USC–SIPI REPORT #254

Robust Scale/Time-Limited Extrapolation with Denoising

by

Li-Chien Lin and C.-C. Jay Kuo

March 1994

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 404
Los Angeles, CA 90089-2564 U.S.A.

Robust Scale/Time-Limited Extrapolation with Denoising*

Li-Chien Lin† and C.-C. Jay Kuo†

March 21, 1994; EDICS - SP 2.5.2 Extrapolation

Abstract

We examine a new signal modeling technique using wavelets, and propose an iterative scale/time-limited extrapolation algorithm. The ill-posedness of the problem is discussed. To obtain a robust extrapolation scheme for handling noisy data, we incorporate the extrapolation algorithm with a denoising procedure which exploits the time-localization property of the wavelet transform. A numerical experiment is given to compare the proposed algorithm with the regularized extrapolation method.

1 Introduction

Extrapolating a signal $f(t)$ from its partially known data in a finite interval $[-T, T]$ with some a priori knowledge of the signal is a fundamental problem in signal reconstruction. Numerical algorithms have been developed for band-limited signal extrapolation. One famous example is the Papoulis-Gerchberg (PG) iterative algorithm [3], [7]. The performance of an extrapolation algorithm is highly depend on a proper modeling of the underlying signal. There are however signals which are not band-limited such as time-limited signals. Thus, it is important to seek a more general signal extrapolation algorithm by relaxing the band-limited condition.

More recently, multiresolution wavelet bases with a good time-frequency localization property have been used for signal modeling, and a generalized PG algorithm based on the wavelet representation has been proposed by Xia, Kuo and Zhang [12]. Instead of using the band-limited signal model, Xia *et al.* considered a class of scale-limited signals contained by a certain wavelet subspace. We use two examples to illustrate the modeling power of the scale-limited model. First, by choosing the wavelet basis to be the sinc functions, the scale-limited model reduces to the band-limited model. Thus, the scale-limited signal model does include the band-limited as a special case. Second, consider the cubic cardinal B-spline wavelet [1], where we approximate a function by the linear combination of a set of basis functions which are second-order polynomials between the knots with continuous first-order derivate at the knots. For the function $f(t)$ with only continuous first-order derivate, it is better to represent the function with the cardinal B-spline wavelet basis rather than the conventional Fourier basis.

*This work was supported by the National Science Foundation Presidential Faculty Fellow (PFF) Award ASC-9350309.
†The authors are with the Signal and Image Processing Institute and the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, California 90089-2564. E-mail: llin@sipi.usc.edu and cckuo@sipi.usc.edu.

Due to the ill-posed nature of the extrapolation problem, a small amount of noise may lead to very different extrapolated solutions. It is often to use a regularization scheme to extrapolate noisy data in band-limited extrapolation. The major disadvantage of the regularization method is that it tends to oversmooth the features of a signal. Multiscale (scale-space) filtering has been studied in the last decade [11]. A nonlinear filtering technique called denoising [6] was developed to remove the noise of a signal without affecting important signal features such as sharp edges . In this correspondence, we incorporate the denoising process in the iterative scale/time-limited extrapolation procedure to obtain a robust signal extrapolation algorithm. A numerical experiment is given to show the superior performance of the proposed algorithm over that of the regularized extrapolation method.

2 Scale/Time-Limited Extrapolation

2.1 Signal Modeling with Wavelets

We review some basic results of orthogonal wavelet theory below, and refer to [2] for more detailed discussion. A multiresolution analysis (MRA) consists of a sequence of successive approximation space $\mathcal{P}_j$ of $L^2(\mathbf{R})$. That is, the $\mathcal{P}_j$ with $j = \cdots, -1, 0, 1 \cdots$ satisfy

$$\cdots \subset \mathcal{P}_{-2} \subset \mathcal{P}_{-1} \subset \mathcal{P}_0 \subset \mathcal{P}_1 \subset \mathcal{P}_2 \cdots, \quad \text{with} \quad \overline{\bigcup_j \mathcal{P}_j} = L^2(\mathbf{R}), \quad \bigcap_j \mathcal{P}_j = \{0\}.$$

Let $\phi(t)$ be the scaling function associated with an MRA. Then, for an arbitrary integer j, $\{\phi_{jk}(t)\}_{k\in\mathbf{Z}}$, where $\phi_{jk}(t) = 2^{j/2}\phi(2^j t - k)$, is an orthonormal basis of the wavelet subspace $\mathcal{P}_j$. The mother wavelet function corresponding to $\phi(t)$ is denoted by $\psi(t)$ and $\{\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k),\ j,k \in \mathbf{Z}\}$ forms an orthonormal basis in $L^2(\mathbf{R})$. For any $f(t) \in L^2(\mathbf{R})$, we have

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} b_{j,k}\psi_{jk}(t). \tag{1}$$

The projection $f_J(t)$ of $f(t) \in \mathcal{P}_J$ can be written as

$$f_J(t) = \sum_{k=-\infty}^{\infty} c_{J,k}\phi_{Jk}(t) = \sum_{j<J} \sum_{k=-\infty}^{\infty} b_{j,k}\psi_{jk}(t).$$

The multiresolution analysis leads naturally to a hierarchical and fast scheme for the computation of wavelet coefficients $b_{j,k}$ with $j < J$ from coefficients $c_{J,k}$ by recursive formulas called the discrete wavelet transform (DWT). The inverse process is called the inverse discrete wavelet transform (IDWT).

We represent $f(t)$ by the wavelet basis ψ_{jk} in (1). Let us assume that $\psi(t)$ is centered around 0 in time and $\pm\xi_0$ in frequency and is well localized in both time and frequency. By using the scaling property, $\psi_{jk}(t)$ is localized around $2^{-j}k$ in time and $\pm 2^j\xi_0$ in frequency. Thus, we may interpret the wavelet coefficient $b_{j,k} =< f, \psi_{j,k} >$ as the "information content" of f near $2^{-j}k$ in time and $\pm 2^j\xi_0$ in frequency, which is plotted in the $t - w$ plane in Fig. 1. Now, suppose that the energy of $f(t)$ is concentrated in two rectangle regions as shown in Fig. 1(b), i.e.

$$\mathcal{R} = [-T, T] \times \ [(-2^J\xi_0, -2^{J_0}\xi_0) \cup (2^{J_0}\xi_0, 2^J\xi_0)],$$

in the sense that we can find a small ϵ so that

$$\int_{2^{J_0}\xi_0 \leq |\xi| \leq 2^J \xi_0} |\hat{f}|^2 \, d\xi \geq (1-\epsilon)\|f\|^2, \quad \text{and} \quad \int_{|x|\leq T} |f|^2 \, dx \geq (1-\epsilon)\|f\|^2.$$

Then, only the wavelet coefficients $b_{j,k}$, $(j,k) \in \mathbf{S}$ are needed for a good approximation of $f(t)$, where

$$\mathbf{S} = \{(j,k) \in \mathbf{Z}^2; J_0 \leq j \leq J \text{ and } 2^{-j} \mid k \mid \leq T + t_\epsilon\}$$

is the set of dydaic points which is enclosed by the dashed shape in Fig. 1(a). We refer to [2] for a more detailed discussion of this approximation. Thus, the following space

$$\mathbf{V_S} = \{f(t): \; f(t) = \sum_{(j,k)\in\mathbf{S}} b_{j,k}\psi_{jk}(t)\}.$$

provides a good model for signals concentrated in the region $\mathcal{R}$. We say that $\mathbf{V_S}$ provides a scale/time-limited signal model.

For the rest of the paper, we assume that the scale/time-limited signal model $\mathbf{V_S}$ is known, and will focus on the development of a robust extrapolation algorithm based on such a model. It is worthwhile to point out that the scale/time-limited function $f(t)$ is in general not analytic and, therefore, we can only extrapolate the signal up to a certain finite interval. In contrast, in the band-limited extrapolation context, it is possible in theory to extrapolate the signal to $\pm\infty$ by using the analytic property of band-limited functions. Since we can only obtain a robust extrapolation result within a certain finite interval in practice, the scale/time-limited model is in fact more consistent with our experience.

2.2 Scale/Time-Limited Extrapolation Algorithm

Consider a sequence $x[k]$ obtained from a sampled signal $f(k/2^{J_s})$ and providing a good approximation to the finest wavelet coefficients $c_{J_s,k}$. Let $\mathbf{x}$ be the column vector representation of the sequence $x[k]$. We call $x[k]$ a *scale/time-limited sequence* if

$$\mathbf{x} = \mathbf{D}^{-1}\mathbf{P}_S\mathbf{D}\mathbf{x}, \tag{2}$$

where $\mathbf{D}^{-1}$ and $\mathbf{D}$ are the DWT and IDWT operators and $\mathbf{P}_S$ has the elements

$$p_{jk} = \begin{cases} 1, & \text{if } (j,k) \in \mathbf{S}, \\ 0, & \text{otherwise.} \end{cases}$$

Let L be the cardinality of $\mathbf{S}$. Then, without loss of generality, we can express $\mathbf{P}_S$ in block matrix form as,

$$\mathbf{P}_S = \begin{pmatrix} \mathbf{I}_L & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_L \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{I}_L & \mathbf{0} \end{pmatrix} = \mathbf{U}^T\mathbf{U},$$

where $\mathbf{U}$ is of dimension $L \times \infty$.

The signal extrapolation problem can be stated as: given a segment of the scale/time-limited sequence $x[n]$ with $|n| \leq N$, we want to determine $x[n]$ for $|n| > N$. We use the column vector $\mathbf{y}$ to denote the observed sequence of length $(2N+1)$. We can relate $\mathbf{y}$ and $\mathbf{x}$ via

$$\mathbf{y} = \mathbf{P}_N\mathbf{x},$$

where

$$\mathbf{P}_N x[n] = \begin{cases} x[n], & |n| \le N, \\ 0, & \text{otherwise.} \end{cases}$$

Since $\{x[k]\}$ is a scale/time-limited sequence, we have from (2) that

$$\mathbf{y} = \mathbf{P}_N \mathbf{D}^{-1} \mathbf{P}_S \mathbf{D} \mathbf{x} = \mathbf{P}_N \mathbf{D}^{-1} \mathbf{U}^T \mathbf{U} \mathbf{D} \mathbf{x},$$

which is to be solved for $\mathbf{x}$ with a given $\mathbf{y}$. It is helpful to consider a two step process for its solution. Let $\mathbf{q} = \mathbf{UDx}$ be a vector consisting of L wavelet coefficients of $\mathbf{x}$. We can first solve

$$\mathbf{y} = \mathbf{P}_N \mathbf{D}^{-1} \mathbf{U}^T \mathbf{q} \tag{3}$$

for $\mathbf{q}$. Since $\mathbf{x} = \mathbf{D}^{-1} \mathbf{P}_S \mathbf{D} \mathbf{x} = \mathbf{D}^{-1} \mathbf{U}^T \mathbf{U} \mathbf{D} \mathbf{x}$ and $\mathbf{q} = \mathbf{UDx}$, we obtain $\mathbf{x}$ by performing the inverse wavelet transform on $\mathbf{U}^{\mathbf{T}} \mathbf{q}$, i.e.

$$\mathbf{x} = \mathbf{D}^{-1} \mathbf{U}^T \mathbf{q}.$$

Various iterative algorithms for solving the extrapolation problem have been examined by the authors in [5].

2.3 Ill-posedness and Regularization

It is well known that band-limited extrapolation is an ill-posed problem. That is, that addition of a small amount of noise to known data may render the inverse process unstable. In this section, we show that scale/time-limited extrapolation is also ill-posed. Let us rewrite (3) as

$$\mathbf{y} = \mathbf{W} \mathbf{q}, \quad \text{where} \quad \mathbf{W} = \mathbf{P}_N \mathbf{D}^{-1} \mathbf{U}^T \in \mathbf{R}^{(2N+1)\times L}. \tag{4}$$

It is assumed that the matrix $\mathbf{W}$ is of rank L. We are interested in the singular values of $\mathbf{W}$ which are by definition the square root of the eigenvalues of $\mathbf{W}\mathbf{W}^T$. Let

$$\mathbf{W}\mathbf{W}^T \mathbf{u}_i = \lambda_i \mathbf{u}_i,$$

where the eigenvalues λ_i's are ordered such that

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_L \ge \lambda_{L+1} = \cdots = \lambda_{2N+1} = 0,$$

and vectors $\mathbf{u}_i \in \mathbf{R}^{2N+1}$ are the corresponding orthonormal eigenvectors.

Let $\mathcal{S}_N = \{\mathbf{y} : \mathbf{y} = \mathbf{P}_N \mathbf{x}\}$. It is clear that $\mathcal{S}_N$ is a linear subspace of $\mathbf{R}^{2N+1}$. Since $\mathbf{W}$ and $\mathbf{W}\mathbf{W}^T$ have the same range space and vectors $\mathbf{u}_i$ span the column space of $\mathbf{W}$, vectors $\mathbf{u}_i$ with $1 \le i \le L$ form an orthonormal basis of $\mathcal{S}_N$. We define

$$(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_L) = \mathbf{W}^T (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_L).$$

Since $\mathbf{u}_i$, $1 \le i \le L$, are linearly independent and $\mathbf{W}^T$ has a rank L, we know that $\mathbf{v}_i$, $1 \le i \le L$, are linearly independent and form a basis of $\mathbf{R}^L$. Any $\mathbf{q} \in \mathbf{R}^L$ can be written as

$$\mathbf{q} = \sum_{i=1}^{L} a_i \mathbf{v}_i = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_L)\mathbf{a},$$

where $\mathbf{a} = (a_1, a_2, \cdots, a_L)^T$. Thus, the solution of the system (4) is equivalent to the computation of coefficients a_i of the vector $\mathbf{q}$.

To compute coefficients a_i, we have

$$\begin{aligned} \mathbf{y} &= \mathbf{W}\mathbf{q} = \mathbf{W}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_L)\mathbf{a} = \mathbf{W}\mathbf{W}^T(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_L)\mathbf{a} \\ &= a_1\lambda_1\mathbf{u}_1 + a_2\lambda_2\mathbf{u}_2 + \cdots + a_L\lambda_L\mathbf{u}_L, \end{aligned}$$

and therefore

$$a_i = \frac{1}{\lambda_i} < \mathbf{y}, \mathbf{u}_i >,$$

where the orthogonality of $\mathbf{u}_i$ is used. It is easy to see from the above formula that the ill-posedness of this problem depends primarily on the value of the smallest nonzero eigenvalue of the matrix $\mathbf{W}\mathbf{W}^T$.

For the case $\lambda_L \approx 0$, computation error or noise in the data can be greatly magnified by the division of λ_L so that the solution procedure is unstable. In practice, the known portion of the data will be accompanied by noise so that the solution will be sensitive to noise. Numerous techniques have been proposed to handle the noise in band-limit extrapolation, including regularization theory [8] and truncated singular value decomposition [10]. Although all of these approachs can be applied to the scale/time-limited extrapolation context, they have to meet two conflicting goals, i.e. accuracy and stability. This is because that all of them can be viewed as some kind of smoothing filter and stability can be achieved by smoothing noise at the expense of signal accuracy.

Regularization can be obtained by adding a regularizing term to the operator $\mathbf{W}$ in (4). For example, we can modify (4) to be

$$\mathbf{y} = (\mathbf{W} + \alpha\mathbf{I})\mathbf{q}.$$

Here, we use the result of Ross [8] by choosing α to be the inverse of the signal to noise ratio (SNR). Then, an iterative extrapolation algorithm with regularization can be written as:

Discrete Generalized PG algorithm with regularization:

$$\begin{aligned} \mathbf{x}^{(0)} &= \mathbf{P}_N\mathbf{x}, \\ \mathbf{x}^{(l+1)} &= \mathbf{P}_N\mathbf{x} + [(1-\alpha)\mathbf{I} - \mathbf{P}_N\mathbf{D}^{-1}\mathbf{P}_S\mathbf{D}]\mathbf{x}^{(l)}, \quad \text{for } l = 0, 1, 2, \cdots. \end{aligned}$$

When $\alpha = 0$, the above algorithm reduces to the generalized PG algorithm proposed by Xia, Kuo and Zhang [12].

3 Robust Extrapolation with Denoising

We use a block diagram as given in Fig. 2 to illustrate the idea of integrating the denoising process and the iterative scale/time-limited extrapolation procedure. These two building components are enclosed by dashed boxes in the figure. We discussed the iterative extrapolation algorithm in the previous section, and will focus on the denoising box in this section.

We know from the scale-space filtering literature [6], [11] that local extrema and zero-crossings provide significant features of signals. In this research, we use the local extrema as the feature to detect noise. By searching along the time axis, local maxima of the modulus of wavelet coefficients can be detected by considering the local properties such as first-order derivatives of the wavelet coefficients in each scale. Generally speaking, the modulus maxima come from two kinds of singularities: the feature of the signal and large noise. Hence, the denoising algorithm should be able to discriminate large noise from meaningful features of the signal. Based on a priori knowledge and estimation of the local regularity of the data, two criteria are developed. The first criterion is simple. For fine-scale wavelet coefficients, since the amplitude of the noise is in general larger than the amplitude of the signal, we can simply choose a amplitude threshold for the modules maxima for discrimination. The threshold depends on some a priori information of the signal, say, the SNR value. The second criterion called the cross-scale criterion is more complicated and detailed in the next subsection. Finally, if a high noisy point is predicted, we replace it by a smoothed data via

$$\bar{y}[n] = \begin{cases} x_k[n] & \text{if noise point is detected} \\ y[n] & \text{otherwise} \end{cases}$$

The cross-scale criterion examines the modulus maxima across scales. To understand it, we have to study the relationship between the modulus maxima in different scales. Some results from wavelet theory can be used for this purpose. In the context of signal processing, it is reasonable to assume that the signal has positive Lipschitz regularity [6]. In contrast, it can be shown that the wide-sense stationary white noise creates singularity whose regularity is negative. Thus, the local maxima of the wavelet coefficient modulus provide enough information to characterize the singularities. By integrating information across scales to estimate the local regularities of the function, we can differentiate large noise from the meaningful feature of the signal.

Let $\alpha = n + \beta, n \in \mathrm{Z}, 0 < \beta < 1$. The function $f(t)$ is said to be Lipschitz α at t_0, if and only if there exist two constants C and h_0, and a polynomial of order n, P_n, such that for $h < h_0$

$$| f(t_0 + h) - P_n(h) | \leq A \mid h \mid^{\alpha} .$$

It can be shown that if $f(t)$ is Lipschitz α at t_0, then $f(t) \in \mathrm{C}^n$, that is n-time continuously differentiable. The following property was mentioned by G. Strang in [9]. Given an arbitrary function $f(t) \in C^n(\mathrm{R})$ and a mother wavelet $\psi(t)$ with compact support, p-vanishing moment (i.e. $\int t^k \psi(t) dt = 0$, for $k = 0, \cdots, p-1, p \geq n$), the wavelet coefficients of a smooth function decay like

$$| Wf(s,k) | = | f_{jk} | = \int f(t)\psi(2^j t - k) dt \leq C(2^{-j})^n = C(s)^n,$$

where $s = 2^{-j}$ is the scale parameter. This result allows us to measure the Lipschitz regularity n by analyzing the ratio change of the modulus maxima during the alternation of the scale.

Now, let us consider the behavior of the modulus maxima for the wavelet transform of the noise. We can use the result in [4] by Grossmann et al. Consider a wide sense stationary white noise $n(t)$ with zero mean and variance σ^2. We denote the wavelet transform of $n(t)$ by $Wn(s,t)$. They showed that the expectation of the wavelet transform $E(|Wn(s,t)|^2)$ is proportional to $1/s$. In mathematics,

$$|Wn(s,t)|^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(u)n(v)\psi(\frac{t-u}{s})\psi(\frac{t-v}{s})dudv.$$

Since $E(n(u),n(v)) = \sigma^2\delta(u-v)$, hence

$$E(|Wn(s,t)|^2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \sigma^2\delta(u-v)\psi(\frac{t-u}{s})\psi(\frac{t-v}{s})dudv.$$

Consequently, we have

$$E(|Wn(s,t)|^2) = \frac{|\psi|^2\sigma^2}{s}.$$

Since almost all noise creates singularities with negative regularity, if the modulus maxima decrease in the amplitude as the scales decrease (from fine to coarse scales), the singularity is most likely contributed from noise. Conversely, if the amplitudes of the modulus maxima increase as scales decrease, it should be created by the feature of the signal. Consequently, this property allow us to discriminate the noise from the feature of signal by the magnitude information across scale. Nontheless, there are two difficulties in practice. First, the modulus maxima will shift when the scale changes. In [6], they showed that the modulus maxima are actually constrainted by a corn in the time-scale domain. This observation can be used to overcome this difficulty. When detecting the modulus maxima in a fine scale, we also check the modulus maxima in its nearest neighboring coefficients at the next coarse scale. Second, a singularity in general generate more than one maximum. In addition, the number of maxima increases when the vanishing moments of wavelets increase. To reduce the computational complexity, one should choose wavelets with lower vanishing moments. But we need sufficient vanishing moments to be able to discriminate the signal from noise. These two difficulties are related to the appropriate choice of good wavelet basis.

4 Experimental Results

A numerical example is given to illustrated the performance of the proposed scale/time-limited extrapolation algorithm with denoising. We use the coiflet basis of order $L = 10$ [2] in this experiment due to the following nice properties. The compact support property makes the implementation easy. The basis function is nearly symmetric so that the filter bank implementation consists of almost linear-phase filters. The scale/time-limited signal model is assumed to be known a priori. Note that the estimation of the scale/time-limited model is important, but out of the scope of this work.

Consider a scale/time-limited sequence $x[n]$ generated by randomly choosing the wavelet coefficients $b_{j,k}$ with $j = 1$ and $-3 \leq k \leq 4$ (while other wavelet coefficients are set to zero) for the coiflet basis functions and observed at the scale $J_s = 4$. The synthesized signal is plotted in Fig. 3 (a) The signal is corrupted by the zero-mean additive white Gaussian noise with SNR = 9.0. There are 35 noisy data points observed as given in Fig. 3(b). We show the detected points with high additive noise values in Fig. 4(b), and the extrapolated signal of applying the denoising processing in the iterative generalized PG algorithm in Fig. 4(c). Almost all points with high noise in the known data interval are detectable from our denoising algorithm. Based on this information, we can remove the detected noise data points without affecting the futures of the signal. In contrast, the traditional regularization process tends to oversmooth the extrapolated result as given in Fig. 4(d). Note that since high noisy points can be detectable in only a few iterations, the computational cost of incorporating the denoising algorithm into the iterative generalized PG algorithm is not high.

5 Conclusion

In this research we examined the signal extrapolation problem based on the wavelet representation of signals. A denoising algorithm has been integrated into the iterative generalized PG algorithm. It was shown numerically that the denoising algorithm performs well when applied to low SNR data. Thus, extrapolation with wavelet representation provides a new promising approach in signal recovery.

References

[1] C. K. CHUI, *An Introduction to Wavelets*, Academic Press, New York, 1992.

[2] I. DAUBECHIES, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[3] R. W. GERCHBERG, *Super-resolution through error energy reduction*, Optica Acta, 21 (1974), pp. 709–720.

[4] A. GROSSMANN, *Wavelet transform and edge detection*, in Stochastic Process in Physics and engineering, M. Hazewinkel, ed., Reidel, Dodrecht, 1986.

[5] L.-C. LIN AND C. C. J. KUO, *On the convergence of signal extrapolation using wavelets.* Submitted to SIAM Matrix Analysis and Applications.

[6] S. G. MALLAT AND W. L. HWANG, *Singularity detection and processing with wavelets*, IEEE Trans. on Information Theory, 38 (1992), pp. 617–643.

[7] A. PAPOULIS, *A new algorithm in spectral analysis and band limited extrapolation*, IEEE Trans. on Circuits and Systems, 22 (1975), pp. 735–742.

[8] G. ROSS, *iterative methods in information processing for object restoration*, Optica Acta, 29(11) (1982), pp. 1523–152.

[9] G. STRANG, *Wavelet and dilation equation: a brief introduction*, SIAM Review, 31 (1989), pp. 614–627.

[10] B. J. SULLIVAN AND B. LIU, *On the use of singular value decomposition and decimation in discrete-time band-limited signal extrapolation*, IEEE Trans. on Acoustic, Speech, and Signal Processing, 32 (1984), pp. 1201–1212.

[11] A. P. WITKIN, *Scale space filtering*, Proc. Int. Joint Conf. Artificial Intell., (1983), pp. 1019–1021.

[12] X. G. XIA, C.-C. J. KUO, AND Z. ZHANG, *Signal extrapolation in wavelet subspaces.* to appear in SIAM J. on Scientific Computing.

Figure Captions

Figure 1: Lattice structure of the wavelet coefficients: (a) wavelet lattice points and (b) the corresponding time-frequency domain.

Figure 2: Block diagram of iterative generalized PG algorithm with denoising.

Figure 3: Test problem: (a) the original scale-time limited sequence, and (b) 35 observed data points with additive white Gaussian noise (SNR=9.0dB).

Figure 4: Experimental results: (a) convergence history of two iterative algorithms, (b) noise and noise location detected by denoising algorithm at the second iteration, (c) extrapolated result with denoising process after 200 iterations, (d) extrapolated result with regularization after 200 iterations.

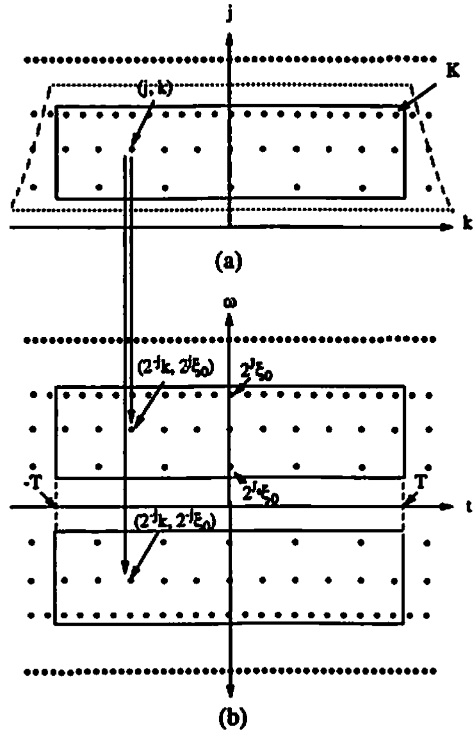


Figure 1: Lattice structure of the wavelet coefficients: (a) wavelet lattice points and (b) the corresponding time-frequency domain.

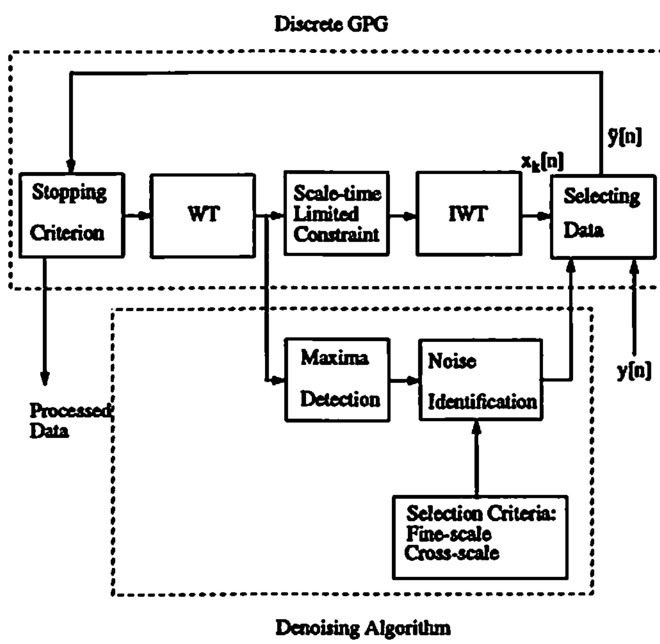
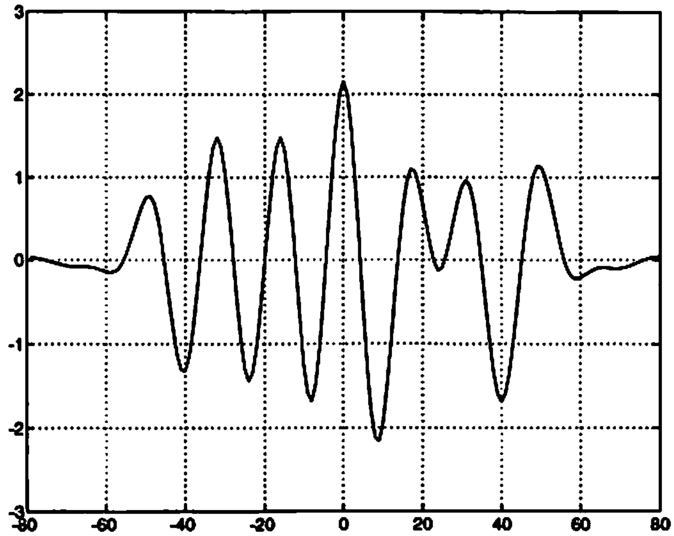
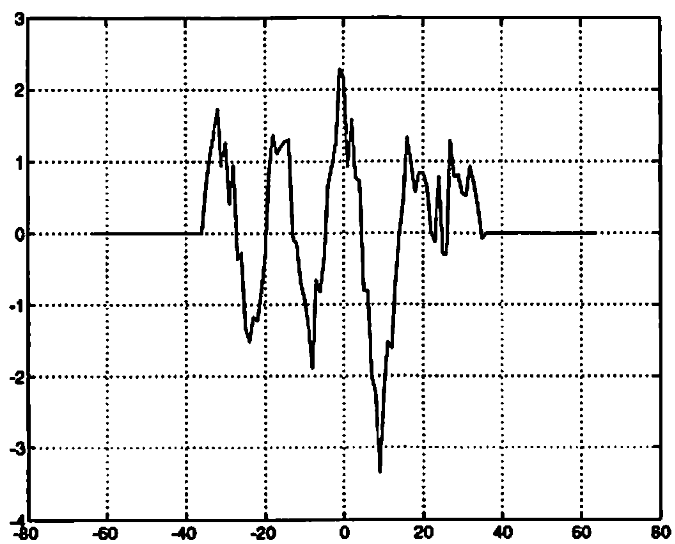


Figure 2: Block diagram of iterative generalized PG algorithm with denoising.



(a)



(b)

Figure 3: Test problems: (a) the original scale-time limited sequence, and (b) 35 observed data points with additive white Gaussian noise (SNR=9.0dB).

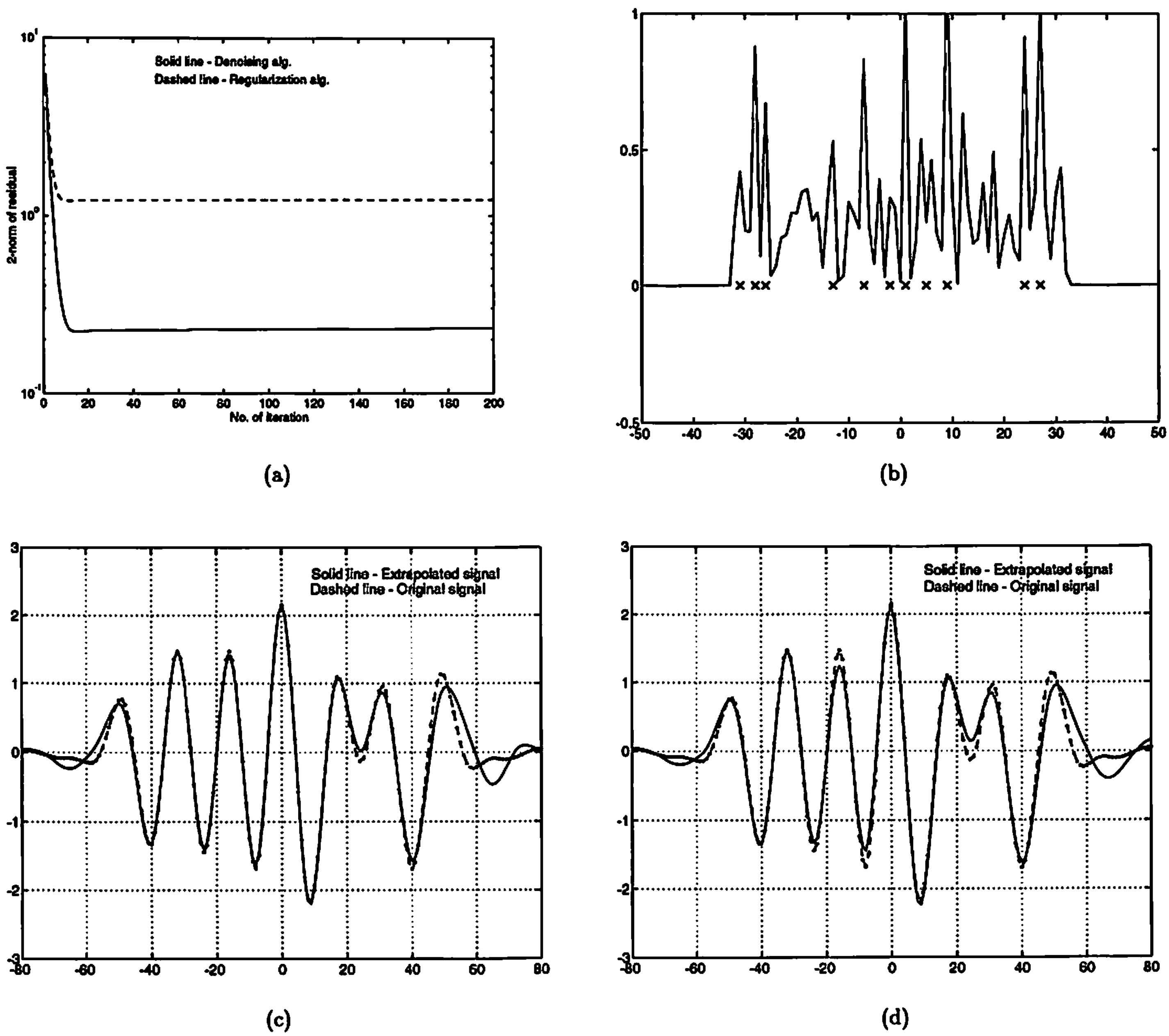


Figure 4: Experimental results: (a) convergence history of two iterative algorithms, (b) noise and noise location detected by denoising algorithm at the second iteration, (c) extrapolated result with denoising process after 200 iterations, (d) extrapolated result with regularization after 200 iterations.