Science in China Series F: Information Sciences

© 2009 SCIENCE IN CHINA PRESS
Springer

www.scichina.com
info.scichina.com
www.springerlink.com

Wavelet denoising via sparse representation

ZHAO RuiZhen[1,2,3†], LIU XiaoYu[2,3], Ching-Chung LI[3], Robert J. SCLABASSI[2,3] & SUN MinGui[2,3]

[1] Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China;
[2] Department of Neurosurgery, University of Pittsburgh, Pittsburgh, PA 15261, USA;
[3] Department of Electrical & Computer Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA

Wavelet threshold denoising is a powerful method for suppressing noise in signals and images. However, this method often uses a coordinate-wise processing scheme, which ignores the structural properties in the wavelet coefficients. We propose a new wavelet denoising method using sparse representation which is a powerful mathematical tool recently developed. Instead of thresholding wavelet coefficients individually, we minimize the number of non-zero coefficients under certain conditions. The denoised signal is reconstructed by solving an optimization problem. It is shown that the solution to the optimization problem can be obtained uniquely and the estimates of the denoised wavelet coefficients are unbiased, i.e., the statistical means of the estimates are equal to the noise-free wavelet coefficients. It is also shown that at least a local optimal solution to the denoising problem can be found. Our experiments on test data indicate that this new denoising method is effective and efficient for a wide variety of signals including those with low signal-to-noise ratios.

signal processing, denoising, sparse representation, wavelet transform

1 Introduction

There exist three major forms of wavelet denoising methods. The first method, provided by Mallat and Zhong[1], is based on the extrema in the wavelet coefficients. These extrema reflect the propagation properties of the signal and noise across wavelet decomposition scales. The second method is called the correlation method[2] in which wavelet coefficients are kept or eliminated according to the correlation of wavelet coefficients across neighboring scales. The third method relies on the thresholding operation[3]. The wavelet coefficients whose amplitude values are larger than a given threshold are kept or shrunk while the smaller coefficients are eliminated. This operation is commonly referred to as "hard" or "soft" thresholding.

Among the three wavelet denoising methods, the thresholding method is most widely utilized in practice because of its effectiveness and simplicity. This method, however, may exhibit Gibbs phenomena in the neighborhood of discontinuities[4]. In addition, its performance often deteriorates substantially as the signal-to-noise ratio (SNR) of the input signal decreases. It operates on wavelet coefficients individually, ignoring the structural features of the wavelet coefficients which carry im-

Received November 18, 2008; accepted March 2, 2009; published online July 2, 2009
doi: 10.1007/s11432-009-0116-7
†Corresponding author (email: rzhzhao@bjtu.edu.cn)
Supported by the U.S. National Institutes of Health (Grant No. U01 HL91736), and the National High-Tech Research & Development Program of China (Grant No. 2007AA01Z175)

Citation: Zhao R Z, Liu X Y, Li C C, et al. Wavelet denoising via sparse representation. Sci China Ser F-Inf Sci, 2009, 52(8): 1371–1377, doi: 10.1007/s11432-009-0116-7

portant information of the input signal. Many improved forms of thresholding schemes have been proposed in the past few years. For example, Cai et al.[5] proposed a NeighShrink method for processing 1D signal by incorporating neighboring coefficients in the thresholding process. Chen et al.[6] extended NeighShrink to the 2D case for image denoising. In recent years, compressive sensing and sparse representation have become an active field of research in both mathematics and engineering[7,8], providing a powerful tool for many applications including signal denoising. Several signal reconstruction algorithms from noisy compressive measurements have been reported[9−12]. However, the noise that was focused on is a special type resulting from the compressive sensing process. In this work, we investigate a new denoising approach by means of sparse representation. We first compute the wavelet transform of the noisy input signal and obtain a set of wavelet coefficients. We then apply a sparse representation model to these coefficients. Instead of thresholding wavelet coefficients individually, we introduce a measurement matrix to obtain a compressed form of representation and then estimate the ideal noise-free coefficients by minimizing the number of non-zero coefficients under certain conditions. Finally the denoised signal is constructed using the inverse wavelet transform. We will show that this approach produces considerably improved denoising results.

This paper is organized as follows. The denoising model based on sparse representaion is presented in section 2 where we also discuss the existence and uniqueness of the solution to the model under certain conditions. In section 3, we show that, with appropriately chosen parameters, at least a local optimal solution to the model can be found. In section 4, the new denoising algorithm is outlined for practical implementation. Experimental results are given in section 5 and finally conclusions are drawn in section 6.

2 Denoising model based on sparse representation

Let us assume that a signal is contaminated with additive noise

$$f(k) = s(k) + n(k), \tag{1}$$

where $f(k)$ and $s(k)$ are, respectively, the observed and noise-free signals, and $n(k)$ is a random noise with normal distribution $N(0,\sigma^2)$. Let $w_{j,k}$ be the wavelet coefficients of $f(k)$ resulting from the discrete wavelet transform with scale levels $j = 1 \sim J$ and time index $k = 1 \sim N$. The Donoho wavelet denoising algorithm (i.e., the SURE shrinkage[3]) produces the denoised wavelet coefficients $\hat{w}_{j,k}$ from the noisy coefficients $w_{j,k}$ by

Soft-thresholding:

$$\hat{w}_{j,k} = \begin{cases} \operatorname{sign}(w_{j,k}) \cdot (|w_{j,k}| - \lambda), & |w_{j,k}| \geqslant \lambda, \\ 0, & |w_{j,k}| < \lambda; \end{cases} \tag{2}$$

Hard-thresholding:

$$\hat{w}_{j,k} = \begin{cases} w_{j,k}, & |w_{j,k}| \geqslant \lambda, \\ 0, & |w_{j,k}| < \lambda. \end{cases} \tag{3}$$

It is clear from eqs. (2) and (3) that the operations involved are coordinate-wise without using the correctional structure in the wavelet coefficients.

For most practical signals, when the wavelet is appropriately chosen which well matches input signal features, the wavelet coefficients should have only a small number of large-amplitude values (sparse). However, if the signal is heavily contaminated with noise, the sparseness of the wavelet coefficients may decrease significantly. Thus it is reasonable to consider an alternative wavelet-domain operation which can reconstruct the sparseness of the noise-free coefficients. We approach this problem by employing the powerful sparsest representation technique[13], aiming at faithfully estimating the dominant wavelet coefficients without losing significant information. Our approach can be formulated into the following optimization problem.

Without loss of generality, let us omit the subscripts j and k in $w_{j,k}$ and let w denote an $N \times 1$ column vector at a given scale level. $\Phi_{M\times N}$ (with M less than N) is a measurement matrix in sparse representation. As a fundamental requirement, Φ must obey the uniform uncertainty principle

(UUP)[14], i.e., for any S-sparse vector x, if

$$0.8\frac{M}{N}\|x\|_2^2 \leqslant \|\Phi x\|_2^2 \leqslant 1.2\frac{M}{N}\|x\|_2^2, \tag{4}$$

then we call $\Phi_{M\times N}$ obeys UUP for sets of size S with $S \leqslant M\log N$.

Supposing $y = \Phi w$, we can obtain the sparsest coefficient vector $\hat{w}$ by solving

$$\hat{w} = \arg\min_z \|z\|_0 \quad \text{s.t. } \|y - \Phi z\|_2 \leqslant \varepsilon, \tag{5}$$

where $\|\cdot\|_0$ represents the zeroth-order norm defined as the number of nonzero entries[7]. Eq. (5) is in general difficult to implement because it is an NP-hard problem in combinatorial mathematics. However, in the denoising application, we have found a simple way to solve it. Let us modify eq. (5) into the following form[12]:

$$\hat{w} = \arg\min_z (\|y - \Phi z\|_2^2 + \gamma\|z\|_0). \tag{6}$$

In the following, we will verify that, under certain nonrestrictive conditions, the solution of eq. (6) is unique and statistically unbiased, i.e., the mean estimates are equal to the ideal wavelet coefficients.

Let $\bar{w}$ denote the noise-free wavelet coefficients. Then, by the additive assumption, wavelet coefficients of the observed signal can be represented by $w = \bar{w} + w_n$, where w_n denotes wavelet coefficients of random noise $n(k)$. Since $n(k) \sim N(0, \sigma^2)$, w_n is also normal with a zero mean. So we have

$$y = \Phi w = \Phi(\bar{w} + w_n) = \Phi\bar{w} + \Phi w_n. \tag{7}$$

If we denote

$$g(z) = \|y - \Phi z\|_2^2 + \gamma\|z\|_0, \tag{8}$$

then we will verify that $E(g(\bar{w})) < E(g(u))$ holds for any $u \neq \bar{w}$, which leads to $E(\hat{w}) = \bar{w}$.

Let Φ be a random matrix with its entries drawn from Gaussian distribution $N(0, \sigma_\Phi^2)$. Using eqs. (7) and (8), we consider

$$\begin{aligned} & g(u) - g(\bar{w}) \\ & = (\|y - \Phi u\|_2^2 - \|y - \Phi\bar{w}\|_2^2) + \gamma(\|u\|_0 - \|\bar{w}\|_0) \\ & = (\|\Phi(\bar{w} + w_n - u)\|_2^2 - \|\Phi w_n\|_2^2) \\ & \quad + \gamma(\|u\|_0 - \|\bar{w}\|_0) \\ & = (\|\Phi(\bar{w} - u) + \Phi w_n\|_2^2 - \|\Phi w_n\|_2^2) \\ & \quad + \gamma(\|u\|_0 - \|\bar{w}\|_0) \\ & = (\|\Phi(\bar{w} - u)\|_2^2 + 2[\Phi(\bar{w} - u)]^{\mathrm{T}}\Phi w_n) \\ & \quad + \gamma(\|u\|_0 - \|\bar{w}\|_0). \end{aligned} \tag{9}$$

Suppose that $\Phi(\bar{w} - u)$ and Φw_n are independent. Then we can easily get

$$\begin{aligned} & E(g(u) - g(\bar{w})) \\ & = E(\|\Phi(\bar{w} - u)\|_2^2) + \gamma E(\|u\|_0 - \|\bar{w}\|_0). \end{aligned} \tag{10}$$

In the following we will discuss eq. (10) in three cases:

i) For $\|u\|_0 > \|\bar{w}\|_0$, we have $E(g(u) - g(\bar{w})) > 0$ obviously.

ii) For $\|u\|_0 = \|\bar{w}\|_0$, assume $\|u\|_0 = \|\bar{w}\|_0 = K$. We denote $v = \bar{w} - u$, so $\|v\|_0 \leqslant 2K$. Let $\Gamma = \{l | v_l \neq 0\}$ and the size of set Γ is $|\Gamma| \leqslant 2K$. Therefore we can rewrite $\Phi(\bar{w} - u)$ as $\Phi v = \sum_{l\in\Gamma} v_l \Phi_l$, where Φ_l are column vectors of matrix $\Phi_{M\times N}$. Any M columns of Φ are nearly linearly independent since Φ satisfies the property of UUP. Therefore, if $2K \leqslant M$, we can verify that $\sum_{l\in\Gamma} v_l \Phi_l \neq \vec{0}$ is almost true since $v_l \neq 0$ for $l \in \Gamma$. So the first term on the right side of eq. (10) is greater than zero and thus $E(g(u) - g(\bar{w})) > 0$.

iii) For $\|u\|_0 < \|\bar{w}\|_0$, suppose $\|u\|_0 = \|\bar{w}\|_0 - \eta$, then

$$E(g(u) - g(\bar{w})) = E(\|\Phi(\bar{w} - u)\|_2^2) - \gamma\eta. \tag{11}$$

We also denote $v = \bar{w} - u$ and $\Gamma = \{l | v_l \neq 0\}$, and then we have $\eta \leqslant \|v\|_0 = |\Gamma| \leqslant 2K - \eta$. Therefore,

$$\begin{aligned} \|\Phi(\bar{w} - u)\|_2^2 & = \|\Phi v\|_2^2 \\ & = \|\sum_{l\in\Gamma} v_l \Phi_l\|_2^2 \\ & = \sum_{i=1}^{M}\left(\sum_{l\in\Gamma} v_l \Phi_{il}\right)^2. \end{aligned} \tag{12}$$

Due to the statistical property of random matrix Φ, we have $E(\Phi_{i,l}) = 0$, $E(\Phi_{i,l}^2) = D(\Phi_{i,l}) = \sigma_\Phi^2$ and $E(\Phi_{i,l}\Phi_{i,\tilde{l}}) = 0$ for $l \neq \tilde{l}$. Therefore, the expectation of eq. (12) is

$$E(\|\Phi(\bar{w} - u)\|_2^2) = M\sigma_\Phi^2 \sum_{l\in\Gamma} v_l^2 = M\sigma_\Phi^2\|v\|_2^2. \tag{13}$$

Combined with eq. (11), if v satisfies the following condition:

$$\|v\|_2^2 > \frac{\gamma\eta}{M\sigma_\Phi^2}, \tag{14}$$

then we have $E(g(u) - g(\bar{w})) > 0$. Since $\|u\|_0 = \|\bar{w}\|_0 - \eta$, there are at least η non-zero coefficients in vector v such that $v_l = \bar{w}_l$ ($l \in \Gamma$). Besides, each non-zero element in v is practically greater than $\gamma/(\sqrt{M}\sigma_\Phi)$ because M is usually much larger than 1 and the value of σ_Φ can be selected appropriately when matrix Φ is constructed. Therefore we point out that eq. (14) can be easily satisfied and represents a weak condition in practice.

Summarizing all three cases, we have $E(g(u) - g(\bar{w})) > 0$ for any $u \neq \bar{w}$. Due to the fact that $\hat{w} = \arg\min_z g(z)$, we conclude that $E(\hat{w}) = \bar{w}$.

3 Solution to the optimization problem

Let $g_1(z)$ denote the first term on the right side of eq. (8). The necessary condition for minimization of $g_1(z)$ can be found by taking derivative of $g_1(z)$ and setting the result at zero:

$$\frac{\partial g_1(z)}{\partial z} = -2\Phi^{\mathrm{T}}(y - \Phi z) = 0. \tag{15}$$

The steepest decent algorithm that iteratively solves the problem is thus given by

$$w^{(t+1)} = w^{(t)} + \mu\Phi^{\mathrm{T}}(y - \Phi w^{(t)}), \tag{16}$$

where $\mu > 0$ is a constant to be determined as follows. Eq. (16) can be rewritten as

$$w^{(t+1)} = \mu\Phi^{\mathrm{T}}y + (I - \mu\Phi^{\mathrm{T}}\Phi)w^{(t)}. \tag{17}$$

Therefore, the convergence condition of eq. (17) is that $(I - \mu\Phi^{\mathrm{T}}\Phi)$ should be a contractive linear operator, i.e.

$$0 < \|I - \mu\Phi^{\mathrm{T}}\Phi\|_2 < 1. \tag{18}$$

Alternatively, the above algorithm converges if the eigenvalues of the linear operator $(I - \mu\Phi^{\mathrm{T}}\Phi)$ satisfy the following sufficient condition[15], which is stronger than (18):

$$0 < \mathrm{eig}(I - \mu\Phi^{\mathrm{T}}\Phi) < 1. \tag{19}$$

In order to show the validity of this condition, we use b to represent any eigenvalue of matrix $I - \mu\Phi^{\mathrm{T}}\Phi$. Then, we have $(I - \mu\Phi^{\mathrm{T}}\Phi)\alpha = b\alpha$, or $\mu\Phi^{\mathrm{T}}\Phi\alpha = (1-b)\alpha$ for eigenvector $\alpha \neq 0$. It is obvious that b should be less than 1 because $\Phi^{\mathrm{T}}\Phi$ is positive definite with probability 1. Therefore, the upper bound of (19) is satisfied. To satisfy the lower bound, $\|\sqrt{\mu}\Phi\|_2^2 < 1$ is required. Unfortunately, matrix Φ obeying the UUP property (4) does not automatically satisfy $\|\sqrt{\mu}\Phi\|_2^2 < 1$ for any μ. However, if we define s as the largest eigenvalue of $\Phi^{\mathrm{T}}\Phi$, then for any vector $\alpha \neq 0$, we have

$$\|\sqrt{\mu}\Phi\|_2^2 = \mu\max_\alpha \frac{||\Phi\alpha\|_2^2}{\|\alpha\|_2^2}$$

$$= \mu\max_\alpha \frac{\alpha^{\mathrm{T}}\Phi^{\mathrm{T}}\Phi\alpha}{\|\alpha\|_2^2} = \mu\frac{\alpha^{\mathrm{T}}(s\alpha)}{\|\alpha\|_2^2} = \mu s. \tag{20}$$

Therefore, μ in eq. (16) can be chosen as $\mu < \frac{1}{s}$ to guarantee the convergence of the iterative algorithm.

However, the above algorithm does not necessarily converge to the minimum of eq. (6) because the iterative formula (16) does not take $\gamma\|w\|_0$ into consideration. To minimize $\gamma\|w\|_0$, we modify eq. (16) by including a thresholding step:

$$w^{(t+1)} = H_\delta(w^{(t)} + \mu\Phi^{\mathrm{T}}(y - \Phi w^{(t)})), \tag{21}$$

where H_δ is an element-wise hard-thresholding operator

$$H_\delta(w_i) = \begin{cases} w_i, & \text{if } |w_i| \geqslant \delta \\ 0, & \text{otherwise} \end{cases} \quad i = 1 \sim N, \tag{22}$$

where δ is a chosen threshold. The hard-thresholding step of this algorithm guarantees that the number of nonzero entries of w is minimized and the total number of entries is determined by parameter δ.

In the following we will verify that the fixed point $\hat{w}$ of the iterative algorithm in eq. (21) is a local minimum of eq. (6); that is, there exists $\xi > 0$ such that $g(\hat{w} + \Delta w) > g(\hat{w})$ always holds for any Δw with $|\Delta w_i| < \xi$.

We introduce a new surrogate objective function[16]

$$\begin{aligned} G(w,u) &= g(w) - \mu\|\Phi w - \Phi u\|_2^2 + \|w - u\|_2^2 \\ &= \|y - \Phi w\|_2^2 + \gamma\|w\|_0 - \mu\|\Phi w - \Phi u\|_2^2 \\ &\quad + \|w - u\|_2^2. \end{aligned} \tag{23}$$

If $\|\sqrt{\mu}\Phi\|_2^2 < 1$, we find $G(w,u) \geqslant G(w,w) = g(w)$. From eq. (23) we can also get

$$\begin{aligned} g(\hat{w} + \Delta w) = G(\hat{w} + \Delta w, \hat{w}) \\ - \|\Delta w\|_2^2 + \mu\|\Phi\Delta w\|_2^2. \end{aligned} \tag{24}$$

Eq. (23) can be rewritten as

$$\begin{aligned} G(w,u) = &\sum_i [w_i^2 - 2w_i(u_i + \Phi_i^{\mathrm{T}}y - \mu\Phi_i^{\mathrm{T}}\Phi u) \\ &+ \gamma|w_i|^0] + (1-\mu)\|\Phi w\|_2^2 + \|y\|_2^2 \\ &+ \|u\|_2^2 - \mu\|\Phi u\|_2^2, \end{aligned} \tag{25}$$

where $|w_i|^0$ is one if $w_i \neq 0$ and zero otherwise. Using eq. (25), we consider

$$G(\hat{w} + \Delta w, \hat{w}) - G(\hat{w}, \hat{w})$$

$$
\begin{aligned}
&= \sum_i [(\hat{w}_i + \Delta w_i)^2 - 2(\hat{w}_i + \Delta w_i) \\
&\quad \times (\hat{w}_i + \Phi_i^{\mathrm{T}} y - \mu \Phi_i^{\mathrm{T}} \Phi \hat{w}) + \gamma |\hat{w}_i + \Delta w_i|^0] \\
&\quad - \sum_i [\hat{w}_i^2 - 2\hat{w}_i(\hat{w}_i + \Phi_i^{\mathrm{T}} y - \mu \Phi_i^{\mathrm{T}} \Phi \hat{w}) + \gamma |\hat{w}_i|^0] \\
&\quad + (1-\mu)(\|\Phi(\hat{w} + \Delta w)\|_2^2 - \|\Phi \hat{w}\|_2^2) \\
&= \sum_i [(\Delta w_i)^2 - 2\Delta w_i \Phi_i^{\mathrm{T}}(y - \mu \Phi \hat{w}) \\
&\quad + \gamma(|\hat{w}_i + \Delta w_i|^0 - |\hat{w}_i|^0)] \\
&\quad + (1-\mu)\sum_i 2\Delta w_i \Phi_i^{\mathrm{T}} \Phi \hat{w} + (1-\mu)\|\Phi \Delta w\|_2^2 \\
&= \|\Delta w\|_2^2 + (1-\mu)\|\Phi \Delta w\|_2^2 \\
&\quad + \sum_i [-2\Delta w_i \Phi_i^{\mathrm{T}}(y - \Phi \hat{w}) \\
&\quad + \gamma(|\hat{w}_i + \Delta w_i|^0 - |\hat{w}_i|^0)].
\end{aligned}
$$

We split the above summation into two parts, one part for $\Gamma_0 = \{i|\hat{w}_i = 0\}$ and the other for $\Gamma_1 = \{i|\hat{w}_i \neq 0\}$. If $\Delta w_i \neq -\hat{w}_i$ for $i \in \Gamma_1$, we get

$$
\begin{aligned}
&G(\hat{w} + \Delta w, \hat{w}) - G(\hat{w}, \hat{w}) \\
&= \|\Delta w\|_2^2 + (1-\mu)\|\Phi \Delta w\|_2^2 \\
&\quad - \sum_{\Gamma_0} [2\Delta w_i \Phi_i^{\mathrm{T}}(y - \Phi \hat{w}) - \gamma |\Delta w_i|^0] \\
&\quad - \sum_{\Gamma_1} [2\Delta w_i \Phi_i^{\mathrm{T}}(y - \Phi \hat{w})]. \qquad (26)
\end{aligned}
$$

According to our iterative equation (21), the solution $\hat{w}_i = \hat{w}_i + \mu \Phi_i^{\mathrm{T}}(y - \Phi \hat{w})$ holds for $i \in \Gamma_1$. Therefore, we have $\Phi_i^{\mathrm{T}}(y - \Phi \hat{w}) = 0$ for $i \in \Gamma_1$, which means that the second summation term in eq. (26) is equal to zero. Let $\xi = \min_{i\in\Gamma_0} \frac{\gamma}{2|\Phi_i^{\mathrm{T}}(y - \Phi\hat{w})|}$. And if we choose $|\Delta w_i| \leqslant \xi$ for $i \in \Gamma_0$ and $\Delta w_i \neq -\hat{w}_i$ for $i \in \Gamma_1$, we can easily get

$$
\begin{aligned}
&G(\hat{w} + \Delta w, \hat{w}) - G(\hat{w}, \hat{w}) \\
&\geqslant \|\Delta w\|_2^2 + (1-\mu)\|\Phi \Delta w\|_2^2. \qquad (27)
\end{aligned}
$$

Substituting (27) to (24), we get

$$
\begin{aligned}
g(\hat{w} + \Delta w) &= G(\hat{w} + \Delta w, \hat{w}) - \|\Delta w\|_2^2 + \mu \|\Phi \Delta w\|_2^2 \\
&\geqslant G(\hat{w}, \hat{w}) + \|\Phi \Delta w\|_2^2 \geqslant g(\hat{w}). \qquad (28)
\end{aligned}
$$

Hence, $g(\hat{w})$ is a local minimum.

4 Denoising algorithm

Since the wavelet coefficients, $\bar{w}_j$, of the pure signal are strongly correlated across scales, it is likely that the nonzero coefficients are located in the same neighborhood at different scales. Therefore, if we have already processed the coefficients at a certain scale level, we can make use of the position information of the processed coefficients to predict the locations of the significant coefficients at other scales. This property can be used to accelerate convergence in the optimization process by iteratively selecting initial values. For scale level $j = 1 \sim J$, we choose the iterative initial values as

$$
\beta_j = \begin{cases} w_J, & \text{if } j = J, \\ \hat{w}_{j+1}, & \text{if } j < J. \end{cases} \qquad (29)
$$

The wavelet denoising algorithm based on sparse representation can thus be described as follows:

1) Perform an undecimated wavelet transform on noisy observation $f(k)$ and obtain wavelet coefficients $w_j (j = 1 \sim J)$.

2) Choose appropriate $\Phi_{M\times N}$ satisfying both UUP property (4) and condition (14). Compute $y_j = \Phi w_j (j = 1 \sim J)$ and then let $j = J$.

3) Let $w_j^{(0)} = 0$ and $w_j^{(1)} = \beta_j$ by using eq. (29).

4) If $\|w_j^{(1)} - w_j^{(0)}\|_2 > \varepsilon$, then let $w_j^{(0)} = w_j^{(1)}$ and use (21) to compute $w_j^{(1)}$, repeat step 4); else, let $\hat{w}_j = w_j^{(1)}$, go to step 5).

5) If $j = 1$, go to step 6); else, $j = j - 1$, go to step 3);

6) Perform the inverse wavelet transform on the estimated wavelet coefficients $\hat{w}_j$.

5 Experimental results

We selected Blocks, Bumps, HeaviSine and Doppler, which are popular library signals utilized, for example, in ref. [3], as the test signals for denoising algorithms. The length and SNR of each test signal were, respectively, $N = 2048$ and 6.1320 dB. The cubic spline wavelet was chosen because of its simplicity, finite support, and smoothness. The largest decomposition scale level was $J = 4$, chosen experimentally. We used the median value method to estimate the standard deviation of the noise[17]

$$
\sigma = \mathrm{Median}(|w_{1,k}|)/0.6745, \qquad (30)
$$

where $w_{1,k}$ represents the wavelet coefficients at scale level $j = 1$. In our experiments we empirically chose $\varepsilon = 1$ and threshold $\delta = \sigma/4$ in eq. (22).

Table 1 SNR (in dB) of the results obtained from different denoising algorithms

	Blocks	Bumps	Heavisine	Doppler
Soft-thresh.	11.4994	11.8633	21.6421	13.8678
Hard-thresh.	12.7075	12.4993	21.5754	13.8762
NeighShrink	15.2111	14.9829	21.6857	14.8955
Our method ($M = N/8$)	16.6193	17.4104	21.7276	16.2274
Our method ($M = N/4$)	17.0667	18.8759	21.9509	16.5522

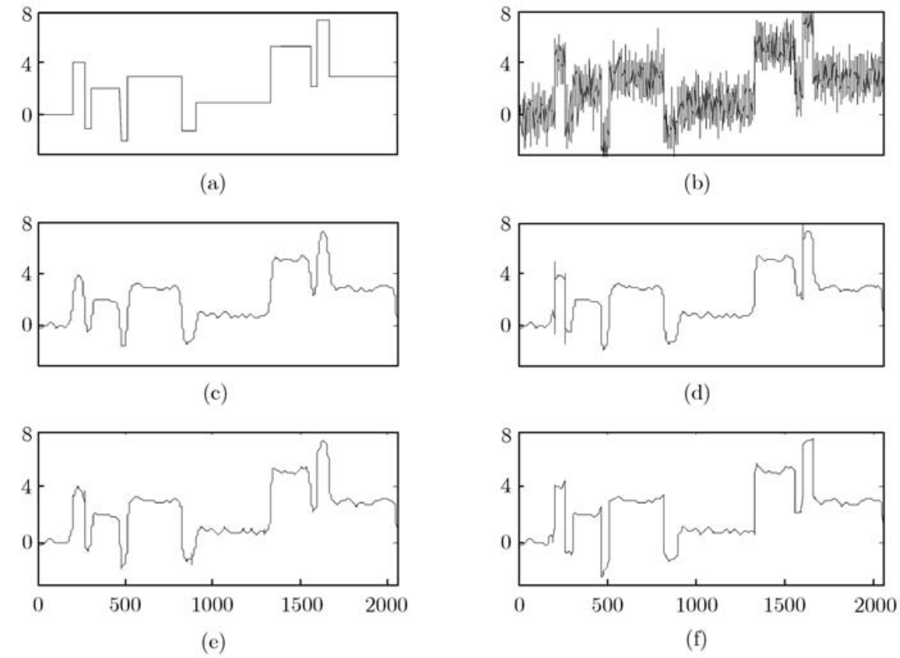

Figure 1 Simulation example using Blocks signal with different denoising methods. (a) Noise-free signal; (b) noisy signal with SNR=6.1320 dB; (c)–(f) are denoised signals respectively by (c) soft-thresholding; (d) hard-thresholding; (e) NeighShrink; (f) proposed method with $M = N/8$.

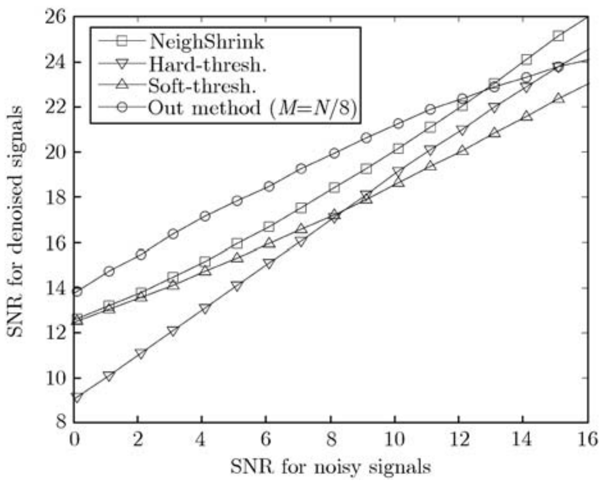

Figure 2 SNR comparison of different methods.

From Table 1, it can be observed that the new algorithm (with two choices of M values) achieves higher SNR gains than the hard-thresholding, soft-thresholding and NeighShrink methods. It can also be observed that the performance of our algorithm improves as M increases. However, this improvement is achieved at the cost of more entries in y_j. As an example, the denoising result of the Blocks signal using the new method is compared with the results of several existing methods in Figure 1. It can be observed that the new method is clearly more superior.

In order to show the robustness of our method, we calculate the average SNR variation between the input and the output. The results of different methods are plotted in Figure 2 where the new method shows the highest output SNR for input SNR less than 12.5 dB. This indicates that the new method is more suitable in low SNR cases where other methods cannot denoise signals effectively.

6 Conclusion

Although the standard wavelet threshold denoising method is simple and practical, it does not work well for signals with a low SNR. Besides, it uses coordinate-wise processing without taking the structural information of the wavelet coefficients into account. Our new denoising method based on wavelet transform and sparse representation overcomes these problems. It estimates the noise-free wavelet coefficients by solving an optimization problem and is effective and efficient for a wide variety of signals including those with low SNRs.

1 Mallat S, Zhong S. Characterization of signals from multiscale edges. IEEE Trans PAMI, 1992, 14(7): 710–732

2 Xu Y, Weaver J, Healy M, et al, Wavelet transform domain filters: a spatially selective noise filtration technique. IEEE Trans Image Process, 1994, 3(6): 747–758

3 Donoho D L. De-noising by soft-thresholding. IEEE Trans Inf Theory, 1995, 41(3): 613–627

4 Kelly S E. Gibbs phenomenon for wavelets. Appl Comput Harmon Anal, 1996, 3(2): 72–81

5 Cai T T, Silverman B W. Incorporating information on neighboring coefficients into wavelet estimation. Indian J Statist, 2001, 63(B): 127–148

6 Chen G Y, Bui T D, Krzyzak A. Image denoising using neighboring wavelet coefficients. ICASSP, 2004, 2: 917–920

7 Candès E J, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans Inf Theory, 2006, 52(2): 489–509

8 Donoho D L. Compressed sensing. IEEE Trans Inf Theory, 2006, 52(4): 1289–1306

9 Boufounos P, Duarte M, Baraniuk R. Sparse signal reconstruction from noisy compressive measurements using cross validation. In: Proc IEEE Statistical Signal Processing Workshop (SSP) 2007, Vol. 15, Iss. 1, August 26-29, Madison, WI, 2007. 358–368

10 Elad M, Aharon M. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process, 2006, 15(12): 3736–3745

11 Donoho D L, Elad M, Temlyakov V. Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans Inf Theory, 2006, 52(1): 6–18

12 Haupt J, Nowak R. Signal reconstruction from noisy random projections. IEEE Trans Inf Theory, 2006, 52(9): 4036–4048

13 Baraniuk R G. Compressive sensing. IEEE Signal Proc Mag, 2007, 24(4): 118–120

14 Candès E J, Tao T. Near-optimal signal recovery from random projection: universal encoding strategies? IEEE Trans Inf Theory, 2006, 52(12): 5406–5426

15 Blumensath T, Davies M E. Iterative thresholding for sparse approximations. J Fourier Anal Appl, 2008, 14(5-6): 629–654

16 Lange K, Hunter D R, Yang I. Optimization transfer using surrogate objective functions. J Comput Graph Statist, 2006, 9: 1–20

17 Pizurica A, Philips W, Lemahieu I, et al. A joint inter- and intrascale statistical model for Bayesian wavelet based image denoising. IEEE Trans Image Proc, 2002, 11(5): 545–557