Signal denoising by constraining the residual to be statistically noise-similar

Lorina Dascal, Michael Zibulevsky and Ron Kimmel
Department of Computer Science, Technion, Haifa, Israel
email:{lorina,mzib,ron}@cs.technion.ac.il

January 6, 2008

Abstract

We propose to solve the signal/image denoising problem by minimizing the total variation of the signal and forcing the residual between the estimated and the measured signal to be statistically noise-similar; we thus project the residual on a set of directions in signal space and require these projections to be 4σ limited and to have appropriate empirical nonlinear moments. Experimental results on denoising 1-D signals demonstrate the efficiency of the method.

1 Introduction and previous works

The purpose of signal denoising is to recover an unknown signal x from its noisy observation y

$$y = x + \xi, \tag{1}$$

where ξ is a white noise with known statistics.

In recent years, different approaches have been studied for solving reconstruction problems of type (1), among them non-parametric statistics methods, the variational approach and the wavelet-like thresholding type approach. In [8] a non-parametric statistics method was applied to noisy observations of images and time-dependent signals. The methods based on variational approach are often related to the total variation model, introduced by Rudin, Osher and Fatemi [10]. At the other end, wavelet soft thresholding methods have been introduced in [3]. These techniques have been extended in different papers [4], [5], [1]. Combining variational techniques and thresholding techniques can be seen in several works see [2], [12].

2 The proposed model

The main idea of the proposed model relies on the following observation: given ξ a random vector with normally distributed $N(0, \sigma^2)$ independent entries then, for all vectors $w : ||w||_2 = 1$, the projection $\langle \xi, w \rangle$ is also normally distributed with the same variance[1]

$$\langle \xi, w \rangle \sim N(0, \sigma^2).$$

First we introduce some notations. We denote by $J(x)$ the discretization of the total variation (TV) of the signal x:

$$J(x) = \sum_{l=1}^{N-1} |(\nabla x)_l|,$$

where N is the number of samples in x and

$$(\nabla x)_l = x_{l+1} - x_l, \quad l = 1, 2, ..., N-1.$$

Keeping in mind that for a good reconstruction, we would like the residual to be as close as possible to noise, and using the above observation, we propose the following constrained minimization

$$\min_x \{ \sum_{l=1}^{N-1} |(\nabla x)_l| \} \tag{2}$$

$$\text{Subject to:} \quad |\langle y - x, w_i \rangle| \leq 4\sigma, \quad \forall w_i \in D, \quad i = 1, ..., T. \tag{3}$$

where D is a given set (dictionary) of windows $w_i : ||w_i|| = 1$, T is the total number of windows and the above scalar product is the Euclidean scalar product between the residual vector and a window vector.

[1] If noise is not white, then

$$Var\langle \xi, w \rangle = \int_\Omega H(\omega) |\bar{w}|^2(\omega) d\omega,$$

where H is noise power spectrum, $\bar{w}$ - Fourier transform of w .

In addition to the demand that these projections would be 4σ limited, we also require some of their empirical moments to be close to those of $N(0, \sigma^2)$:

$$E_T\{\psi_k(\langle y-x, w_i\rangle)\} \equiv \frac{1}{T}\sum_{i=1}^{T}\psi_k(\langle y-x, w_i\rangle) \leq E\{\psi_k(N(0,\sigma^2))\} + \Delta_k, \quad (4)$$

where some reasonable choices of the non-linear functions $\psi_k(t)$ may be (but not limited to) $\psi_k(t) = |t|^{p_k}$, with exponents $p_k = 2, 3, ...$, but also with $1 < p_k < 2$.

The value Δ_k should be chosen from the requirement

$$\mathrm{Prob}\Big(E_T\{\psi_k(\langle \xi, w_i\rangle)\} > E\{\psi_k(N(0,\sigma^2))\} + \Delta_k\Big) < \epsilon,$$

where ξ is a random vector following the noise distribution; $\epsilon = 10^{-3} - 10^{-8}$.

Note that for non-overlapping windows, the value of Δ_k is easy to compute. For overlapping windows, it can be determined empirically.

In our simulations, the dictionary D was a set of one-dimensional windows of width $1, 2, 4, 8, ..$, with all possible shifts. Other choices for D are also possible, e.g. wavelet bases, wavelet packet bases, etc. (see for example [6]).

A model close to (2), (3) was first proposed and analyzed in [9](see also [8]). Later (and independently) similar models were designed and studied in the context of image decompression [2]. Problem (2), (3) was also explored in [1], [6] and [7].

The classical TV minimization model proposed by Rudin, Osher, Fatemi [10] is given by

$$\min_{x\in BV(\Omega)} \int_\Omega |\nabla x|\, dx,$$

subject to

$$||y-x||_2^2 = \sigma^2,$$

and it will be abbreviated as the ROF model.

The hypothesis on the functional space of functions used the ROF model is that functions of bounded variations (the BV space model) are a reasonable functional model for many problems in image processing, in particular for restoration problems. Here we also adopt the BV space as the natural space of functions needed in the reconstruction. Functions of bounded variations have discontinuities along rectifiable curves, being continuous in some sense (in the measure theoretic sense) away from the discontinuities.

In order to solve the constrained optimization problem (2), (3), (4) we use the penalty technique. We mention that in our simulations we limit ourselves to second order moments, $E_T\{\langle w_i, y-x\rangle^2\}$ in order to shape gaussianity of the residual projections, but as mentioned above, other nonlinear moments can also be employed. We therefore solve Problem (2), (3) with the additional constraints

$$E_T\{\langle w_i, y-x\rangle^2\} \approx \sigma^2.$$

This can be solved by the following unconstrained minimization problem:

$$\arg\min_x \{J(x) + \mu_1 \sum_{i=1}^{T} \phi_{4\sigma}(\langle w_i, y-x\rangle) + \sum_{i=1}^{T} \mu_2 \langle w_i, y-x\rangle^2\}. \qquad (5)$$

Here the parameter μ_1 is chosen to be large and the parameter μ_2 is tuned such that

$$E_T\{\langle w_i, y-x\rangle^2\} \approx \sigma^2.$$

The penalty function ϕ_τ used is the following linear dead-zone function:

$$\phi_\tau(x) = \max\{0, |x| - \tau\}. \qquad (6)$$

To improve convergence of the optimization procedure, we use a smoothed version of the function Φ_τ as seen in the Figure 1. We notice that we can write the function ϕ_τ as

$$\phi_\tau(x) = \frac{1}{2}(\eta(x-\tau) + \eta(x+\tau)) - \tau, \qquad (7)$$

where $\eta(x) = |x|$.

In order to smooth out the piecewise smooth function ϕ_τ, we use (7) and a smoother for $\eta(x)$, namely

$$|x| \approx \sqrt{x^2 + \beta^2},$$

where β is a small positive number.

The smoothed version we employed is

$$\phi_\tau^\beta(x) = \frac{1}{2}\Big(\sqrt{(x-\tau)^2 + \beta^2} + \sqrt{(x+\tau)^2 + \beta^2}\Big) - \tau,$$

where the parameter β was typically $\beta = 10^{-3}$.

We remark that the function we consider in our optimization is more stable to outliners than the quadratic convex function chosen in [6].

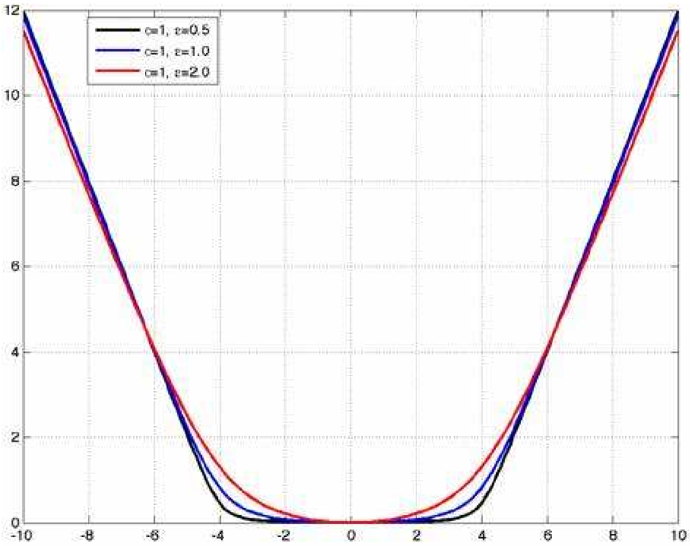

Figure 1: Smoothed penalty function used in the minimization.

Another possibility to better shaping the Gaussianity of the residual is to use more general moments, such as non-quadratic moments, $E_T\{|\langle w_i, y - x\rangle|^p\}$, $p > 0$ in addition to the quadratic term given in (5).

3 Experimental results

In the experiments we consider a signal composed of two pulses and a ramp, corrupted with noise of standard deviation $\sigma = 0.9$. We used the algorithm based on rectangular normalized 1-D windows and compared its performance to the classical ROF model. The windows we chose are 1-D normalized overlapped windows of lengths $1, 2, 4, 8, 16$ and the parameters used are $\mu_1 =$

1000 and $\mu_2 = 0.02$. Optimization of the functional (5) was performed by the gradient descent.

The best way to see the subtle differences between the methods is to look at the residual. A residual that contains less structure and looks more like random noise is an indication of a better reconstruction. In Fig. 4 we show the residual for each algorithm. Notice that for the projections-based algorithm, the residual looks more like noise, rather than in the case of classical total variation (ROF) model.

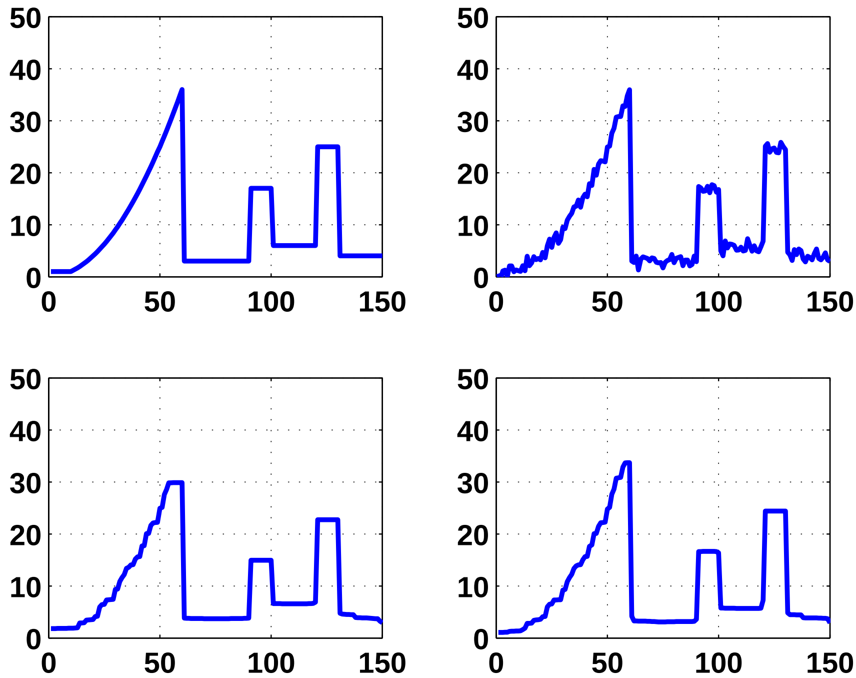

Figure 2: Top left: original signal. Top right: noisy signal ($\sigma = 0.9$, PSNR=25.32 db). Bottom left : ROF reconstruction. (PSNR=27.09 db). Bottom right: Signal reconstructed using windows ($\mu_1 = 1000$, $\mu_2 = 0.02$, PSNR=28.89 db).

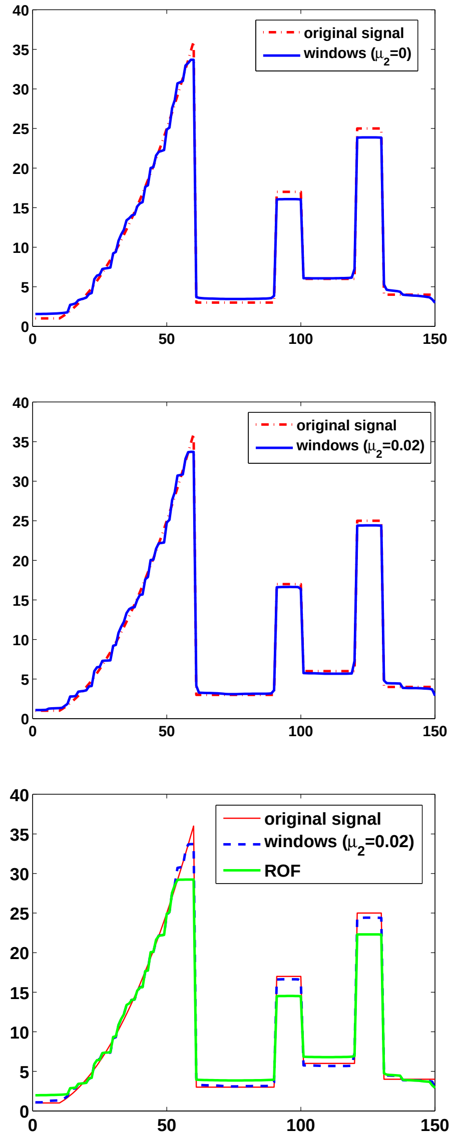

Figure 3: Top: Original signal and reconstructed signal based on projections without using the second order moment ($\mu_1 = 1000$, $\mu_2 = 0$, PSNR=28.09 db). Middle: Original signal and reconstructed signal based on projections with the second order moment ($\mu_1 = 1000$, $\mu_2 = 0.02$). Bottom: Comparison between the original signal, reconstructed signal based on ROF model and reconstructed signal based on projections using the second order moment.

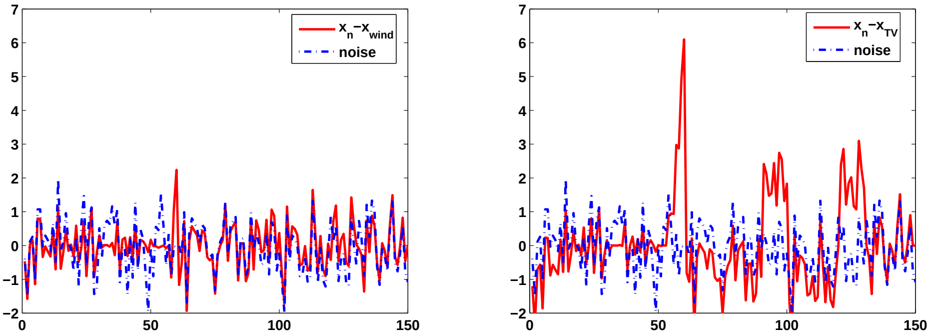

Figure 4: Left: Comparison between noise and residual for the algorithm based on projections using second order moment ($\mu_1 = 1000, \mu_2 = 0.02$). Right: Comparison between noise and residual for the ROF algorithm. Notice that the residual in the left better captures the behaviour of the noise compared to the one on the right.

4 Conclusions

We proposed an improvement of the classical ROF Total Variation model, by combining total variation minimization with convex constraint terms based on residual projection which forces the difference between the measured and the estimated signal to be noise-similar. Experiments on denoising of 1-D signals using rectangular windows demonstrate the efficiency of this method. The dictionary made of rectangular windows is simple and can certainly be replaced by a dictionary made of other kind of elements which would better capture the structure of the signals. We believe that using a dictionary made of wavelets or wavelets packets for 1-D signals, or curvelets/ridgelets for 2-D signals could significantly improve the results.

References

[1] E. J. *Candès* and F. Guo, "New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction", *Signal Processing*, 82, 1519-1543, 2001.

[2] R. R. Coifman and A. Sowa,“Combining the calculus of variation and wavelets for image enhancement”, *Applied and Computational Harmonic Analysis*, vol. 9, pp. 1-18, 2000.

[3] D. Donoho and I. M. Johnstone, “ Ideal spatial adaptation by wavelet shrinkage”, *Biometrica*, 81(3): 425-455, 1994.

[4] J. Kalifa, “Restauration minimax et deconvolution dans une base d'ondelettes miroirs”, Ph. D. Thesis, Ecole Polytechnique, 1999.

[5] F. Malgouyres, “A framework for image deblurring using wavelet packet bases ”, *Applied and Computational Harmonic Analysis*, vol. 12, no. 3, pp. 309-331, 2002.

[6] F. Malgouyres, “Minimizing the total variation under a general convex constraint for image restoration”, *IEEE. Trans. Image Processing*, vol. 11 (12), pp. 1450-1456, 2002.

[7] F. Malgouyres, “Mathematical analysis of a model which combines total variation and wavelet for image restoration”, *Journal of Information Processes*, vol. 2, pp. 1-10, 2002.

[8] A. Nemirovski, “Topics in Non-Parametric Statistics”, in: M. Emery, A. Nemirovski, D. Voiculescu, *Lectures on Probability Theory and Statistics, Ecole d'ete de Probabilities de Saint-Flour XXVIII - 1998*, Editor: P. Bernard, Lecture Notes in Mathematics 1738, Springer.

[9] A. Nemirovski, ” On nonparametric estimation of smooth regression functions”, *Sov. J. Comput. Syst. Sci.*, vol. 23, No. 6, pp. 1-11, 1985.

[10] L. Rudin, S. Osher and E. Fatemi, “Non Linear Total Variation Based Noise Removal Algorithms”, *Physica D 60 (1992)*, pp. 259-268.

[11] S. Tramini, “Problemes inverse et EDP pour le decodage et la deconvolution d'images”, Ph.D. Thesis, Nice-Sophia-Antipolis, I3S, Nov. 1999.

[12] Woog, L. , “Adapted waveform algorithm for denoising”, Ph.D. Thesis, Yale University.