USC–SIPI REPORT #240

On The Convergence of Wavelet-Based Iterative Signal Extrapolation Algorithms

by

Li-Chien Lin, Xiang-Gen Xia and C.-C. Jay Kuo

September 1993

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.

ON THE CONVERGENCE OF WAVELET-BASED ITERATIVE SIGNAL EXTRAPOLATION ALGORITHMS*

LI-CHIEN LIN†, XIANG-GEN XIA† AND C.-C. JAY KUO†

Abstract. A generalized Papoulis-Gerchberg (PG) algorithm for signal extrapolation based on the wavelet representation has been recently proposed by Xia, Kuo and Zhang. In this research, we examine the convergence property and the convergence rate of several signal extrapolation algorithms in wavelet subspaces. We first show that the generalized PG algorithm converges to the minimum norm solution when the wavelet bases are semi-orthogonal (or known as the prewavelet). However, the generalized PG algorithm converges slowly in numerical implementation. To accelerate the convergence rate, we formulate the discrete signal extrapolation problem as a two-step process and apply the steepest descent and conjugate gradient methods for its solution. Numerical experiments are given to illustrate the performance of the proposed algorithms.

Key words. signal extrapolation, Papoulis-Gerchberg algorithm, wavelet.

AMS(MOS) subject classifications. 41, 65D, 65F.

1. Introduction. Extrapolating a band-limited signal $f(t)$ from its values in a finite interval $[-T, T]$ is a fundamental problem in signal reconstruction. Possible applications of signal extrapolation include spectrum estimation, synthetic aperture radar, limited-angle tomography, beamforming and high resolution image restoration. In 70's, Papoulis [13] and Gerchberg [8] proposed an iterative procedure for band-limited signal extrapolation. Numerous techniques to extend the interpolation scheme have been proposed, including the minimum norm least square (MNLS) solution [9], the discrete prolate spheroidal sequence (DPSS) expansion [16], [17], and the weighted norm least square solution [1], [2], [14]. However, all of them were derived from the Fourier transform viewpoint.

More recently, multiresolution wavelet bases with a nice time-frequency localization property have been extensively studied [3], [6], [7], [12] and a generalized PG algorithm based on the wavelet representation has been proposed by Xia, Kuo and Zhang [19]. Instead of using the band-limited signal model, Xia et al. considered a general class of scale-limited signal contained in a certain wavelet subspace. One potential advantage of the generalized PG algorithm is that it provides a more general class of bases for signal modeling. The time-localized wavelet basis should be more suitable than the Fourier basis in modeling signals with interesting transient information such as those arising from the electrocardiogram and radar applications. Furthermore, the band-limited PG algorithm is very sensitive to noise even in the case where only a small amount of extrapolated data are desired [15]. In contrast, noise in the generalized PG algorithm can be detected via the time-localization property of wavelet bases and can be more easily removed [10].

In implementing an iterative extrapolation algorithm, it is natural to ask two basic questions: whether the algorithm converges and what is the converged result. In [19], the convergence of the generalized PG algorithm with orthogonal wavelets were examined. The convergence proof given there only applies to a subset of orthogonal wavelets which excludes some popular bases such as the Daubechies bases. In this work, we provide more complete answers to the above questions. We give a convergence proof for the generalized PG algorithm with semi-orthogonal wavelets, which include all orthogonal wavelets as a subset and are known as the prewavelets, by utilizing the alternating projection theorem. The convergence of the generalized PG algorithm with biorthogonal wavelet bases

* This research was supported by the National Science Foundation Young Investigator (NYI) Award ASC-9258396 and Presidential Faculty Fellow (PFF) Award GER 93-50309.

† Signal and Image Processing Institute, Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-2564. E-mail: llin@sipi.usc.edu, xianggen@solar.usc.edu and cckuo@sipi.usc.edu.

however does not hold in general. Furthermore, we show that the generalized PG algorithm converges to the minimum norm solution of the extrapolation problem.

The generalized PG algorithm converges slowly for the ill-conditioned problem in numerical implementation. To accelerate the convergence rate, we formulate the discrete signal extrapolation problem as a two-step process and apply the steepest descent and conjugate gradient methods for the solution. As a result, we obtain two new iterative algorithms with a better convergence performance for discrete signal extrapolation. The convergence rate of these algorithms is analyzed. Numerical experiments are also given to illustrate the performance of the proposed algorithms.

This paper is organized as follows. We briefly review some basic results of wavelet theory in §2. In §3, a new signal model based on the wavelet representation is presented and used to derive a signal extrapolation algorithm called the generalized PG algorithm (or the scale-time limited extrapolation). Then, we investigate the convergence of the generalized PG algorithm for semi-orthogonal wavelets. Two new iterative algorithms with faster convergence rates are proposed and some convergence analysis is presented in §4. Numerical experiments are given in §5 to show the convergence behavior of the proposed algorithms. Concluding remarks and possible extensions are given in §6.

2. Results from wavelet theory. We review some basic results of biorthogonal wavelet theory below, and refer interesting readers to [4], [5], [7] for more detailed discussion. Let $\phi(t)$ be a biorthogonal scaling function and $\psi(t)$ and $\mathcal{P}_j$ be its associated wavelet function and multiresolution analysis (MRA), respectively. We also use $\tilde{\phi}(t)$, $\tilde{\psi}(t)$ and $\tilde{\mathcal{P}}_j$ to denote their duals. Then,

$$\cdots \subset \mathcal{P}_0 \subset \mathcal{P}_1 \subset \mathcal{P}_2 \cdots$$

with

$$\overline{\bigcup_j \mathcal{P}_j} = L^2(\mathbf{R}), \quad \bigcap_j \mathcal{P}_j = \{0\},$$

and

$$f(t) \in \mathcal{P}_j \text{ if and only if } f(2t) \in \mathcal{P}_{j+1},$$

and for a fixed integer j, $\phi_{jk}(t) \triangleq 2^{j/2}\phi(2^j t - k)$, $j \in \mathbf{Z}$, form a biorthogonal basis of $\mathcal{P}_j$ as

$$\langle \phi_{jk_1}, \tilde{\phi}_{jk_2} \rangle = \delta_{k_1 k_2},$$

where $\delta_{k_1 k_2} = 1$ when $k_1 = k_2$ and 0 otherwise. If we let $\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$, then

$$\langle \psi_{j_1 k_1}, \tilde{\psi}_{j_2 k_2} \rangle = \delta_{k_1 k_2}\delta_{j_1 j_2}.$$

When $\psi(t) = \tilde{\psi}(t)$, the wavelet basis is orthogonal. A wavelet basis is called *semi-orthogonal* (or known as the *prewavelet*), if the wavelet basis function only satisfies

$$\langle \psi_{j_1 k_1}, \psi_{j_2 k_2} \rangle = 0, \quad \text{for} \quad j_1 \neq j_2.$$

Clearly, the set of prewavelet includes the set of orthogonal wavelets as a subset, and the relationship $\mathcal{P}_j = \tilde{\mathcal{P}}_j$ holds for both cases. For more details, see [4], [5].

Any $f(t) \in L^2(\mathbf{R})$ can be decomposed by

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} b_{j,k}\psi_{jk}(t). \tag{2.1}$$

For any $f(t) \in \mathcal{P}_J$, we have

$$f(t) = \sum_{k=-\infty}^{\infty} c_{J,k}\phi_{Jk}(t) = \sum_{j<J} \sum_{k=-\infty}^{\infty} b_{j,k}\psi_{jk}(t), \tag{2.2}$$

where $b_{j,k} \triangleq \langle f, \tilde{\psi}_{jk} \rangle$ and $c_{j,k} \triangleq \langle f, \tilde{\phi}_{jk} \rangle$. The $b_{j,k}$ in (2.1) are called the wavelet series transform (WST) coefficients of $f(t)$, and (2.1) provides the inverse wavelet series transform of $b_{j,k}$. Multiresolution analysis leads naturally to a hierarchical and fast scheme for the computation of wavelet coefficients $b_{j,k}$ with $j < J$ which can be obtained from coefficients $c_{J,k}$ by the recursive formulas:

$$\begin{array}{rcl} c_{j-1,k} & = & \sqrt{2}\sum_n \tilde{h}_{n-2k} c_{j,n}, \\ b_{j-1,k} & = & \sqrt{2}\sum_n \tilde{g}_{n-2k} c_{j,n}, \end{array} \tag{2.3}$$

for $j = J, J-1, J-2, \cdots$. The synthesis formula which compute coefficients $c_{J,k}$ from $c_{J_0,k}$ and $b_{j,k}$ with $J_0 \leq j < J$ is

$$c_{j+1,n} = \sqrt{2}\left(\sum_k h_{n-2k} c_{j,k} + \sum_k g_{n-2k} b_{j,k}\right), \quad \text{for} \quad j = J_0, J_0+1, \cdots, J-1. \tag{2.4}$$

In practice, $c_{J,n}$ can be viewed as a sequence of $x[n]$, sampled of a signal $f(n/2^J)$. Then (2.3) and (2.4) are called the discrete wavelet transform (DWT) and the inverse discrete wavelet transform (IDWT), respectively.

3. Scale-time limited signal extrapolation. A new signal modeling scheme based on wavelet representation is described and applied to the signal extrapolation problem in this section.

3.1. Scale-time limited signals. We represent $f(t)$ with the wavelet basis ψ_{jk} in (2.1). 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 domains. Then, by using the scaling property, $\psi_{jk}(t)$ is localized around $2^{-j}k$ in the time domain and $\pm 2^j \xi_0$ in the frequency domain. 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. This concept is illustrated in Fig. 1, where the dot (j,k) in the upper plot denotes the time and scale indices of a certain wavelet coefficient while the dots in the lower plot denote its influence in the time and frequency domains. Now, suppose that the energy of $f(t)$ is well concentrated in two rectangle regions as shown in Fig. 1, i.e.

$$[-T_0, T_0] \times \ [(-2^J\xi_0, -2^{J_0}\xi_0) \cup (2^{J_0}\xi_0, 2^J\xi_0)], \tag{3.1}$$

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

$$\int_{2^{J_0}\xi_0 \leq |\xi| \leq 2^J \xi_0} |\hat{f}|^2 \, d\xi \geq (1-\epsilon)\|f\|^2,$$

and

$$\int_{|x| \leq T_0} |f|^2 \, dx \geq (1-\epsilon)\|f\|^2.$$

Then, only the wavelet coefficients $b_{j,k}, (j,k) \in D$, where

$$D = \{(j,k) \in \mathbf{Z}^2 : J_0 \leq j < J \text{ and } 2^{-j} \mid k \mid \leq T_0 + t_\epsilon\}$$

is the set of dyadic points enclosed by the solid rectangle in Fig. 1 and t_ϵ is a constant, are needed for a good approximation of $f(t)$. We refer to [7] for a more detailed discussion of this approximation.

It is easy to see from Fig. 1 that the wavelet representation gives a higher resolution to sharp varying components than smooth components. This feature is ideal for signals composed of high frequency variations in short duration and low frequency variations in long duration. This kind of signals occurs frequently in practice. Let $\mathcal{K} = \{k : |k| \leq \mathrm{K} = 2^J(T_0 + t_\epsilon)\}$. We have from the above analysis that

$$f(t) \approx \sum_{J_0 \leq j \leq J} \sum_{k \in \mathcal{K}} b_{j,k} \psi_{jk}(t).$$

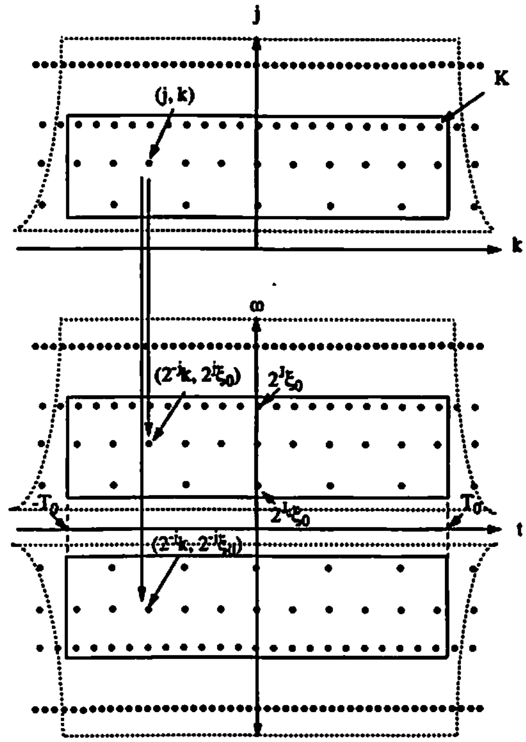


FIG. 1. *Lattice structure of the wavelet coefficients.*

Thus, the following space

$$\mathbf{V}_{J_0,J,\mathcal{K}} = \{f(t): \ f(t) = \sum_{J_0 \le j \le J} \sum_{k \in \mathcal{K}} b_{j,k} \psi_{jk}(t)\}. \tag{3.2}$$

provides a good model for signals concentrated in (3.1).

Assume that $f(t) \in \mathbf{V}_{J_0,J,\mathcal{K}}$ and $c_{J,k} = \int_{-\infty}^{\infty} f(t)\phi_{Jk}(t)dt$. Then, the DWT coefficients of $c_{J,k}$ are b_{jk} for $J_0 \le j < J$ and $k \in \mathcal{K}$, and 0 otherwise. In other words, for $x[k] = c_{J,k}$, it satisfies

$$x[k] = (\mathbf{D}_{J_0,J}^{-1} \mathbf{T}_{J,\mathcal{K}} \mathbf{D}_{J_0,J} x)[k], \qquad k \in \mathbf{Z},$$

where $\mathbf{D}_{J_0,J}^{-1}$ and $\mathbf{D}_{J_0,J}$ are, respectively, the DWT and IDWT operators and $\mathbf{T}_{J,\mathcal{K}}$ is the following projection operator

$$\mathbf{T}_{J,\mathcal{K}} u_{j,k} = \begin{cases} u_{j,k}, & \text{if } J_0 \le j < J, k \in \mathcal{K}, \\ 0, & \text{otherwise.} \end{cases}$$

In general, since the behavior of $\phi(t)$ is like a low pass filter, $c_{J,k} \approx 2^{-J/2} f(\frac{k}{2^J})$. By setting $x[k] = 2^{-J/2} f(\frac{k}{2^J})$, we have

$$x[k] \approx (\mathbf{D}_{J_0,J}^{-1} \mathbf{T}_{J,\mathcal{K}} \mathbf{D}_{J_0,J} x)[k].$$

We conclude from the above discussion that the signal set

$$\mathbf{S}_{J_0,J,\mathcal{K}} = \{x[k]: \ x[k] = (\mathbf{D}_{J_0,J}^{-1} \mathbf{T}_{J,\mathcal{K}} \mathbf{D}_{J_0,J} x)[k], \quad k \in \mathbf{Z}\} \tag{3.3}$$

provides a good approximation model for discrete-time signals with energy concentrated in (3.1).

Let $\mathbf{x}$ denote the vector of the sequence $x[k]$. We call $x[k]$ a $(J_0, J; \mathcal{K})$ *scale-time limited sequence* if

$$x[k] = \mathbf{D}_{J_0,J}^{-1} \mathbf{T}_{J,\mathcal{K}} \mathbf{D}_{J_0,J} x[k], \quad \text{or} \quad \mathbf{x} = \mathbf{D}_{J_0,J}^{-1} \mathbf{T}_{J,\mathcal{K}} \mathbf{D}_{J_0,J} \mathbf{x}, \tag{3.4}$$

where $\mathbf{D}^{-1}$, $\mathbf{D}$ and $\mathbf{T}_{J,\mathcal{K}}$ in the second expression are all matrices of dimension $\infty \times \infty$. Let

$$L \triangleq |\mathcal{K}| \cdot (J - J_0 + 1),$$

which is the number of possible non-zero wavelet transform coefficients of x[k]. Then, without loss of generality, we can express $\mathbf{T}_{J,\mathcal{K}}$ as

$$\mathbf{T}_{J,\mathcal{K}} = \mathbf{U}^T\mathbf{U},$$

where $\mathbf{U} = \{u_{ij}\}$ is a $L \times \infty$ matrix operator

$$u_{ij} = \begin{cases} 1, & \text{if } i = j, \text{ and } 1 \le i, j \le L, \\ 0, & \text{otherwise.} \end{cases}$$

3.2. Generalized PG algorithm. The generalized PG algorithms for continuous- and discrete-time proposed by Xia, Kuo and Zhang in [19] are summarized below.

We first examine the continuous-time case. Let $\mathbf{P}_J, \mathbf{P}_T, \mathbf{Q}_J, \mathbf{Q}_T$ denote the projection operators which project functions onto the subspaces $\mathcal{P}_J$, $\mathcal{P}_T$, $\mathcal{P}_J^{\perp}$, and $\mathcal{P}_T^{\perp}$, respectively, where $\mathcal{P}_J$ is the wavelet subspace as defined in §2 and $\mathcal{P}_T$ a set consisting of all functions $f(t) \in L^2(\mathbf{R})$ with $f(t) = 0$ outside $[-T, T]$. We see from the representations (2.1)-(2.2) that for any $f \in L^2(\mathbf{R})$

$$\mathbf{P}_J f(t) = \sum_{k=-\infty}^{\infty} c_{J,k}\phi_{Jk}(t) = \sum_{j<J}\sum_{k=-\infty}^{\infty} b_{j,k}\psi_{jk}(t),$$

and

$$\mathbf{Q}_J f(t) = \sum_{j\ge J}\sum_{k=-\infty}^{\infty} b_{j,k}\psi_{jk}(t).$$

Now, given a scale-limited function $f(t) \in \mathbf{V}_{J_0,J,\mathcal{K}}$, the generalized PG algorithm recovers $f(t)$ from its segment $g(t) = \mathbf{P}_T f(t)$ with $T < T_0$ via the following iteration:

$$\begin{aligned} f^{(0)}(t) &= g(t), \\ f^{(l+1)}(t) &= \mathbf{Q}_T\mathbf{P}_J f^{(l)}(t) + f^{(0)}(t), \quad l = 0, 1, 2, \cdots. \end{aligned} \tag{3.5}$$

In [19], the following result on the convergence of the generalized PG algorithm was obtained for orthogonal wavelets.

PROPOSITION 1. *Let $\phi(t)$ be an orthogonal scaling function. If*

$$Q_J(s,t) \triangleq \sum_k \phi_{Jk}(s)\phi_{Jk}(t). \tag{3.6}$$

is continuous and positive definite in the region $[-T,T] \times [-T,T]$ and moreover $\phi(t)$ can be uniquely determined in V_J by any one of its segments $\phi(t)$, $t \in [-2^J T - k, 2^J T - k]$, $k \in \mathbf{Z}$, then then $\|f^{(l)} - f\| \to 0$ as $l \to \infty$.

The proof was based on theory of adjoint operators corresponding to symmetric kernels $Q_J(s,t)$. It is in general not easy to check the convergence condition of this proposition for a given arbitrary orthogonal wavelet. Some non-trivial orthogonal wavelet bases were verified to satisfy the convergence condition [19]. However, the convergence condition described only applies to a subset of orthogonal wavelets which excludes some popular bases such as the Daubechies bases.

Next, we examine the discrete-time case. Let

$$\mathbf{T}_{\mathcal{N}} x[n] = \begin{cases} x[n], & n \in \mathcal{N}, \\ 0, & n \notin \mathcal{N}, \end{cases}$$

be the projection operator and I be the identity operator. Given a segment $\mathbf{T}_{\mathcal{N}} x[n]$, $n \in \mathcal{N}$, of a scale-limited sequence $x[n] \in \mathbf{S}_{J_0,J,\mathcal{K}}$, the discrete generalized PG algorithm determines $x[n]$ with $n \notin \mathcal{N}$ and can be stated as:

$$\begin{aligned} x^{(0)}[n] &= \mathbf{T}_{\mathcal{N}} x[n], \\ x^{(l+1)}[n] &= \mathbf{T}_{\mathcal{N}} x[n] + (\mathbf{I} - \mathbf{T}_{\mathcal{N}})\mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}\mathbf{D}_{J_0,J} x^{(l)}[n], \quad l = 0, 1, 2, \cdots, \end{aligned} \tag{3.7}$$

A condition for the convergence of the above iterative procedure was also provided in [19].

3.3. Convergence of the Generalized PG algorithm. We will provide a more general convergence condition of the generalized PG algorithm, and examine the uniqueness of the corresponding extrapolated signal for semi-orthogonal wavelet bases in this subsection.

For the continuous-time case, we have the following convergence theorem.

THEOREM 1. *Let $\phi(t)$ be a semi-orthogonal scaling function and $f^{(l)}(t)$ be the sequence of functions generated via iteration (3.5) with $f \in \mathcal{P}_J$ for a certain $J > 0$. Then, when $l \to \infty$, $f^{(l)}(t)$ converges to the minimum norm solution $f^{\dagger}$ satisfying*

$$f^{\dagger} \in \mathcal{P}_J, \quad \text{and} \quad \|f^{\dagger}\| = \min\{\|h\| : h \in \mathcal{P}_J, \mathbf{P}_T h = g\}.$$

Proof. Any $f(t) \in \mathcal{P}_J$ and $g(t) = \mathbf{P}_T f(t)$ can be written as

$$f(t) = g + h_1, \tag{3.8}$$

where

$$g \in \mathcal{P}_T, \quad \text{and} \quad h_1 \in \mathcal{P}_T^{\perp}.$$

Since $h_1 = \mathbf{Q}_T f$ and $f = \mathbf{P}_J f$, we can rewrite (3.8) as

$$f = g + \mathbf{Q}_T \mathbf{P}_J (g + h_1).$$

By substituting h_1 with $\mathbf{Q}_T \mathbf{P}_J f$ and decomposing f into $g + h_1$ repeatedly, we have

$$\begin{aligned} f &= g + \mathbf{Q}_T\mathbf{P}_J g + (\mathbf{Q}_T\mathbf{P}_J)^2 g \cdots + (\mathbf{Q}_T\mathbf{P}_J)^l g + (\mathbf{Q}_T\mathbf{P}_J)^l h_1 \\ &= f^{(l)} + (\mathbf{Q}_T\mathbf{P}_J)^l \mathbf{Q}_T f, \quad l \to \infty, \end{aligned}$$

where the last equality is due to (3.5). By the definition of semi-orthogonal wavelets, the operator $\mathbf{P}_J$ is an orthogonal projection. The operator $\mathbf{Q}_T$ is also clearly an orthogonal projection. By using the Alternating Projection Theorem (Theorem 13.7 in [18]), we have

$$\lim_{l\to\infty} (\mathbf{Q}_T\mathbf{P}_J)^l \mathbf{Q}_T f = \mathbf{G} f,$$

where $\mathbf{G}$ is the orthogonal projection operator onto $\mathcal{P}_J \cap \mathcal{P}_T^{\perp}$, which is a linear subspace of $L^2(\mathbf{R})$. Therefore, the generalized PG algorithm converges to

$$f^{\dagger} = lim_{l\to\infty} f^{(l)} = f - \mathbf{G}f.$$

This proves the convergence. Due to the orthogonality of $\mathbf{G}$, we get

$$\|f - \mathbf{G}f\| = \min_{f_1 \in \mathcal{P}_J \cap \mathcal{P}_T^{\perp}} \|f - f_1\|.$$

For any $h \in \mathcal{P}_J$ and $\mathcal{P}_T h = \mathcal{P}_T f = g$, $f - h \in \mathcal{P}_J \cap \mathcal{P}_T^{\perp}$. Then, since $\|f^{\dagger}\| = \|f - \mathbf{G}f\| \leq \|f - (f-h)\| = \|h\|$, $f^{\dagger}$ is the minimum norm solution. □

With this theorem, we have the following straightforward corollary.

COROLLARY 1. *If $\mathcal{P}_J \cap \mathcal{P}_T^{\perp} = \{0\}$, $f^{(l)}(t)$ converges to $f(t)$ in $L^2(\mathbf{R})$ as $l \to \infty$.*

If the semi-orthogonal scaling function $\phi(t)$ is band-limited, all signals in $\mathcal{P}_J$ are band-limited and therefore analytic. For this case, the condition in Corollary 1 is satisfied, and another corollary follows.

COROLLARY 2. *If the scaling function is semi-orthogonal and band-limited, $f^{(l)}(t)$ converges to $f(t)$ in $L^2(\mathbf{R})$ as $l \to \infty$.*

Note also that since all Meyer wavelets satisfy Corollary 2, Theorem 4 in [19] is in fact a special case of the result derived above.

Next, we will examine the discrete-time case. When the wavelet basis is orthogonal, $\mathbf{D}_{J_0,J}^{-1}$ in (3.7) is the transpose of $\mathbf{D}_{J_0,J}$. Thus, the operator $\mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}\mathbf{D}_{J_0,J}$ is an orthogonal projection. By applying the Alternating Projection Theorem to (3.7), we can prove the convergence of the iterative procedure

(3.7). However, the operator $\mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}\mathbf{D}_{J_0,J}$ may not be an orthogonal projection when the wavelet basis is non-orthogonal. This is stated in the following theorem.

THEOREM 2. *For any fixed J_0, J and $\mathcal{K}$ the operator $\mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}\mathbf{D}_{J_0,J}$ is an orthogonal projection if and only if the wavelet basis is orthogonal.*

Proof. Since the sufficient part is straightforward, we only show the necessary part, i.e, to prove that if $\mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}\mathbf{D}_{J_0,J}$ is an orthogonal projection for any fixed J_0, J and $\mathcal{K}$, the wavelet basis is orthogonal. This is equivalent to prove that $\mathbf{D}_{J_0,J}^{-1}$ is equal to the transpose $\mathbf{D}_{J_0,J}^{T}$ of $\mathbf{D}_{J_0,J}$ for any fixed J_0 and J.

Recall (3.2) that we use $\mathbf{V}_{J_0,J,\mathcal{K}}$ to denote the set of all $(J_0, J; \mathcal{K})$ scale-time limited sequences. Then, the operator $\mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}\mathbf{D}_{J_0,J}$ is an orthogonal projection onto $\mathbf{V}_{J_0,J,\mathcal{K}}$. Thus, any $x \in l^2$ can be decomposed as

$$x = x_1 + x_2, \quad \text{where} \quad x_1 = \mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}y, \text{ and } x_2 = \mathbf{D}_{J_0,J}^{-1}(y - \mathbf{T}_{J,\mathcal{K}}y).$$

for certain $y \in l^2$. By using the orthogonality of the operator $\mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}\mathbf{D}_{J_0,J}$, we have

$$\langle x_1, x_2\rangle = 0, \quad \text{for any } y \in l^2.$$

Let b_{ij} denote the (i,j) element of the matrix $\mathbf{D}_{J_0,J}^{-1}$. Let us partition all integers into two nonoverlapping groups $\mathcal{N}_1$ and $\mathcal{N}_2$, i.e. $Z = \mathcal{N}_1 \cup \mathcal{N}_2$ and $\mathcal{N}_1 \cap \mathcal{N}_2 = \emptyset$. Then, this leads to

$$\sum_i \sum_{j_1 \in \mathcal{N}_1} b_{ij_1} y[j_1] \sum_{j_2 \in \mathcal{N}_2} b_{ij_2} y[j_2] = 0,$$

or

$$\sum_{j_1 \in \mathcal{N}_1} \sum_{j_2 \in \mathcal{N}_2} \left(\sum_i b_{ij_1} b_{ij_2} \right) y[j_1] y[j_2] = 0.$$

This equality implies

$$\sum_i b_{ij_1} b_{ij_2} = 0, \quad \text{for } j_1 \neq j_2, \quad j_1, j_2 \in Z, \tag{3.9}$$

since y can be any element in l^2 and $\mathcal{N}_1$ can be any subset of $\mathbf{Z}$. As a direct consequence of (3.9), $\mathcal{D}_{J_0,J}$ is orthogonal. □

For a general biorthogonal wavelet basis, the projection $\mathbf{P}_J$ in algorithm (3.5) or (3.7) is a nonorthogonal projection. The generalized PG algorithm in both continuous- and discrete-time cases assume the form

$$f^{(l)} = f - (\mathbf{TP})^l f, \; l = 0, 1, 2, \cdots,$$

where $\mathbf{T}$ is a truncation operator and $\mathbf{P}$ is a nonorthogonal projection. To check the convergence of the above iterative procedure, it is important to examine the norm of the operator $\mathbf{TP}$ as stated in the following proposition.

PROPOSITION 2. *If the operator $\mathbf{P}$ is a nonorthogonal projection, there is a truncation operator $\mathbf{T}$ such that $\|\mathbf{TP}\| > 1$.*

Proof: We only have to prove that there is a nonzero signal x such that $\|\mathbf{P}x\| > \|x\|$. Since $\mathbf{P}$ is a nonorthogonal projection, we can find an $x = \mathbf{P}x + y$ with $\langle \mathbf{P}x, y\rangle < 0$ such that $2\langle \mathbf{P}x, y\rangle + \|y\|^2 < 0$. Then,

$$\|x\|^2 = \|\mathbf{P}x\|^2 + \|y\|^2 + 2\langle \mathbf{P}x, y\rangle < \|\mathbf{P}x\|^2,$$

and the proposition is proved. □

Because of Proposition 2, we do not expect the convergence of the generalized PG algorithm for continuous-time signal extrapolation with biorthogonal wavelet bases. Furthermore, because of Theorem 2 and Proposition 2, we do not expect the convergence of the generalized PG algorithm for discrete-time signal extrapolation with nonorthogonal (including biorthogonal and semi-orthogonal) wavelet bases.

4. Iterative algorithms for discrete signal extrapolation. In this section, we formulate the discrete-time extrapolation problem as a two-step process and apply more efficient numerical algorithms such as the steepest descent and conjugate gradient methods to improve the convergence rate of the iteration.

4.1. Problem formulation. Let $\mathbf{y} = \mathbf{T}_{\mathcal{N}}\mathbf{x}$ be a given segement of $\mathbf{x}$. According to the discussion in §3.1, we have

$$\begin{aligned} \mathbf{y} &= \mathbf{T}_{\mathcal{N}}\mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}\mathbf{D}_{J_0,J}\mathbf{x} \\ &= \mathbf{T}_{\mathcal{N}}\mathbf{D}_{J_0,J}^{-1}\mathbf{U}^T\mathbf{U}\mathbf{D}_{J_0,J}\mathbf{x}, \end{aligned}$$

which is to be solved for $\mathbf{x}$ with a given $\mathbf{y}$. Let $\mathbf{p} = \mathbf{U}\mathbf{D}_{J_0,J}\mathbf{x}$ be a vector consisting of L wavelet coefficients of $\mathbf{x}$, and $\mathbf{W} = \mathbf{D}_{J_0,J}^{-1}\mathbf{U}^T$. We can rewrite the above equation as

$$\mathbf{y} = \mathbf{T}_{\mathcal{N}}\mathbf{W}\mathbf{p}.$$

By multipling both sides with $(\mathbf{T}_{\mathcal{N}}\mathbf{W})^T$, we obtain the normal equation

$$\mathbf{W}^T\mathbf{y} = \mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}\mathbf{p}, \tag{4.1}$$

where the equlalities $\mathbf{T}_{\mathcal{N}}\mathbf{y} = \mathbf{y}$ and $(\mathbf{T}_{\mathcal{N}})^T\mathbf{T}_{\mathcal{N}} = \mathbf{T}_{\mathcal{N}}$ are used to simplify the result. Furthmore, note that since $\mathbf{x}$ is a scale-time limited sequence, i.e.

$$\mathbf{x} = \mathbf{D}_{J_0,J}^{-1}\mathbf{T}_{J,\mathcal{K}}\mathbf{D}_{J_0,J}\mathbf{x} = \mathbf{D}_{J_0,J}^{-1}\mathbf{U}^T\mathbf{U}\mathbf{D}_{J_0,J}\mathbf{x}$$

we can obtain $\mathbf{x}$ from $\mathbf{p}$ via

$$\mathbf{x} = \mathbf{W}\mathbf{p}. \tag{4.2}$$

Therefore, we can divide the solution procedure of determining $\mathbf{x}$ into two steps: first solving the normal equation for $\mathbf{p}$ and then determining $\mathbf{x}$ from $\mathbf{p}$ as described by (4.1) and (4.2), respectively.

In the following discussion, we assume that the $L \times L$ matrix $\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}$ is of full rank and that the wavelet basis under considertation is orthogonal. Some useful properties of the operators in (4.1)-(4.2) are summarized below:

Property 1: For the orthogonal wavelet bases, $\mathbf{D}_{J_0,J}^{-1} = \mathbf{D}_{J_0,J}^T$ so that the scale-time limited sequence $\mathbf{x}$ can be written as

$$\mathbf{x} = \mathbf{D}_{J_0,J}^{-1}\mathbf{U}^T\mathbf{U}\mathbf{D}_{J_0,J}\mathbf{x} = \mathbf{W}\mathbf{W}^T\mathbf{x}. \tag{4.3}$$

Property 2: From the definition of $\mathbf{W}$, we have

$$\mathbf{W}^T\mathbf{W} = \mathbf{U}\mathbf{D}_{J_0,J}\mathbf{D}_{J_0,J}^{-1}\mathbf{U}^T = \mathbf{U}\mathbf{U}^T = \mathbf{I}_L, \tag{4.4}$$

where $\mathbf{I}_L$ is an identity matrix of dimension $L \times L$.

Property 3: The operator $\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}$ is symmetric positive definite. The symmetric semipositive definiteness of $\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}$ can be easily seen. The positiveness is due to the assumption that $\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}$ is of full rank.

Property 4: Let $\lambda_{\max}(\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W})$ denote the largest eigenvalue of $\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}$. Then,

$$\lambda_{\max}(\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}) \leq 1. \tag{4.5}$$

This can be proved by noting that

$$\lambda_{\max}(\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}) = \max_{\mathbf{z}\neq 0} \frac{(\mathbf{z}^*)^T\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}\mathbf{z}}{(\mathbf{z}^*)^T\mathbf{z}},$$

and

$$(\mathbf{z}^*)^T\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}\mathbf{z} = (\mathbf{z}^*)^T\mathbf{U}\mathbf{D}_{J_0,J}\mathbf{T}_{\mathcal{N}}\mathbf{D}_{J_0,J}^{-1}\mathbf{U}^T\mathbf{z} \leq (\mathbf{z}^*)^T\mathbf{U}\mathbf{D}_{J_0,J}\mathbf{D}_{J_0,J}^{-1}\mathbf{U}^T\mathbf{z} = (\mathbf{z}^*)^T\mathbf{U}\mathbf{U}^T\mathbf{z} = (\mathbf{z}^*)^T\mathbf{z}.$$

There are two reasons to avoid solving (4.1) and (4.2) with direct methods. First, direct computation of the matrix $\mathbf{W}$ is expensive. Second, if the matrix is ill-conditioned, the direct method is usually unstable. Therefore, we consider the solution of (4.1) and (4.2) with iterative algorithms.

4.2. Steepest descent method. The iterative process based on the steepest descent method to solve (4.1)-(4.2) can be stated as: with any given $\mathbf{x}_0$, we perform the following iteration:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{r}_k, \quad \text{for} \quad k = 0, 1, 2, \cdots, \tag{4.6}$$

where

$$\mathbf{r}_k = \mathbf{W}\mathbf{W}^T(\mathbf{T}_{\mathcal{N}}\mathbf{x}_k - \mathbf{y}) \tag{4.7}$$

and

$$\alpha_k = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_k^T \mathbf{T}_{\mathcal{N}} \mathbf{r}_k}. \tag{4.8}$$

We show below how the algorithm given by (4.6)-(4.8) is directly related to the well-known steepest descent method. The application of the steepest descent method [11, pp215] to the solution of the normal equation

$$\mathbf{W}^T\mathbf{y} = \mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}\mathbf{p}$$

is equivalent to the minimization of the cost functional

$$f(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T(\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W})\mathbf{p} - \mathbf{p}^T(\mathbf{W}^T\mathbf{y}).$$

The result can be written as

$$\mathbf{p}_{k+1} = \mathbf{p}_k - \alpha_k \mathbf{d}_k, \tag{4.9}$$

where the vector

$$\mathbf{d}_k = \mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}\mathbf{p}_k - \mathbf{W}^T\mathbf{y},$$

is the gradient direction of the cost functional at point $\mathbf{p}_k$ and

$$\alpha_k = \frac{\mathbf{d}_k^T\mathbf{d}_k}{\mathbf{d}_k^T\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}\mathbf{d}_k} \tag{4.10}$$

is determined by $\min_\alpha f(\mathbf{p}_k + \alpha\mathbf{d}_k)$. Premultiplying both side of (4.9) with $\mathbf{W}$ and applying (4.2), we have

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\mathbf{W}\mathbf{W}^T(\mathbf{T}_{\mathcal{N}}\mathbf{x}_k + \mathbf{y}). \tag{4.11}$$

Thus, we can justify (4.7). Furthermore, due to $\mathbf{W}^T\mathbf{W} = \mathbf{I}_L$ (Property 2), we have

$$\begin{aligned}
\mathbf{d}_k^T\mathbf{d}_k &= (\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{x}_k - \mathbf{W}^T\mathbf{y})^T(\mathbf{W}^T\mathbf{W})(\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{x}_k - \mathbf{W}^T\mathbf{y}) \\
&= (\mathbf{W}\mathbf{W}^T(\mathbf{T}_{\mathcal{N}}\mathbf{x}_k - \mathbf{y}))^T(\mathbf{W}^T\mathbf{W}(\mathbf{T}_{\mathcal{N}}\mathbf{x}_k - \mathbf{y})) \\
&= \mathbf{r}_k^T\mathbf{r}_k,
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{d}_k^T\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}\mathbf{d}_k &= (\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{x}_k - \mathbf{W}^T\mathbf{y})^T(\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W})(\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{x}_k - \mathbf{W}^T\mathbf{y}) \\
&= (\mathbf{W}\mathbf{W}^T(\mathbf{T}_{\mathcal{N}}\mathbf{x}_k - \mathbf{y}))^T\mathbf{T}_{\mathcal{N}}(\mathbf{W}\mathbf{W}^T(\mathbf{T}_{\mathcal{N}}\mathbf{x}_k - \mathbf{y})) \\
&= \mathbf{r}_k^T\mathbf{T}_{\mathcal{N}}\mathbf{r}_k,
\end{aligned}$$

so that (4.8) can also be justified.

The discrete generalized PG algorithm discribed in §3.2 is in fact a special case of the steepest descent algorithm by choosing $\alpha_k = 1$. To see this, let $\alpha_k = 1$ in (4.7), then we have the iterative process:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{W}\mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{x}_k + \mathbf{W}\mathbf{W}^T\mathbf{y}.$$

Since $\mathbf{x}_k$ is a scale-time limited sequence, the above iteration is equivalent to

$$\left\{\begin{array}{lcl} \mathbf{x}'_{k+1} & = & \mathbf{x}_k - \mathbf{T}_{\mathcal{N}}\mathbf{x}_k + \mathbf{y} \\ \mathbf{x}_k & = & \mathbf{W}\mathbf{W}^T\mathbf{x}'_k \end{array}\right.$$

As a consequence, we have

$$\mathbf{x}'_{k+1} = (\mathbf{I} - \mathbf{T}_{\mathcal{N}})\mathbf{W}\mathbf{W}^T\mathbf{x}'_k + \mathbf{y}$$

This is exactly the discrete generalized PG algorithm. Although both the discrete generalized PG algorithm (3.6) and the steepest descent method (4.6) search along the gradient direction of the cost functional, the steepest descent method adjusts the step size α_k at each iteration for minimization so that the convergence rate can be improved. This is confirmed by numerical experiments as given in §5.

4.3. Conjugate gradient method. It is well known that the convergence rate of the steepest descent method can be further improved by that of the conjugate gradient method.

We summarize the conjugate gradient method for solving the system $\mathbf{Q}\mathbf{z} = \mathbf{b}$, where Q is symmetric positive definite below [11, pp244]. Given any $\mathbf{z}_0$ and $\mathbf{d}_0 = \mathbf{b} - \mathbf{Q}\mathbf{z}_0$, we perform the following iteration for $k = 0,\ 1,\ 2, \cdots$:

$$\begin{array}{rcl} \mathbf{g}_k & = & \mathbf{Q}\mathbf{z}_k - \mathbf{b} \\ \mathbf{z}_{k+1} & = & \mathbf{z}_k + \alpha_k \mathbf{d}_k \\ \alpha_k & = & -\dfrac{\mathbf{g}_k^T \mathbf{g}_k}{\mathbf{d}_k^T \mathbf{Q} \mathbf{d}_k} \\ \mathbf{d}_{k+1} & = & -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k \\ \beta_k & = & \dfrac{\mathbf{g}_{k+1}^T \mathbf{g}_{k+1}}{\mathbf{g}_k^T \mathbf{g}_k}. \end{array}$$

We can derive the conjugate gradient method by setting $\mathbf{z} = \mathbf{p}$, $\mathbf{Q} = \mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W}$, and $\mathbf{b} = \mathbf{W}^T\mathbf{y}$. Since the derivation is straightforward, we simply summarize the result below.
Initialization:

$$\mathbf{x}_0 = 0, \quad \tilde{\mathbf{d}}_0 = \mathbf{W}\mathbf{W}^T\mathbf{y}, \quad \tilde{\mathbf{g}}_0 = -\tilde{\mathbf{d}}_0,$$

For $k = 0,\ 1,\ 2, \cdots$,

$$\begin{array}{rcl} \mathbf{x}_{k+1} & = & \mathbf{x}_k + \alpha_k \tilde{\mathbf{d}}_k \\ \alpha_k & = & \dfrac{\tilde{\mathbf{g}}_k^T \tilde{\mathbf{g}}_k}{\tilde{\mathbf{d}}_k^T \mathbf{T}_{\mathcal{N}} \tilde{\mathbf{d}}_k} \\ \tilde{\mathbf{g}}_{k+1} & = & \mathbf{W}\mathbf{W}^T(\mathbf{T}_{\mathcal{N}}\mathbf{x}_{k+1} - \mathbf{y}) \\ \tilde{\mathbf{d}}_{k+1} & = & \beta_k \tilde{\mathbf{d}}_k - \tilde{\mathbf{g}}_{k+1} \\ \beta_k & = & \dfrac{\tilde{\mathbf{g}}_{k+1}^T \tilde{\mathbf{g}}_{k+1}}{\tilde{\mathbf{g}}_k^T \tilde{\mathbf{g}}_k} \end{array}$$

4.4. Convergent rate analysis. To simplify the discussion, we use the notation:

$$\mathbf{Q} = \mathbf{W}^T\mathbf{T}_{\mathcal{N}}\mathbf{W} \quad \text{and} \quad \mathbf{b} = \mathbf{W}^T\mathbf{y}.$$

Let us first examine the convergence rate of the discrete generalized PG algorithm. To solve $\mathbf{Q}\mathbf{p} = \mathbf{b}$ is equivalent to

$$\min_{\mathbf{p}} \frac{1}{2}\mathbf{p}^T\mathbf{Q}\mathbf{x} - \mathbf{p}^T\mathbf{b},$$

which is again equivalent to

$$\min_{\mathbf{p}} E(\mathbf{p}) = \min_{\mathbf{p}} \frac{1}{2}(\mathbf{p} - \hat{\mathbf{p}})^T\mathbf{Q}(\mathbf{p} - \hat{\mathbf{p}}), \tag{4.12}$$

where $\hat{\mathbf{p}}$ be the solution vector $\mathbf{Q}\hat{\mathbf{p}} = \mathbf{b}$. It is easier to analyze the convergent rate for (4.12). The gradient of E is $\mathbf{d} = \mathbf{Q}\mathbf{p} - \mathbf{b}$. Since the discrete generalized PG which can be regarded as the special case of the steepest descent with $\alpha_k = 1$, we have from (4.9)

$$\mathbf{p}_{k+1} = \mathbf{p}_k - \mathbf{d}_k.$$

By direct computation, we have

$$\frac{E(\mathbf{p}_k) - E(\mathbf{p}_{k+1})}{E(\mathbf{p}_k)} = \frac{2\mathbf{d}_k^T \mathbf{Q}\mathbf{u}_k - \mathbf{d}_k^T \mathbf{Q}\mathbf{d}_k}{\mathbf{u}_k^T \mathbf{Q}\mathbf{u}_k},$$

where $\mathbf{u}_k = \mathbf{p}_k - \hat{\mathbf{p}}$. By defining the convergent rate

$$r \triangleq \frac{E(\mathbf{p}_{k+1})}{E(\mathbf{p}_k)},$$

and using the equality $\mathbf{d}_k = Q\mathbf{u}_k$, we have

$$r = \frac{\mathbf{u}_k^T \mathbf{Q}\mathbf{u}_k - 2\mathbf{u}_k^T \mathbf{Q}^2 \mathbf{u}_k + \mathbf{u}_k^T \mathbf{Q}^3 \mathbf{u}_k}{\mathbf{u}_k^T \mathbf{Q}\mathbf{u}_k}.$$

Since $\mathbf{Q}$ is symmetric, it is unitarily diagonalizable with ordered digonals denoted by $\lambda_{\min} = \lambda_1 \le \lambda_2 \cdots \le \lambda_L = \lambda_{\max}$, which are also eigenvalues of $\mathbf{Q}$. Therefore, we have

$$\begin{aligned} r &= \frac{\sum_{i=1}^L \lambda_i u_i^2 - 2\sum_{i=1}^L \lambda_i^2 u_i^2 + \sum_{i=1}^L \lambda_i^3 u_i^2}{\sum_{i=1}^L \lambda_i u_i^2} \\ &= \frac{\sum_{i=1}^L \lambda_i u_i^2 (1-\lambda_i)^2}{\sum_{i=1}^L \lambda_i u_i^2} \\ &= \sum_{i=1}^L p_i (1-\lambda_i)^2, \end{aligned}$$

where

$$p_i = \frac{\lambda_i u_i^2}{\sum_{i=1}^L \lambda_i u_i^2},$$

so that $\sum_{i=1}^L p_i = 1$. Consequently, the convergent rate r can be bounded by

$$r \le \max_{\lambda_i}(1-\lambda_i)^2 = (1-\lambda_{\min})^2$$

For the convergence rate results of the steepest descent and conjugate gradient method, we can take them directly from [11]. They are listed below for comparison. The rate of the steepest descent is bounded by

$$r(\text{steepest descent}) \le \frac{(\lambda_{\max} - \lambda_{\min})^2}{(\lambda_{\max} + \lambda_{\min})^2},$$

where $\lambda_{\max}$ and $\lambda_{\min}$ is the maximum and minimum eigenvalues of Q. The conjugate gradient method converges in at most L steps, where L is the rank of the matrix Q,

5. Experimental results. Numerical examples are given in this section to illustrate the convergence performance of the three iterative signal extrapolation algorithms. We use the orthogonal and compact coiflet basis of order $N = 10$ [7] in our experiments. The basis function is nearly symmetric around the y-axis so that the filter bank implementation consists of almost linear-phase filters. The high order of vanishing moments (i.e. 10) implies its smoothness, and the compact support property makes its implementation easy. Since the convergence behavior heavily depends on the minimum and

TABLE 1

The maximum and minimum eigenvalues and the convergence rate bounds for Test Problem 1.

	λ_{min}	λ_{max}	r_{gpg}	r_{sd}
$M = 15$	0.0000081	0.6263491	0.9999838	0.9999482
$M = 50$	0.2212905	0.9985423	0.6063885	0.4059970

TABLE 2

The maximum and minimum eigenvalues and the convergence rate bounds of the Test problem 2.

	λ_{min}	λ_{max}	r_{gpg}	r_{sd}
$M = 25$	0.0000033	0.8854048	1.0	1.0
$M = 55$	0.2092635	0.9991635	0.6252642	0.4272709

maximum eigenvalues of the matrix $Q = \mathbf{W}^T\mathbf{T}_\mathcal{N}\mathbf{W}$, we consider two test problems with different eigenvalue distributions of Q.

Test problem 1: Consider a scale-time limited sequence $x[n]$ which is 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. 2 (a). For signal modeling, we assume that the scale-time limited information is available to us, i.e. only $b_{j,k}$ with $j = 1$ and $-3 \leq k \leq 4$ are nonzeros. Consequently, the degree of the freedom of the problem is $L = 8$.

We observe $(2M + 1)$ data points $x[n]$ with $|n| \leq M$, and want to extrapolate the values of $x[n]$ for $|n| > M$. By calculating the matrix Q explicitly, we can determine the maximum and minimum eigenvalues of Q and calculate the bounds on the convergence rate of the generalized PG and the steepest descent methods. These values are given in Table 1 with $M = 15$ and 50. It is clear from the table that if we observe more data points, the condition number of the matrix Q becomes smaller and both the generalized PG and the steepest descent methods have faster convergence rates. The improvement of the convergence rate is more significant in the steepest descent case.

The convergence histories of three signal extrapolation algorithms with 31 and 101 (i.e. $M = 15$ and 50) observations are shown in Figs. 3(a) and (b), respectively. For the case $M = 15$, the matrix has a small minimum eigenvalue and a large condition number as indicated in Table 1, the convergence performance of the steepest descent is as poor as that of generalized PG algorithm. In contrast, the conjugate gradient method has a much better convergence performance. For $M = 50$, we see from Fig. 3(b) that the steepest descent method converges more rapidly for this case where the matrix Q has a smaller condition number. It performs better than the generalized PG method as expected from Table 1 and converges almost as fast as the conjugate gradient method. Generally speaking, matrix Q has an decreasing condition number as the number of observations increases, and the convergence rate improvement of the steepest descent and the conjugate gradient methods over the generalized PG method becomes more obvious.

Test problem 2: In this problem, we use the same wavelet basis as in Test Problem 1, but increase the number of nonzero wavelet coefficients so that the degree of freedom of this problem is $L = 12$. The test signal $x[n]$ is generated by randomly choosing the wavelet coefficients $b_{j,k}$ with $J_0 = 0, J = 1$ and $-3 \leq k \leq 4$ (while other wavelet coefficients are set to zero) and observed at the scale $J_s = 4$ as plotted in Fig. 2 (b). The maximum and minimum eigenvalues of Q and the bounds on the convergence rate of the generalized PG and the steepest descent methods for $M = 25$ and 55 are given in Table 2. Finally, the convergence histories of the three methods are given in Fig. 4. We see from Fig. 4(a) that the conjugate gradient method converges much faster than the generalized PG and the steepest descent methods which have about the same convergence rate for small value of M. For $M = 55$, the steepest descent and the conjugate gradient methods have a very similar performance while the generalized PG works poorly as shown in Fig. 4(b).

We may conclude from the two test problems that the conjugate gradient method performs the best among the three methods, the steepest descent method has a good performance when we have more observed data points, and the generalized PG algorithm in general converges very slowly. This

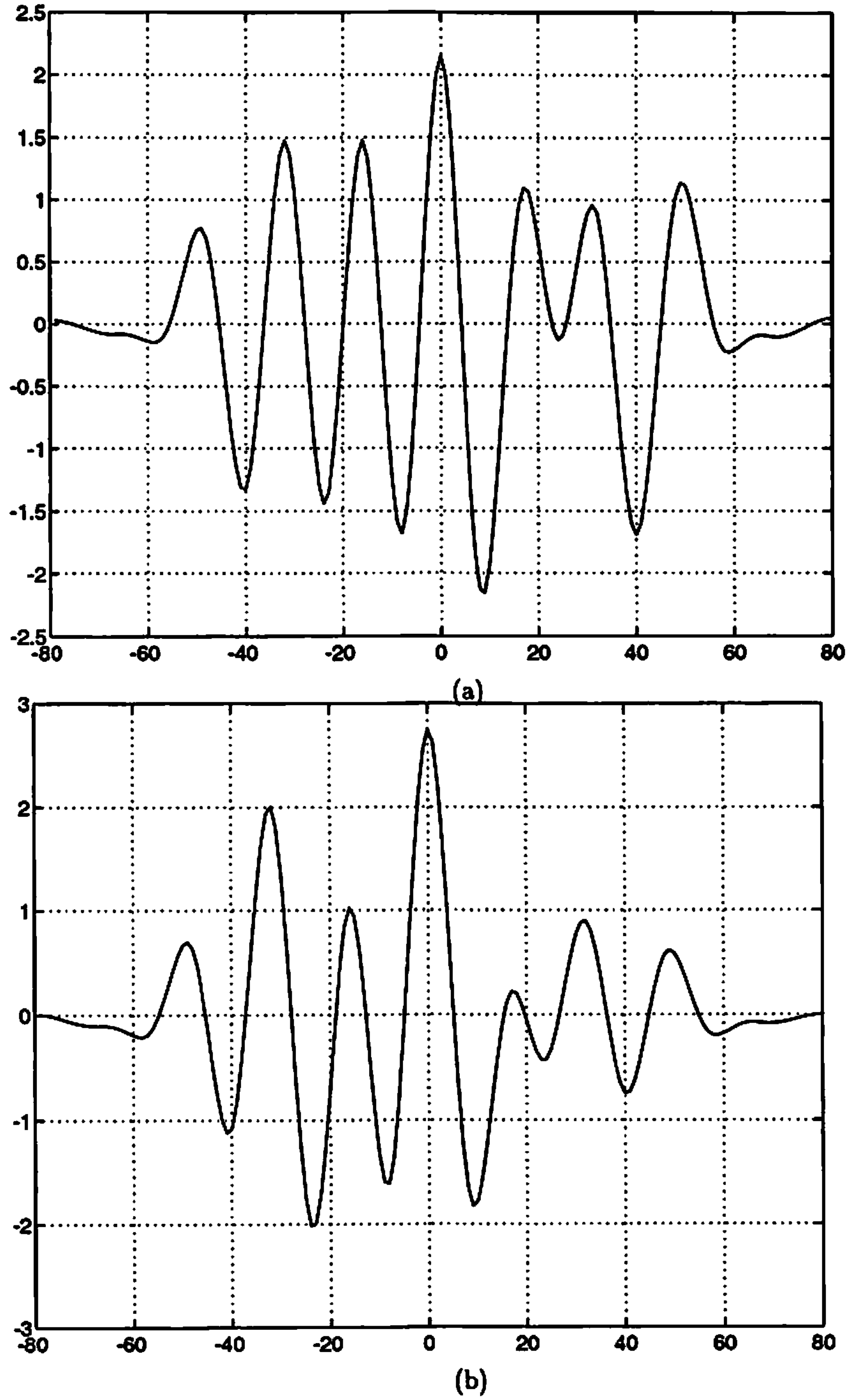
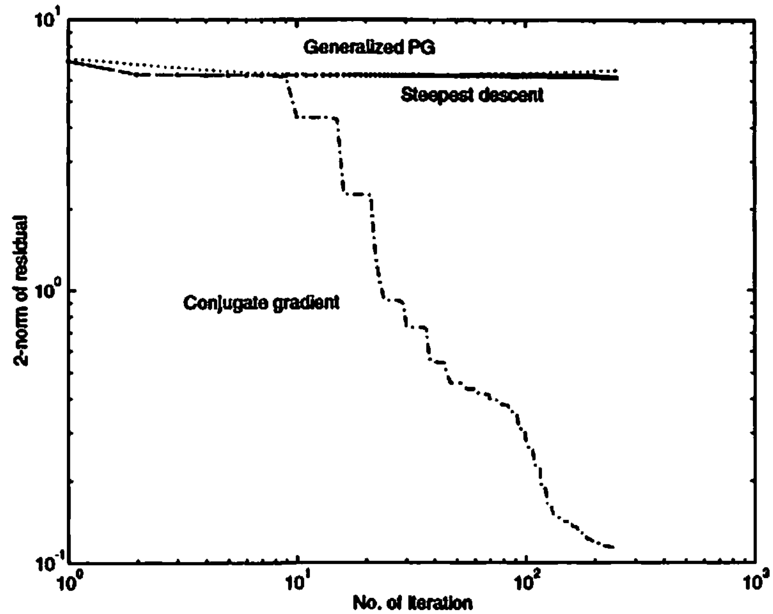
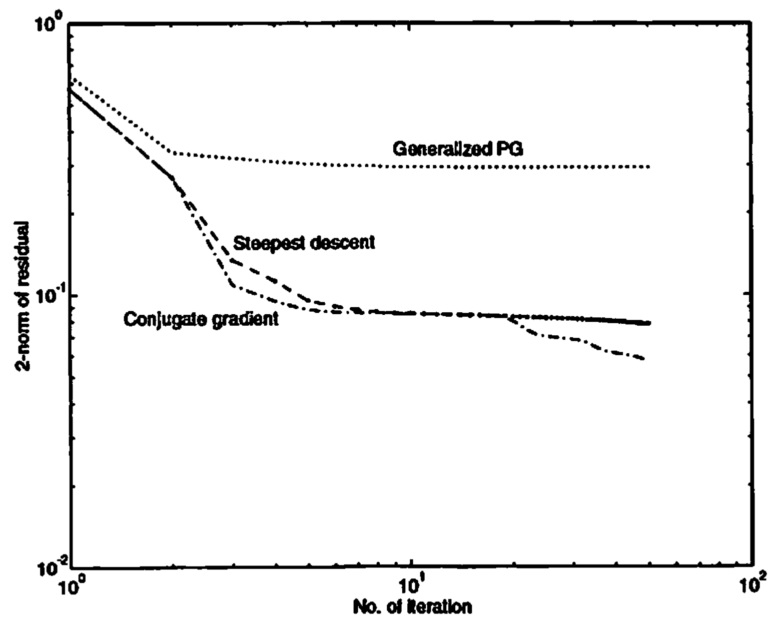


FIG. 2. *The original signals in (a) Test Problem 1 and (b) Test Problem 2.*

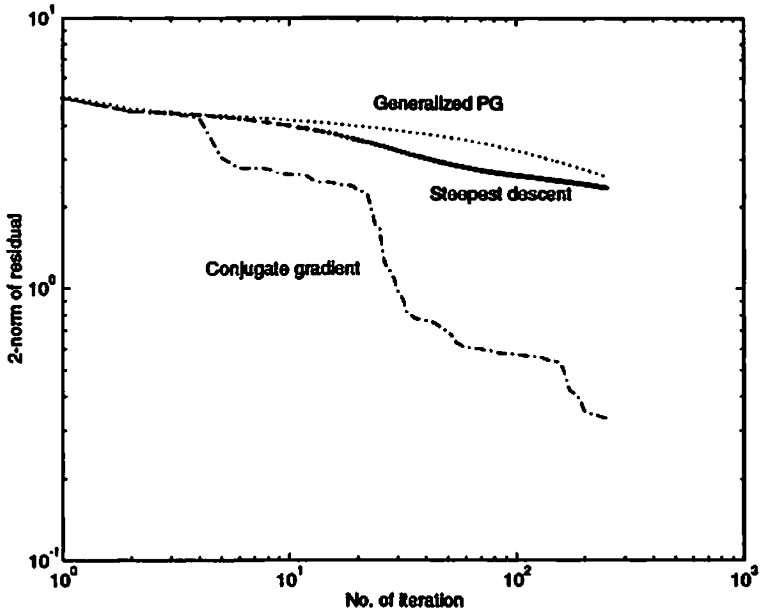


(a)

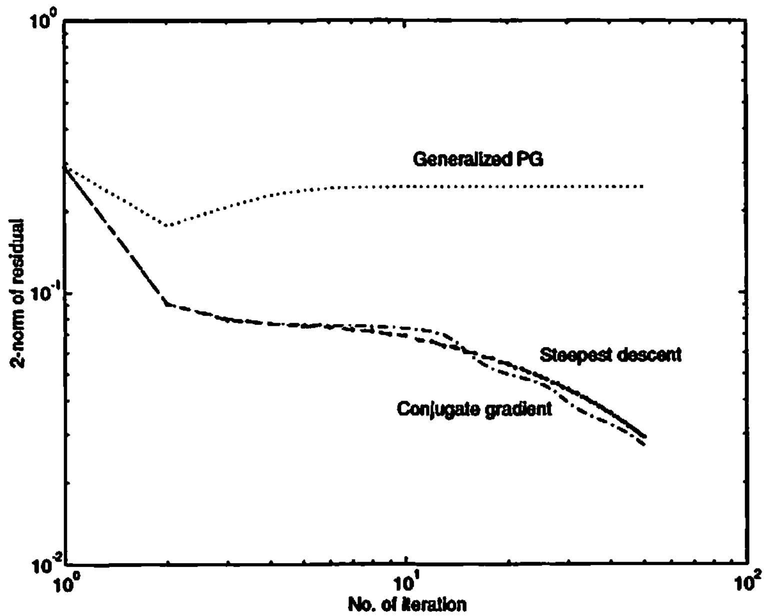


(b)

FIG. 3. *Convergence history of three iterative algorithms for Test Problem 1 with (a) $M = 15$ and (b) $M = 50$, where the number of observed data points is $2M + 1$.*



(a)



(b)

FIG. 4. *Convergence history of three iterative algorithms for Test Problem 2 with (a)* $M = 25$ *and (b)* $M = 55$*, where the number of observed data points is* $2M + 1$.

observation is consistent with the theoretical derivation given in §4.

6. Conclusions and extensions. This research examined signal extrapolation schemes based on the wavelet model of scale-time limited signals. We showed that the generalized PG algorithm converges for semi-orthogonal wavelets in both continue-time and discrete-time cases and the solution can be viewed as a minimum norm solution. Practically, the discrete-time implementation is needed, and two new effective algorithms have been proposed and studied. There are several interesting topics worth further study. For example, it is important to compare the performance of different wavelet bases, and study the optimal basis for some particular applications. Besides, we assume that the J and K values of the scale-time limited sequence are known a priori. However, they are usually not available and have to be estimated in practice.

REFERENCES

[1] C. L. BYRNE AND R. M. FITZGERALD, *Spectral estimators that expand the maximum entropy and maxmum likehood method*, SIAM J. Appl. Math, vol. 44, no. 4 (1984), pp. 425–442.

[2] S. D. CABRERA AND T. W. PARK, *Extrapolation and spectral estimation with iterative weighted norm modification*, IEEE Trans. on Signal Processing, vol. 39, no. 4 (1991), pp. 842–851.

[3] T. CHANG AND C.-C. J. KUO, *Texture analysis and classification with tree-structured wavelet transform*, IEEE Trans. Image Processing, 2 (1993).

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

[5] ——, ed., *Wavelets: A Tutorial in Theory and Applications*, Academic Press, New York, 1992.

[6] R. R. COIFMAN AND M. V. WICKERHAUSER, *Entropy-based algorithms for best basis selection*, IEEE Trans. on Information Theory, 38 (1992), pp. 713–718.

[7] I. DAUBECHIES, *Ten Lectures on Wavelets*, SIAM, Philadephia, 1992.

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

[9] A. K. JAIN AND S. RANGANATH, *Extrapolation algorithm for discrete signal with application in spectral estimation*, IEEE Trans. on Acoustic, Speech, and Signal Processing, 29 (1981), pp. 830–845.

[10] L.-C. LIN AND C.-C. J. KUO, *Signal extrapolation in noisy data with wavelet representation*. SPIE's 1993 Annual Meeting, San Diego, California, July 11-16, 1993.

[11] D. G. LUENBERGER, *Linear and nonlinear programming*, Addison-Wesley Publishing Company Inc., 1984.

[12] S. MALLAT, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Trans. on Pattern Anal. and Mach.Intell., 11 (1989), pp. 674–693.

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

[14] L. C. POTTER AND K. S. ARUN, *Energy concentration in band-limited extrapolation*, IEEE Trans. on Acoustic, Speech, and Signal Processing, vol. 37 (1989), pp. 1027–1041.

[15] J. L. C. SANZ AND T. S. HUANG, *Discrete and continuous bandlimited signal extrapolation*, IEEE Trans. on Acoustics, Speech and Signal Processing, 31 (1983), pp. 1276–1285.

[16] D. SLEPIAN, *Some comments on fourier analysis, uncertainty and modeling*, SIAM Review, vol. 25, no. 3 (1983), pp. 379–393.

[17] 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.

[18] J. VON NEUMANN, *the geometry of orthogonal spaces, vol II*, Princeton University Press, Princeton, NJ, 1950.

[19] X.-G. XIA, C.-C. J. KUO, AND Z. ZHANG, *Signal extrapolation in wavelet subspaces*, Dept. of EE-Systems, University of Southern California, 1992. USC-SIPI Report No. 219.