. Available at: https://eti.pg.edu.pl/katedra-systemow-automatyki/eusipco2019

[34] WAVELAB toolbox [Online]. Available at: http://statweb.stanford.edu/~wavelab/Wavelab_850/