MEDLEY FILTERS – SIMPLE TOOLS FOR EFFICIENT SIGNAL SMOOTHING

Maciej Niedźwiecki and Szymon Gackowski

Faculty of Electronics, Telecommunications and Computer Science
Department of Automatic Control, Gdańsk University of Technology
Narutowicza 11/12, Gdańsk , Poland
Tel: + 48 58 3472519, Fax: +48 58 3416132
e-mail: maciekn@eti.pg.gda.pl, szymon.gackowski@gmail.com

ABSTRACT

Medley filters are defined as convex combinations of elementary smoothing filters (averaging, median) with different smoothing bandwidths. It is shown that when adaptive weights of such a mixture are evaluated using the recently proposed Bayesian rules, one obtains a tool which often outperforms the state-of-the-art wavelet-based smoothing algorithms. Additionally, unlike wavelet-based procedures, medley filters can easily cope with non-Gaussian (Laplacian, uniform) and temporarily inhomogeneous measurement noise.

1. INTRODUCTION

Consider the problem of noncausal estimation of the signal $s(i)$ based on its noisy measurements $y(i)$:

$$y(i) = s(i) + v(i), \quad i = \dots, -1, 0, 1, \dots \tag{1}$$

where i denotes normalized time and $\{v(i)\}$ is the sequence of independent random variables obeying the generalized Gaussian law [1]

$$v \sim \mathscr{G}\mathscr{N}(\mu, \alpha, \beta):$$
$$p(v; \mu, \alpha, \beta) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left\{ -\left(\frac{|v-\mu|}{\alpha} \right)^{\beta} \right\} \tag{2}$$

where μ is the location parameter (we will assume that $\mu = 0$, i.e., that measurement noise is zero-mean), $\alpha > 0$ is the *unknown* scale parameter, $\beta \geq 1$ is the *known* shape parameter, and $\Gamma(\cdot)$ denotes Euler's gamma function.

Generalized Gaussian is a parametric family of symmetric distributions that includes the normal distribution when $\beta = 2$ (with mean μ and variance $\alpha^2/2$), and the Laplace distribution when $\beta = 1$ (with mean μ and variance $2\alpha^2$). When $\beta \to \infty$, the density (2) converges pointwise to a uniform density on $(\mu - \alpha, \mu + \alpha)$ (with mean μ and variance $\alpha^2/3$).

Distributions corresponding to $\beta \in [1, 2)$ are referred to as super-Gaussian or leptokurtic (with positive kurtosis). They have heavier tails than Gaussian distribution, i.e., they assign higher probabilities to extreme values. For this reason Laplace distribution is often used to model mixtures of wide-band noise and impulsive disturbances. When $\beta \in (2, \infty)$ the distribution (2) is called sub-Gaussian or platykurtic (with negative kurtosis). It has lighter tails than those of the normal distribution.

To simplify our further considerations, we will assume that an infinite observation history is available $\mathscr{Y} = \{y(i), i \in (-\infty, \infty)\}$. Note that for a given time instant t, $\mathscr{Y}$ can be decomposed into the set of "past" measurements $\mathscr{Y}_-(t) = \{y(i), i < t\}$, the "present" measurement $y(t)$, and the set of "future" measurements $\mathscr{Y}_+(t) = \{y(i), i > t\}$:

$$\mathscr{Y} = \{\mathscr{Y}_-(t), y(t), \mathscr{Y}_+(t)\}.$$

Our objective will be to find the estimate $\widehat{s}(t) = f[t, \mathscr{Y}]$ that minimizes the mean-squared error

$$\mathrm{E}\left\{ [s(t) - \widehat{s}(t)]^2 \right\} \to \min. \tag{3}$$

It is well known [2] that the optimal, in the mean-square sense, noncausal estimator of $s(t)$ has the form

$$\widehat{s}(t) = \mathrm{E}[s(t)|\mathscr{Y}]. \tag{4}$$

When the estimated signal admits a known state-space description, and when both the signal $s(t)$ and measurement noise $v(t)$ are normally distributed, the conditional mean estimate (4) can be computed using the celebrated Kalman smoother. However, even though mathematically well founded and statistically efficient (under assumptions mentioned above), Kalman smoothers have limited practical applicability. In practice one needs algorithms which are much less demanding in terms of the required prior knowledge about the recovered signal, and which are capable of adapting to the unknown and/or time-varying degree of signal smoothness and noise intensity. To fulfill this demand, several powerful *universal smoothing* schemes were proposed, i.e., schemes that require no, or very little, prior knowledge of signal/noise characteristics. The best-known examples of universal smoothers are those based on kernel regression [3], [4], order statistic filtering [5], [6], and wavelet thresholding (shrinkage) [7], [8], [9].

In this paper we present a very simple multiresolution smoothing procedure combining outputs of several nonlinear median filters, and several linear averaging filters. We show that, when appropriately designed, the resulting *medley* filter often outperforms, on a set of benchmarks signals, the state-of-the-art wavelet-based procedures known for their excellent smoothing capabilities. On the qualitative level, we continue research on, increasingly popular, combination schemes, where the outputs of several filters are mixed together to obtain an overall output of improved quality [10].

2. MEDLEY FILTERS

Denote by $\mathscr{F} = \{\widehat{s}_k(t), k = 1, \dots, K\}$ the family of elementary smoothers

$$\widehat{s}_k(t) = f_k[t, \mathscr{Y}] \tag{5}$$

This work was partially supported by the Foundation for the Polish Science.

with different smoothing bandwidths. The medley filter supported by $\mathscr{F}$ will be defined in the form

$$\widehat{s}(t) = \sum_{k=1}^{K} \mu_k(t)\widehat{s}_k(t) \tag{6}$$

where $\mu_k(t)$ are data-dependent weights, further called credibility coefficients, that obey

$$\sum_{k=1}^{K} \mu_k(t) = 1, \quad \mu_k(t) \geq 0, \; k = 1, \ldots, K, \; \forall t$$

Hence, $\widehat{s}(t)$ is a convex combination of component smoothers.

2.1 Selection of Component Smoothers

As already noted in [11], a robust medley filter can be obtained by combining probably the simplest smoothers used in signal processing: linear averaging filters

$$\widehat{s}_k(t) = \frac{\sum_{i=-n_k}^{n_k} y(t+i)}{2n_k+1} \tag{7}$$

and nonlinear median filters[1]

$$\widehat{s}_k(t) = \text{med}\{y(t-n_k), \ldots, y(t+n_k)\} \tag{8}$$

with local fitting frames $F_k(t) = [t-n_k, t+n_k]$ of different lengths $N_k = 2n_k+1$. Averaging filters efficiently remove Gaussian noise from lowpass signals. Two remarkable features of median filters are their ability to preserve step-like signal features and their resistance to outliers – this makes them an attractive choice for smoothing discontinuous signals and/or signals corrupted by impulsive noise. Both filters have computationally efficient recursive implementations.

In Section 4 we will show that by combining averaging filters and median filters, one obtains quite powerful medley filter, often outperforming wavelet-based solutions. We note that averaging filters and median filters are the simplest representatives of two more general classes of smoothing filters, usually referred to as kernel smoothers [3] and order statistic filters [5], respectively. Generally, good medley filters can be obtained by combining tools from both classes mentioned above.

2.2 Evaluation of Credibility Coefficients

Following [11], credibility coefficients will be evaluated using the following formula

$$\mu_k(t) \propto \left[\sum_{i \in T(t)} |\varepsilon_k^\circ(i)|^\beta \right]^{-M/\beta} \tag{9}$$

where $T(t) = [t-m, t+m]$ denotes the local evaluation frame of lenght $M = 2m+1$ and

$$\varepsilon_k^\circ(i) = y(i) - \widehat{s}_k^\circ(i), \;\; i \in T(t)$$

denote matching errors, i.e., residual errors evaluated for the holey smoother associated with $\widehat{s}_k(t)$. Holey smoother $\widehat{s}_k^\circ(t)$ is identical with $\widehat{s}_k(t)$, except that it excludes the "central" sample $y(t)$ from the set of measurements used for estimation of $s(t)$

$$\widehat{s}_k^\circ(t) = f_k[t, \mathscr{Y}^\circ(t)], \;\; \mathscr{Y}^\circ(t) = \mathscr{Y} - \{y(t)\} \tag{10}$$

In the uniform noise case ($\beta \to \infty$), one should set

$$\mu_k(t) \propto \left[\max_{i \in T(t)} |\varepsilon_k^\circ(i)| \right]^{-M}$$

For many nonlinear smoothing algorithms, including median filters, holey smoothers are either ill-defined or they do not preserve important properties of the original scheme. In cases like this, holey smoothers can be substituted with patched smoothers, obtained by replacing the central sample $y(t)$ with the signal estimate $\widehat{s}_k(t)$, rather than by leaving $y(t)$ out

$$\widehat{s}_k^\bullet(t) = f_k[t, \mathscr{Y}^\bullet(t)], \;\; \mathscr{Y}^\bullet(t) = \mathscr{Y}\,\big|_{y(t):=\hat{s}_k(t)} \tag{11}$$

Similar to the holey smoother, the patched smoother can be used for the purpose of local evaluation of $\widehat{s}_k(\cdot)$. This can be accomplished by replacing matching errors $\varepsilon_k^\circ(i)$ in (9) with the modified matching errors

$$\varepsilon_k^\bullet(i) = y(i) - \widehat{s}_k^\bullet(i) \; .$$

Evaluation of smoothers in terms of the corresponding matching errors is consistent with the long-standing statistical approach known as leave-one-out cross-validatory analysis [12]. Similarly, the modified matching errors are the cornerstone of the so-called full cross-validatory analysis, proposed by Bunke at al. [13]. We note, however, that cross-validation results in competitive, rather than cooperative smoothing schemes (winner-takes-all strategy). Our Bayesian framework brings the notion of filter credibility into cross-validatory analysis.

2.3 Computational Hints

One can easily check that for the linear averaging filter (7) it holds that

$$\begin{aligned} \varepsilon_k^\circ(i) &= \delta_k \varepsilon_k(i), \quad \delta_k = n_k/(n_k-1) \\ \varepsilon_k^\bullet(i) &= \rho_k \varepsilon_k(i), \quad \rho_k = (n_k+1)/n_k \end{aligned} \tag{12}$$

where

$$\varepsilon_k(i) = y(i) - \widehat{s}_k(i)$$

denotes residual error. This means that both errors can be computed without actually implementing the corresponding holey/patched smoothers.

Even though the relationships (12) do not extend to nonlinear filters, we have observed that they usually yield good approximations when applied to validation of median filters.

Some of the quantities involved in computation of credibility coefficients $\mu_k(t)$ may take very large or very small values. The following modified expression, mathematically equivalent to (9), allows one to avoid numerical problems (such as numerical overflow) caused by improper scaling

$$\mu_k(t) = \frac{\exp\{\chi_k(t)\}}{\sum_{k=1}^{K} \exp\{\chi_k(t)\}} \tag{13}$$

[1] $\text{med}\{\cdot\}$ denotes the central value of the ordered sequence of samples

where

$$\begin{aligned}\chi_k(t) &= \psi_k(t) - \psi_{\max}(t)\\ \psi_k(t) &= -(M/\beta)\log r_k(t)\\ \psi_{\max}(t) &= \max_{1\le k\le K} \psi_k(t)\ .\end{aligned}$$

and

$$r_k(t) = \sum_{i\in T(t)} |\varepsilon_k^\circ(i)|^\beta .$$

3. QUALITATIVE COMPARISON WITH WAVELET THRESHOLDING

Since the wavelet thresholding (shrinkage) approach has the reputation for being one of the most efficient denoising tools, we will use it as a benchmark for both qualitative and quantitative evaluation of the proposed approach. Quantitative results will be presented in section 4. In this section we will focus on some qualitative features of medley filters which, at least in some applications, make them an interesting alternative to wavelet-based smoothers:

1. The basic rules of wavelet thresholding were derived under the assumption that the additive measurement noise is Gaussian. In contrast with this, medley filters can be trimmed to the (known) distribution of noise, including practically important non-Gaussian cases such as Laplacian distribution and uniform distribution.
2. While procedures based on wavelet thresholding are block-oriented, i.e., they can be used for fixed-interval smoothing only (suitable for off-line applications), medley filters are fixed-lag smoothers and, as such, they can be used in near real-time applications, where a constant decision delay of τ sampling intervals is tolerable. Note that for the smoother combining averaging filters (7) and median filters (8) such a decision delay, or lag, is given by $\tau = \max\{m+1, n_k+1, k=1,\ldots,K\}$.
3. Due to global thresholding, all wavelet-based denoising procedures fail to work correctly when noise variance changes across the analysis frame. Since the medley filter is a local smoothing algorithm, which does not require information about the local noise variance (note that the credibility coefficients do not depend on the scale parameter α), it can easily cope with temporally inhomogeneous noise.

4. QUANTITATIVE COMPARISON WITH WAVELET THRESHOLDING

The four test signals used in our simulation experiments were proposed by Donoho and Johnstone [8], and are called *Blocks*, *Bumps*, *Doppler* and *HeaviSine*, respectively (see Fig. 1). They exemplify different forms of spatial and temporal inhomogeneity encountered in many real-world signals. Since it is agreeably difficult to design a smoothing algorithm that copes favorably with *all four* signals, they constitute a demanding testbed, commonly used for benchmarking purposes.

Test signals, each containing 2048 samples, were extended by zeros at both ends (to avoid boundary problems) and corrupted with Gaussian ($\beta = 2$), Laplacian ($\beta = 1$), or uniform ($\beta = \infty$) white noise with intensity varying from $\sigma_v^2 = 0.01$ to $\sigma_v^2 = 25$. We note that for a fixed value of σ_v the average signal-to-noise ratio (SNR) is for all test signals the same, e.g. for $\sigma_v^2 = 1$ it is equal to 16.9 dB.

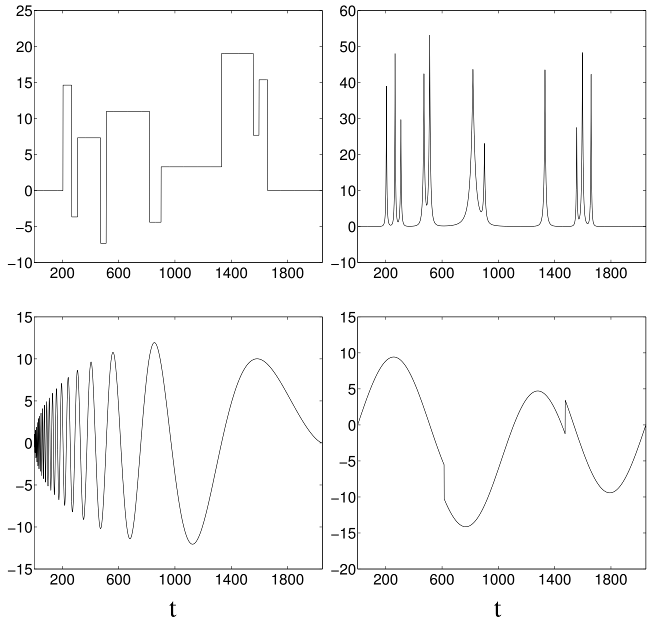

Figure 1: Test signals: *Blocks* (top left), *Bumps* (top right), *Doppler* (bottom left), and *HeaviSine* (bottom right).

The bank of component filters consisted of five averaging filters and five median filters, with fitting frames of lenghts $N_k = 2n_k + 1, k = 1,\ldots,5$, forming (approximately) a geometric progression: $N_1 = 5$, $N_2 = 11$, $N_3 = 23$, $N_4 = 47$, $N_5 = 95$. The width M of the evaluation frame was set equal 31 ($m = 15$).

Fig. 2 shows the performance comparison between the medley filter and its component filters. The plots show dependence of the average MSE scores on standard deviation of an additive Gaussian noise. All results were obtained by ensemble averaging over 100 realizations of measurement noise. Note that the medley filter almost always works better than component filters.

Tables 1, 2, and 3 show the comparison of the medley filter with the state-of-the-art wavelet thresholding procedures *VisuShrink* [8] and *BayesShrink* [9] (for the Daubechies D6 basis). Despite its simplicity, in a majority of cases the proposed scheme yields better results than procedures based on wavelet thresholding. Wavelet-based procedures work better when the average SNR is large (> 30 dB). Since wavelets constitute a complete set of basis functions (i.e., every finite-length sequence has an exact representation in the wavelength domain), this high-SNR advantage of wavelet-based methods is expected and practically impossible to beat.

Finally, Figs. 3–5 illustrate the inhomogeneous noise experiment. Test signals were corrupted by Gaussian noise, the standard deviation of which linearly grew from 0.5 to 2 along the time axis (Fig. 3). Results of smoothing, presented in Fig. 4, confirm that medley filter can successfully adapt to changing conditions – the corresponding MSE scores were equal to 0.1564, 0.7118, 0.1176 and 0.0778 for *Blocks*, *Bumps*, *Doppler* and *HeaviSine*, respectively. Fig. 6 shows the analogous results obtained using the *BayesShrink* procedure (which worked better than *VisuShrink*). As ex-

σ_n	Test signal	VisuShrink	BayesShrink	Medley
0.1	Blocks	0.0019	0.0083	**0.0013**
	Bumps	**0.0038**	0.0084	0.5502
	Doppler	**0.0033**	0.0049	0.0195
	HeaviSine	**0.0008**	0.0028	0.0011
0.5	Blocks	0.0611	0.1057	**0.0269**
	Bumps	**0.0681**	0.1194	0.5804
	Doppler	0.0572	0.0627	**0.0529**
	HeaviSine	0.0247	0.0241	**0.0167**
1.0	Blocks	0.2497	0.2595	**0.1045**
	Bumps	**0.2870**	0.3546	0.6593
	Doppler	0.1832	0.1813	**0.1240**
	HeaviSine	0.0812	0.0602	**0.0589**
2.0	Blocks	0.9776	0.7941	**0.4095**
	Bumps	1.1406	1.0186	**0.9467**
	Doppler	0.6283	0.4795	**0.3521**
	HeaviSine	0.2085	0.1478	**0.2046**
5.0	Blocks	4.2541	2.4839	**2.1203**
	Bumps	5.5822	3.7826	**2.7255**
	Doppler	2.2825	1.6912	**1.6480**
	HeaviSine	0.8924	**0.4817**	1.1219

Table 1: Performance comparison (MSE) of the medley filter with two denoising procedures based on wavelet thresholding: *VisuShrink* and *BayesShrink*. Test signals were corrupted by Gaussian noise with standard deviation σ_n.

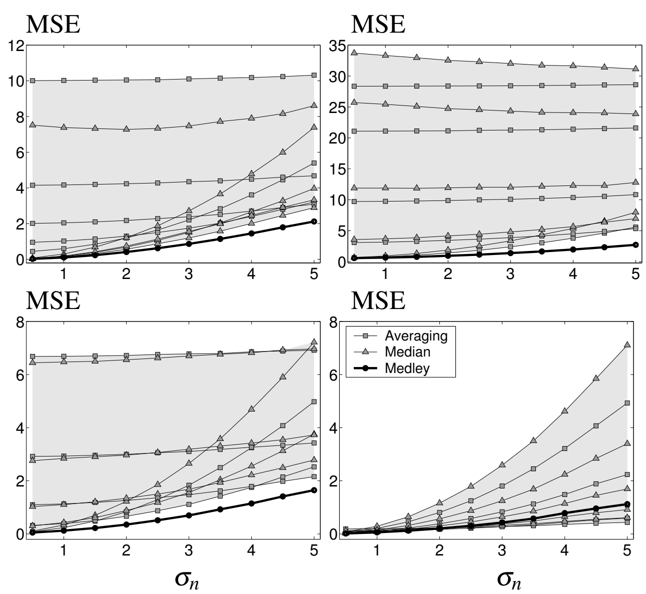

Figure 2: Comparison of MSE errors yielded by component averaging filters (thin lines marked with squares), component median filters (thin lines marked with triangles) and medley filter (thick line marked with circles). Results are shown for four test signals: *Blocks* (top left), *Bumps* (top right), *Doppler* (bottom left), and *HeaviSine* (bottom right), corrupted by Gaussian noise with standard deviation σ_n.

pected, due to global thresholding, in three out of four cases wavelet-based procedures work considerably worse than the medley filter (the corresponding MSE scores were equal to 0.5186, 0.5091, 0.3519 and 0.1056).

σ_n	Test signal	VisuShrink	BayesShrink	Medley
0.1	Blocks	0.0021	0.0087	**0.0007**
	Bumps	**0.0037**	0.0088	0.5800
	Doppler	**0.0033**	0.0057	0.0193
	HeaviSine	0.0012	0.0042	**0.0008**
0.5	Blocks	0.0631	0.1374	**0.0173**
	Bumps	**0.0699**	0.1486	0.6121
	Doppler	0.0625	0.0853	**0.0489**
	HeaviSine	0.0321	0.0427	**0.0121**
1.0	Blocks	0.2622	0.3819	**0.0681**
	Bumps	**0.2807**	0.4435	0.6796
	Doppler	0.2106	0.2568	**0.1086**
	HeaviSine	0.1282	0.1408	**0.0410**
2.0	Blocks	0.9912	1.0119	**0.2813**
	Bumps	1.1481	1.2952	**0.8823**
	Doppler	0.7642	0.7925	**0.2834**
	HeaviSine	0.4310	0.4657	**0.1345**
5.0	Blocks	5.0145	4.0721	**1.4481**
	Bumps	6.0569	5.7126	**2.2950**
	Doppler	3.5126	3.6715	**1.2266**
	HeaviSine	2.2075	2.4523	**0.7349**

Table 2: Performance comparison (MSE) of the medley filter with two denoising procedures based on wavelet thresholding: *VisuShrink* and *BayesShrink*. Test signals were corrupted by Laplacian noise with standard deviation σ_n.

σ_n	Test signal	VisuShrink	BayesShrink	Medley
0.1	Blocks	**0.0021**	0.0079	0.3587
	Bumps	**0.0041**	0.0080	0.5519
	Doppler	**0.0037**	0.0049	0.0220
	HeaviSine	**0.0008**	0.0022	0.0133
0.5	Blocks	**0.0677**	0.0856	0.3514
	Bumps	**0.0770**	0.1054	0.5765
	Doppler	0.0648	0.0573	**0.0494**
	HeaviSine	0.0278	0.0251	**0.0199**
1.0	Blocks	**0.2807**	0.2812	0.3898
	Bumps	**0.3250**	0.3884	0.6476
	Doppler	0.2079	0.1860	**0.1071**
	HeaviSine	0.0858	0.0645	**0.0452**
2.0	Blocks	1.1068	0.8653	**0.6237**
	Bumps	1.2912	0.9808	**0.9032**
	Doppler	0.6861	0.4887	**0.2936**
	HeaviSine	0.2218	0.1577	**0.1234**
5.0	Blocks	4.6546	2.5623	**2.0343**
	Bumps	6.3941	3.6510	**2.4581**
	Doppler	2.4896	1.6804	**1.2102**
	HeaviSine	1.0222	**0.4508**	0.5698

Table 3: Performance comparison (MSE) of the medley filter with two denoising procedures based on wavelet thresholding: *VisuShrink* and *BayesShrink*. Test signals were corrupted by uniform noise with standard deviation σ_n.

5. CONCLUSION

We have shown that by combining several linear averaging filters and several nonlinear median filters, one obtains a powerful smoothing algorithm, called medley filter. Under medium and low SNR conditions, medley filter outperforms wavelet-based procedures, known of their very good

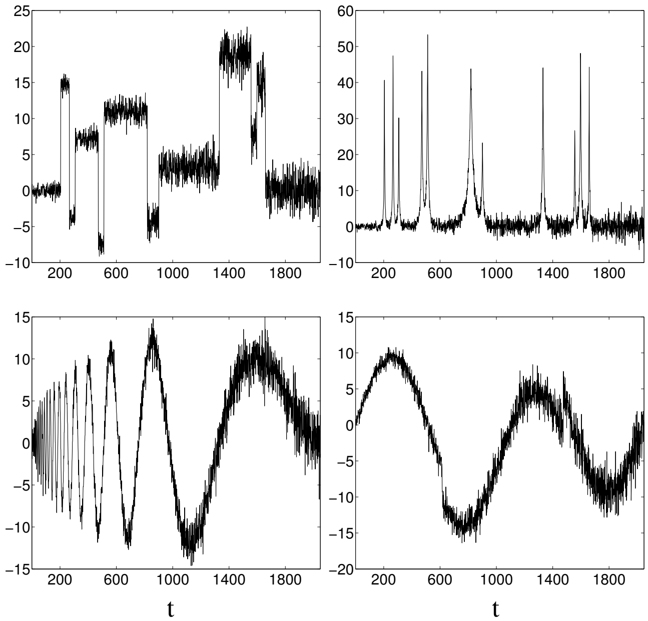

Figure 3: Test signals corrupted by white inhomogeneous Gaussian noise.

smoothing properties. Medley filter accounts for distribution of measurement noise and deals favorably with inhomogeneous noise. Additionally, unlike wavelet-based procedures, it can be used as a fixed-lag smoother.

REFERENCES

[1] N. Saralees, “A generalized normal distribution,” *J. Appl. Stat.*, vol. 32, pp. 685–694, 2005.

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

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

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

[5] 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.

[6] Y.H. Lee and A.S. Kassam, “Generalized median filtering and related nonlinear filtering techniques,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 672-83, 1985.

[7] D. Donoho and I. Johnstone, “Ideal spatial adaptation via wavelet shrinkage,” *Biometrika*, vol. 81, pp. 425–455, 1994.

[8] D. Donoho and I. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *American Statistical Assoc.*, vol. 90, pp. 1200–1224, 1995.

[9] 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.

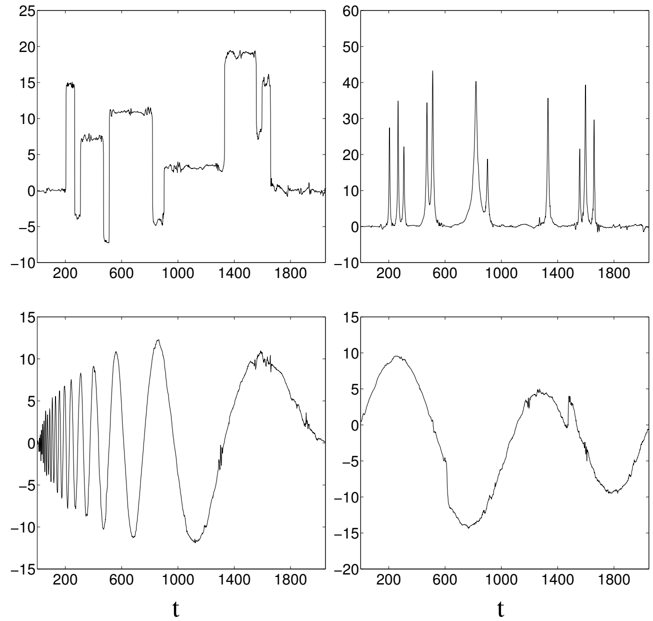

Figure 4: Test signals denoised using the medley filter.

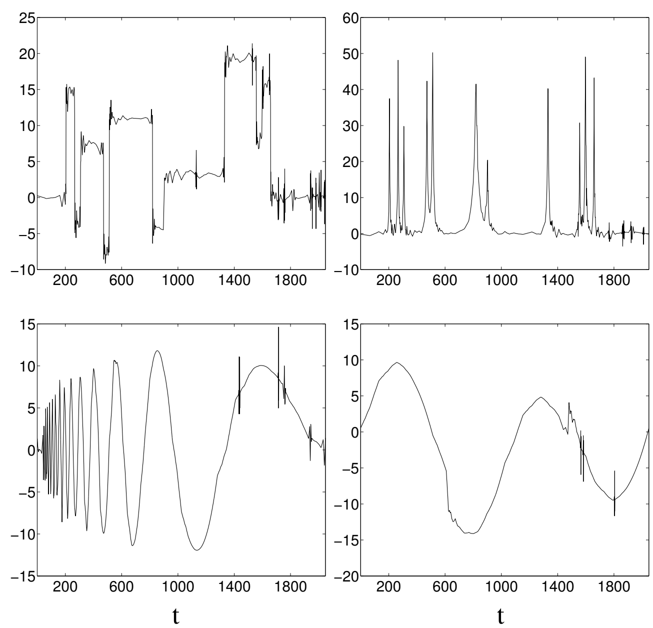

Figure 5: Test signals denoised using the *VisuShrink* procedure.

[10] Y. Yang, “Combining different procedures for adaptive regression,” *J. Multiv. Anal.*, vol. 74, pp. 973–950, 2000.

[11] M. Niedźwiecki, “Easy recipes for cooperative smoothing,” *Automatica*, vol. 46, pp. 716–720, 2010.

[12] H. Friedl and E. Stampfer, “Cross-validation,” in *Encyclopedia of Environmetrics*, A.H. El-Shaarawi & W.W. Piegorsch, Eds., vol. 1, pp. 452–460, New York: Wiley, 2002.

[13] O. Bunke, B. Droge and J. Polzehl, “Model selection, transformations and variance estimation in nonlinear regression,” *Statistics*, vol. 33, pp. 197–240, 1999.