ELSEVIER

Signal Processing 48 (1996) 51–65

SIGNAL PROCESSING

On the convergence of wavelet-based iterative signal extrapolation algorithms[1]

Li-Chien Lin[a], Xiang-Gen Xia[b], C.-C. Jay Kuo[c,]*

[a] *Department of Electrical Engineering, Feng Chia University, Taichung, Taiwan*
[b] *Hughes Research Laboratories, Bldg 254, M.S. RL96, 3011 Malibu Canyon Rd., Malibu, CA 90265, USA*
[c] *Signal and Image Processing Institute and the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089-2564, USA*

Received 9 September 1994; revised 11 May 1995 and 4 September 1995

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.

Zusammenfassung

Kürzlich wurde von Xia, Kuo und Zhang ein verallgemeinerter Papoulis–Gerchberg (PG) Algorithmus zur Signalextrapolation basierend auf der Wavelet-Darstellung vorgeschlagen. In dieser Arbeit untersuchen wir die Konvergenzeigenschaft und die Konvergenzgeschwindigkeit von mehreren Signalextrapolationsalgorithmen in Wavelet-Unterräumen. Wir zeigen zunächst daß der verallgemeinerte PG-Algorithmus gegen die Lösung minimaler Norm konvergiert, wenn die Wavelet-Basen semi-orthogonal sind (auch bekannt als prewavelet). Die Konvergenz von numerischen Implementierungen des verallgemeinerten PG-Algorithmus ist allerdings langsam. Um die Konvergenzgeschwindigkeit zu verbessern, formulieren wir das diskrete Signalextrapolationsproblem als einen zweistufigen Prozeß und verwenden die Steepest-Descent-Methode und die Methode konjugierter Gradienten zu dessen Lösung. Numerische Experimente illustrieren die Leistungsfähigkeit der vorgeschlagenen Algorithmen.

Résumé

Un algorithme de Papoulis–Gerchberg (PG) généralisé pour l'extrapolation de signaux, basé sur une représentation en ondelettes, a été récemment proposé par Xia, Kuo and Zhang. Dans cette étude, nous examinons les propriétés de

* Corresponding author. E-mail: cckuo@sipi.usc.edu.
[1] This research was supported by the National Science Foundation Presidental Faculty Fellow (PFF) Award ASC-9350309.

0165-1684/96/$15.00
SSDI 0165-1684(95)00123-9

convergence et le taux de convergence de plusieurs algorithmes d'extrapolation de signaux dans les sous-espaces d'ondelettes. Nous montrons tout d'abord que l'algorithme PG généralisé converge vers la solution de norme minimale quand les bases d'ondelettes sont semi orthogonales (c.à.d. connues comme pré-ondelettes). Toutefois, l'algorithme PG généralisé converge lentement pour une implementation numérique. Afin d'accélérer le taux de convergence, nous formulons le problème d'extrapolation de signaux discrets comme processus à deux étapes et appliquons les méthodes de descente rapide et de gradient conjugué à sa résolution. Des expérimentations numériques sont fournies pour illustrer les performances des algorithmes proposés.

Keywords: Signal extrapolation; Papoulis–Gerchberg algorithm; Wavelets

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 1970s, 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-squares (MNLS) solution [9], the discrete prolate spheroidal sequence (DPSS) expansion [16, 17], and the weighted norm least-squares solution [1, 2, 4]. 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 et al. [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 was 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 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 Section 2. In Section 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 Section 4. Numerical experiments are given in Section 5 to show the convergence behavior of the proposed algorithms. Concluding remarks and possible extensions are given in Section 6.

2. Results from wavelet theory

We review some basic results of biorthogonal wavelet theory below, and refer interested readers to [4, 5, 7] for a more detailed discussion. Let $\phi(t)$ be a biorthogonal scaling function and $\psi(t)$ and $\mathscr{P}_j$ be its associated wavelet function and multiresolution analysis (MRA), $\tilde{\phi}(t)$, $\tilde{\psi}(t)$ and $\tilde{\mathscr{P}}_j$ be their duals, respectively, as follows:

$$\cdots \subset \mathscr{P}_0 \subset \mathscr{P}_1 \subset \mathscr{P}_2 \subset \cdots \subset L^2(\mathbf{R}),$$

where the union of all subspaces $\mathscr{P}_j$ of $L^2(\mathbf{R})$ is dense in $L^2(\mathbf{R})$ and the intersection of all subspaces $\mathscr{P}_j$ only contains the 0 function, i.e.,

$$\overline{\bigcup_i \mathscr{P}_j} = L^2(\mathbf{R}), \qquad \bigcap_j \mathscr{P}_j = \{0\}.$$

Moreover,

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

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 $\mathscr{P}_j$ with its dual $\tilde{\phi}_{jk} \triangleq 2^{j/2}\tilde{\phi}(2^j t - k)$ 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, and $\langle f, g \rangle = \int f(t) g^*(t)\,\mathrm{d}t$.

If we let $\psi_{jk}(t) \triangleq 2^{j/2}\psi(2^j t - k)$ and $\tilde{\psi}_{jk}(t) \triangleq 2^{j/2}\tilde{\psi}(2^j t - k)$, then ψ_{jk}, $j, k \in \mathbf{Z}$, the set of all integers, form a biorthogonal basis for the space $L^2(\mathbf{R})$ as

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

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

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

If a biorthogonal wavelet basis is not orthogonal or semi-orthogonal, then it is called *non-orthogonal.* Clearly, the set of prewavelet includes the set of orthogonal wavelets as a subset, and the relationship $\mathscr{P}_j = \tilde{\mathscr{P}}_j$ holds for both cases. For more details, see [4, 5].

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

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

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

$$f(t) = \sum_{k=-\infty}^{\infty} c_{J,k} \phi_{Jk}(t) = \sum_{j<J} \sum_{k=-\infty}^{\infty} b_{j,k} \psi_{jk}(t), \tag{2}$$

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 (1) are called the wavelet series transform (WST) coefficients of $f(t)$, and (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:

$$\begin{aligned} c_{j-1,k} &= \sqrt{2} \textstyle\sum_n \tilde{h}_{n-2k} c_{j,n}, \\ b_{j-1,k} &= \sqrt{2} \textstyle\sum_n \tilde{g}_{n-2k} c_{j,n}, \end{aligned} \tag{3}$$

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

$$c_{j+1,n} = \sqrt{2}\left(\sum_k h_{n-2k} c_{j,k} + \sum_k g_{n-2k} b_{j,k}\right)$$
$$\text{for } j = J_0, J_0+1, \ldots, J-1. \tag{4}$$

In practice, $c_{J,n}$ can be viewed as a sequence of $x[n]$, sampled of a signal $f(n/2^J)$. Then (3) and (4) are called the discrete wavelet transform (DWT) and the inverse discrete wavelet transform (IDWT), respectively.

Formulas (3) and (4) can be implemented as a biorthogonal quadrature mirror filter bank [7] and the sequences g, h, $\tilde{g}$ and $\tilde{h}$ are the impulse responses of the corresponding filters. By choosing these sequences carefully, we are lead to different biorthogonal wavelet bases [7]. The biorthogonal wavelet basis is reduced to the orthogonal one if $g = \tilde{g}$ and $h = \tilde{h}$.

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 (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 $b_{j,k} = \langle f, \psi_{j,k}\rangle$ 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.

$$[-T_0, T_0] \times [(-2^J\xi_0, -2^{J_0}\xi_0) \cup (2^{J_0}\xi_0, 2^J\xi_0)], \tag{5}$$

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

$$\int_{2^{J_0}\xi_0 \leqslant |\xi| \leqslant 2^J\xi_0} |\hat{f}|^2 \,\mathrm{d}\xi \geqslant (1-\varepsilon)\|f\|^2$$

and

$$\int_{|x| \leqslant T_0} |f|^2 \,\mathrm{d}x \geqslant (1-\varepsilon)\|f\|^2.$$

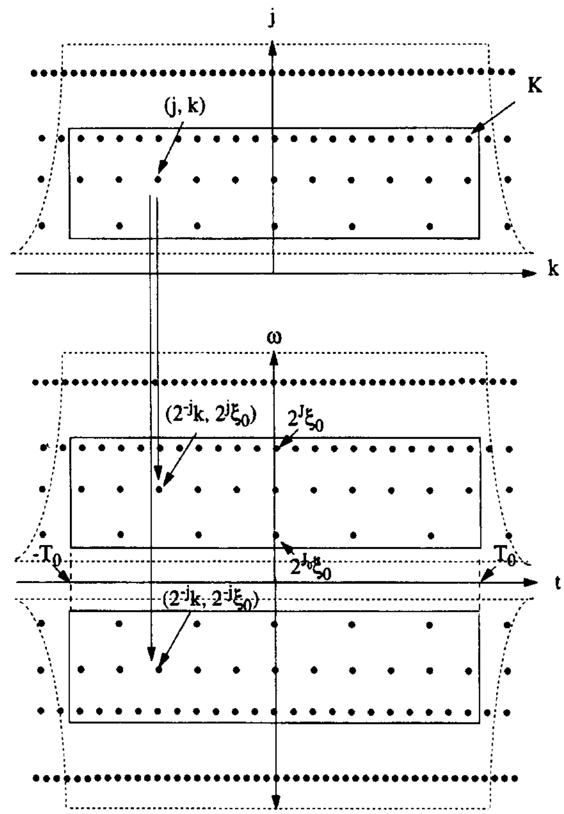


Fig. 1. Lattice structure of the wavelet coefficients.

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

$$D = \{(j,k) \in \mathbf{Z}^2 \colon J_0 \leqslant j < J \text{ and } 2^{-j}|k| \leqslant T_0 + t_\varepsilon\}$$

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 $\mathscr{K} = \{k \colon |k| \leqslant \boldsymbol{K} = 2^J(T_0 + t_\varepsilon)\}$. We have from the above analysis that

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

Thus, the following space:

$$V_{J_0,J,\mathscr{K}} = \left\{ f(t)\colon f(t) = \sum_{J_0 \leqslant j \leqslant J} \sum_{k \in \mathscr{K}} b_{j,k}\psi_{jk}(t) \right\}, \quad (6)$$

provides a good model for signals concentrated in (5).

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

$$x[k] = (D_{J_0,J}^{-1} T_{J,\mathscr{K}} D_{J_0,J} x)[k], \quad k \in \mathbf{Z},$$

where $D_{J_0,J}^{-1}$ and $D_{J_0,J}$ are, respectively, the DWT and IDWT operators and $T_{J,\mathscr{K}}$ is the following projection operator:

$$T_{J,\mathscr{K}} u_{j,k} = \begin{cases} u_{j,k} & \text{if } J_0 \leqslant j < J, k \in \mathscr{K}, \\ 0 & \text{otherwise}. \end{cases}$$

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

$$x[k] \approx (D_{J_0,J}^{-1} T_{J,\mathscr{K}} D_{J_0,J} x)[k].$$

We conclude from the above discussion that the signal set

$$S_{J_0,J,\mathscr{K}} = \{x[k]\colon$$
$$x[k] = (D_{J_0,J}^{-1} T_{J,\mathscr{K}} D_{J_0,J} x)[k], \quad k \in \mathbf{Z}\} \quad (7)$$

provides a good approximation model for discrete-time signals with energy concentrated in (5).

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

$$x[k] = D_{J_0,J}^{-1} T_{J,\mathscr{K}} D_{J_0,J} x[k]$$

or

$$\boldsymbol{x} = D_{J_0,J}^{-1} T_{J,\mathscr{K}} D_{J_0,J} \boldsymbol{x}, \quad (8)$$

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

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

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

$$T_{J,\mathscr{K}} = U^{\mathrm{T}} U,$$

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

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

3.2. Generalized PG algorithm

The generalized PG algorithms for continuous- and discrete-time proposed by Xia et al. in [19] are summarized below.

We first examine the continuous-time case. Let P_J, P_T, Q_J, Q_T denote the projection operators which project functions onto the subspaces $\mathscr{P}_J$, $\mathscr{P}_T$, $\mathscr{P}_J^{\perp}$ and $\mathscr{P}_T^{\perp}$, respectively, where $\mathscr{P}_J$ is the wavelet subspace as defined in Section 2 and $\mathscr{P}_T$ 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 (1)–(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

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

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

$$\begin{aligned} &f^{(0)}(t) = g(t), \\ &f^{(l+1)}(t) = Q_T P_J f^{(l)}(t) + f^{(0)}(t), \quad l = 0, 1, 2, \ldots \end{aligned} \quad (9)$$

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

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

$$Q_J(s,t) \triangleq \sum_k \phi_{Jk}(s)\phi_{Jk}(t) \quad (10)$$

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 $\phi(t)$, $t\in[-2^JT-k,2^JT-k]$, $k\in\mathbf{Z}$, 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

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

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

$$x^{(0)}[n]=\boldsymbol{T}_{\mathcal{N}}x[n],$$

$$x^{(l+1)}[n]=\boldsymbol{T}_{\mathcal{N}}x[n]+(\boldsymbol{I}-\boldsymbol{T}_{\mathcal{N}})\boldsymbol{D}_{J_0,J}^{-1}\boldsymbol{T}_{J,\mathscr{K}}\boldsymbol{D}_{J_0,J}x^{(l)}[n],$$

$$l=0,1,2,\dots \qquad (11)$$

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 (9) with $f\in\mathscr{P}_J$ for a certain $J>0$. Then, when $l\to\infty$, $f^{(l)}(t)$ converges to the minimum norm solution $f^\dagger$ satisfying*

$$f^\dagger\in\mathscr{P}_J$$

and

$$\|f^\dagger\|=\min\{\|h\|: h\in\mathscr{P}_J, \mathbf{P}_Th=\mathbf{P}_Tf\}.$$

Proof. Any $f(t)\in\mathscr{P}_J$ and $g(t)=\boldsymbol{P}_Tf(t)$ can be written as

$$f(t)=g+h_1, \qquad (12)$$

where

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

Since $h_1=\boldsymbol{Q}_Tf$ and $f=\boldsymbol{P}_Jf$, we can rewrite (12) as

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

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

$$\begin{aligned}f&=g+\boldsymbol{Q}_T\boldsymbol{P}_Jg+(\boldsymbol{Q}_T\boldsymbol{P}_J)^2g\cdots\\&\quad+(\boldsymbol{Q}_T\boldsymbol{P}_J)^lg+(\boldsymbol{Q}_T\boldsymbol{P}_J)^lh_1\\&=f^{(l)}+(\boldsymbol{Q}_T\boldsymbol{P}_J)^l\boldsymbol{Q}_Tf, \quad l\to\infty,\end{aligned}$$

where the last equality is due to (9). By the definition of semi-orthogonal wavelets, the operator $\boldsymbol{P}_J$ is an orthogonal projection. The operator $\boldsymbol{Q}_T$ is also clearly an orthogonal projection. By using the Alternating Projection Theorem ([18, Theorem 13.7], we have

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

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

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

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

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

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

With this theorem, we have the following straightforward corollary.

Corollary 1. *If* $\mathscr{P}_J \cap \mathscr{P}_T^{\perp} = \{0\}$, $f^{(l)}(t)$ *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 $\mathscr{P}_J$ 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, $\boldsymbol{D}_{J_0,J}^{-1}$ in (11) is the transpose of $\boldsymbol{D}_{J_0,J}$. Thus, the operator $\boldsymbol{D}_{J_0,J}^{-1}\boldsymbol{T}_{J,\mathscr{K}}\boldsymbol{D}_{J_0,J}$ is an orthogonal projection. By applying the Alternating Projection Theorem to (11), we can prove the convergence of the iterative procedure (11). However, the operator $\boldsymbol{D}_{J_0,J}^{-1}\boldsymbol{T}_{J,\mathscr{K}}\boldsymbol{D}_{J_0,J}$ 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* $\mathscr{K}$ *the operator* $\boldsymbol{D}_{J_0,J}^{-1}\boldsymbol{T}_{J,\mathscr{K}}\boldsymbol{D}_{J_0,J}$ *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 $\boldsymbol{D}_{J_0,J}^{-1}\boldsymbol{T}_{J,\mathscr{K}}\boldsymbol{D}_{J_0,J}$ is an orthogonal projection for any fixed J_0, J and $\mathscr{K}$, the wavelet basis is orthogonal. This is equivalent to prove that $\boldsymbol{D}_{J_0,J}^{-1}$ is equal to the transpose $\boldsymbol{D}_{J_0,J}^{T}$ of $\boldsymbol{D}_{J_0,J}$ for any fixed J_0 and J.

Recall (6) that we use $\boldsymbol{V}_{J_0,J,\mathscr{K}}$ to denote the set of all $(J_0, J; \mathscr{K})$ scale-time-limited sequences. Then, the operator $\boldsymbol{D}_{J_0,J}^{-1}\boldsymbol{T}_{J,\mathscr{K}}\boldsymbol{D}_{J_0,J}$ is an orthogonal projection onto $\boldsymbol{V}_{J_0,J,\mathscr{K}}$. Thus, any $x \in l^2$ can be decomposed as

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

for certain $y \in l^2$. By using the orthogonality of the operator $\boldsymbol{D}_{J_0,J}^{-1}\boldsymbol{T}_{J,\mathscr{K}}\boldsymbol{D}_{J_0,J}$, 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 $\boldsymbol{D}_{J_0,J}^{-1}$. Let us partition all integers into two non-overlapping groups $\mathscr{N}_1$ and $\mathscr{N}_2$, i.e. $Z = \mathscr{N}_1 \cup \mathscr{N}_2$ and $\mathscr{N}_1 \cap \mathscr{N}_2 = \emptyset$. Then, this leads to

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

or

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

This equality implies

$$\sum_i b_{ij_1} b_{ij_2} = 0 \quad \text{for } j_1 \neq j_2,\ j_1, j_2 \in \mathbf{Z}, \tag{13}$$

since y can be any element in l^2 and $\mathscr{N}_1$ can be any subset of $\mathbf{Z}$. As a direct consequence of (13), $\mathscr{D}_{J_0,J}$ is orthogonal. □

For a non-orthogonal wavelet basis, the projection $\boldsymbol{P}_J$ in algorithm (9) or (11) is a non-orthogonal projection. The generalized PG algorithm in both continuous- and discrete-time cases assume the form

$$f^{(l)} = f - (\boldsymbol{TP})^l f, \quad l = 0, 1, 2, \ldots,$$

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

Proposition 2. *If the operator* $\boldsymbol{P}$ *is a non-orthogonal projection, there is a truncation operator* $\boldsymbol{T}$ *such that* $\|\boldsymbol{TP}\| > 1$.

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

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

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 non-orthogonal wavelet bases (not orthogonal or semi-orthogonal). 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 non-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 $y = T_{\mathcal{N}}x$ be a given segment of x. According to the discussion in Section 3.1, we have

$$y = T_{\mathcal{N}} D_{J_0,J}^{-1} T_{J,\mathcal{K}} D_{J_0,J} x$$
$$= T_{\mathcal{N}} D_{J_0,J}^{-1} U^{\mathrm{T}} U D_{J_0,J} x,$$

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

$$y = T_{\mathcal{N}} W p .$$

By multiplying both sides with $(T_{\mathcal{N}}W)^{\mathrm{T}}$, we obtain the normal equation

$$W^{\mathrm{T}} y = W^{\mathrm{T}} T_{\mathcal{N}} W p , \tag{14}$$

where the equalities $T_{\mathcal{N}}y = y$ and $(T_{\mathcal{N}})^{\mathrm{T}}T_{\mathcal{N}} = T_{\mathcal{N}}$ are used to simplify the result. Furthermore, note that since x is a scale–time-limited sequence, i.e.

$$x = D_{J_0,J}^{-1} T_{J,\mathcal{K}} D_{J_0,J} x = D_{J_0,J}^{-1} U^{\mathrm{T}} U D_{J_0,J} x ,$$

we can obtain x from p via

$$x = Wp . \tag{15}$$

Therefore, we can divide the solution procedure of determining x into two steps: first solving the normal equation for p and then determining x from p as described by (14) and (15), respectively.

In the following discussion, we assume that the $L \times L$ matrix $W^{\mathrm{T}}T_{\mathcal{N}}W$ is of full rank and that the wavelet basis under consideration is orthogonal. Some useful properties of the operators in (14)–(15) are summarized below.

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

$$x = D_{J_0,J}^{-1} U^{\mathrm{T}} U D_{J_0,J} x = W W^{\mathrm{T}} x . \tag{16}$$

Property 2. From the definition of W, we have

$$W^{\mathrm{T}} W = U D_{J_0,J} D_{J_0,J}^{-1} U^{\mathrm{T}} = U U^{\mathrm{T}} = I_L , \tag{17}$$

where I_L is an identity matrix of dimension $L \times L$.

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

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

$$\lambda_{\max}(W^{\mathrm{T}} T_{\mathcal{N}} W) \leqslant 1 . \tag{18}$$

This can be proved by noting that

$$\lambda_{\max}(W^{\mathrm{T}} T_{\mathcal{N}} W) = \max_{z \neq 0} \frac{(z^*)^{\mathrm{T}} W^{\mathrm{T}} T_{\mathcal{N}} W_z}{(z^*)^{\mathrm{T}} z}$$

and

$$(z^*)^{\mathrm{T}} W^{\mathrm{T}} T_{\mathcal{N}} W z = (z^*)^{\mathrm{T}} U D_{J_0,J} T_{\mathcal{N}} D_{J_0,J}^{-1} U^{\mathrm{T}} z$$
$$\leqslant (z^*)^{\mathrm{T}} U D_{J_0,J} D_{J_0,J}^{-1} U^{\mathrm{T}} z = (z^*)^{\mathrm{T}} U U^{\mathrm{T}} z = (z^*)^{\mathrm{T}} z .$$

There are two reasons to avoid solving (14) and (15) with direct methods. First, direct computation of the matrix W is expensive. Second, if the matrix is ill-conditioned, the direct method is usually unstable. Therefore, we consider the solution of (14) and (15) with iterative algorithms.

4.2. Steepest descent method

The iterative process based on the steepest descent method to solve (14)–(15) can be stated as: with any given $\boldsymbol{x}_0$, we perform the following iteration:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{r}_k \quad \text{for } k = 0, 1, 2, \dots, \tag{19}$$

where

$$\boldsymbol{r}_k = \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{T}_1 \boldsymbol{x}_k - \boldsymbol{y}) \tag{20}$$

and

$$\alpha_k = \frac{\boldsymbol{r}_k^{\mathrm{T}} \boldsymbol{r}_k}{\boldsymbol{r}_k^{\mathrm{T}} \boldsymbol{T}_1 \boldsymbol{r}_k}. \tag{21}$$

We show below how the algorithm given by (19)–(21) is directly related to the well-known steepest descent method. The application of the steepest descent method [11, pp. 215] to the solution of the normal equation

$$\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y} = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{W}\boldsymbol{p}$$

is equivalent to the minimization of the cost functional

$$f(\boldsymbol{p}) = \tfrac{1}{2}\boldsymbol{p}^{\mathrm{T}}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{W})\boldsymbol{p} - \boldsymbol{p}^{\mathrm{T}}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}).$$

The result can be written as

$$\boldsymbol{p}_{k+1} = \boldsymbol{p}_k - \alpha_k \boldsymbol{d}_k, \tag{22}$$

where the vector

$$\boldsymbol{d}_k = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{W}\boldsymbol{p}_k - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}$$

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

$$\alpha_k = \frac{\boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{d}_k}{\boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{W}\boldsymbol{d}_k} \tag{23}$$

is determined by $\min_\alpha f(\boldsymbol{p}_k + \alpha \boldsymbol{d}_k)$. Premultiplying both side of (22) with $\boldsymbol{W}$ and applying (15), we have

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \alpha_k \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{T}_1 \boldsymbol{x}_k + \boldsymbol{y}). \tag{24}$$

Thus, we can justify (20). Furthermore, due to $\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} = \boldsymbol{I}_L$ (Property 2), we have

$$\begin{aligned}
&\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{d}_k \\
&= (\boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{x}_k - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W})(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{x}_k - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}) \\
&= (\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{T}_1 \boldsymbol{x}_k - \boldsymbol{y}))^{\mathrm{T}}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}(\boldsymbol{T}_1 \boldsymbol{x}_k - \boldsymbol{y})) \\
&= \boldsymbol{r}_k^{\mathrm{T}}\boldsymbol{r}_k
\end{aligned}$$

and

$$\begin{aligned}
&\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{W}\boldsymbol{d}_k \\
&= (\boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{x}_k - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{W})(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{x}_k - \boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}) \\
&= (\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{T}_1 \boldsymbol{x}_k - \boldsymbol{y}))^{\mathrm{T}}\boldsymbol{T}_1(\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{T}_1 \boldsymbol{x}_k - \boldsymbol{y})) \\
&= \boldsymbol{r}_k^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{r}_k,
\end{aligned}$$

so that (21) can also be justified.

The discrete generalized PG algorithm described in Section 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 (20), then we have the iterative process

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_1 \boldsymbol{x}_k + \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{y}.$$

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

$$\begin{cases} \boldsymbol{x}'_{k+1} = \boldsymbol{x}_k - \boldsymbol{T}_1 \boldsymbol{x}_k + \boldsymbol{y}, \\ \boldsymbol{x}_k = \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}'_k. \end{cases}$$

As a consequence, we have

$$\boldsymbol{x}'_{k+1} = (\boldsymbol{I} - \boldsymbol{T}_1)\boldsymbol{W}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{x}'_k + \boldsymbol{y}.$$

This is exactly the discrete generalized PG algorithm. Although both the discrete generalized PG algorithm (10) and the steepest descent method (19) 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 Section 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 $\boldsymbol{Q}\boldsymbol{z} = \boldsymbol{b}$, where Q is symmetric positive definite below [11, p. 244]. Given any

z_0 and $\boldsymbol{d}_0 = \boldsymbol{b} - \boldsymbol{Q}z_0$, we perform the following iteration for $k = 0, 1, 2, \ldots$:

$$g_k = \boldsymbol{Q}z_k - \boldsymbol{b},$$

$$z_{k+1} = z_k + \alpha_k \boldsymbol{d}_k,$$

$$\alpha_k = -\frac{g_k^{\mathrm{T}} g_k}{\boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{d}_k},$$

$$\boldsymbol{d}_{k+1} = -g_{k+1} + \beta_k \boldsymbol{d}_k,$$

$$\beta_k = \frac{g_{k+1}^{\mathrm{T}} g_{k+1}}{g_k^{\mathrm{T}} g_k}.$$

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

$$\boldsymbol{x}_0 = \boldsymbol{0}, \qquad \tilde{\boldsymbol{d}}_0 = \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}} \boldsymbol{y}, \qquad \tilde{g}_0 = -\tilde{\boldsymbol{d}}_0,$$

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

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \tilde{\boldsymbol{d}}_k,$$

$$\alpha_k = \frac{\tilde{g}_k^{\mathrm{T}} \tilde{g}_k}{\tilde{\boldsymbol{d}}_k^{\mathrm{T}} \boldsymbol{T}_{\mathcal{N}} \tilde{\boldsymbol{d}}_k},$$

$$\tilde{g}_{k+1} = \boldsymbol{W}\boldsymbol{W}^{\mathrm{T}} (\boldsymbol{T}_{\mathcal{N}} \boldsymbol{x}_{k+1} - \boldsymbol{y}),$$

$$\tilde{\boldsymbol{d}}_{k+1} = \beta_k \tilde{\boldsymbol{d}}_k - \tilde{g}_{k+1},$$

$$\beta_k = \frac{\tilde{g}_{k+1}^{\mathrm{T}} \tilde{g}_{k+1}}{\tilde{g}_k^{\mathrm{T}} \tilde{g}_k}.$$

4.4. Convergent rate analysis

To simplify the discussion, we use the notation

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

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

$$\min_{\boldsymbol{p}} \tfrac{1}{2} \boldsymbol{p}^{\mathrm{T}} \boldsymbol{Q} \boldsymbol{x} - \boldsymbol{p}^{\mathrm{T}} \boldsymbol{b},$$

which is again equivalent to

$$\min_{\boldsymbol{p}} E(\boldsymbol{p}) = \min_{\boldsymbol{p}} \tfrac{1}{2} (\boldsymbol{p} - \hat{\boldsymbol{p}})^{\mathrm{T}} \boldsymbol{Q} (\boldsymbol{p} - \hat{\boldsymbol{p}}), \tag{25}$$

where $\hat{\boldsymbol{p}}$ be the solution vector $\boldsymbol{Q}\hat{\boldsymbol{p}} = \boldsymbol{b}$. It is easier to analyze the convergent rate for (25). The gradient of E is $\boldsymbol{d} = \boldsymbol{Q}\boldsymbol{p} - \boldsymbol{b}$. Since the discrete generalized PG which can be regarded as the special case of the steepest descent with $\alpha_k = 1$, we have from (22)

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

By direct computation, we have

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

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

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

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

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

Since $\boldsymbol{Q}$ is symmetric, it is unitarily diagonalizable with ordered diagonals denoted by $\lambda_{\min} = \lambda_1 \leqslant \lambda_2 \cdots \leqslant \lambda_L = \lambda_{\max}$, which are also eigenvalues of $\boldsymbol{Q}$. 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

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

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

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

where λ_{max} and λ_{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

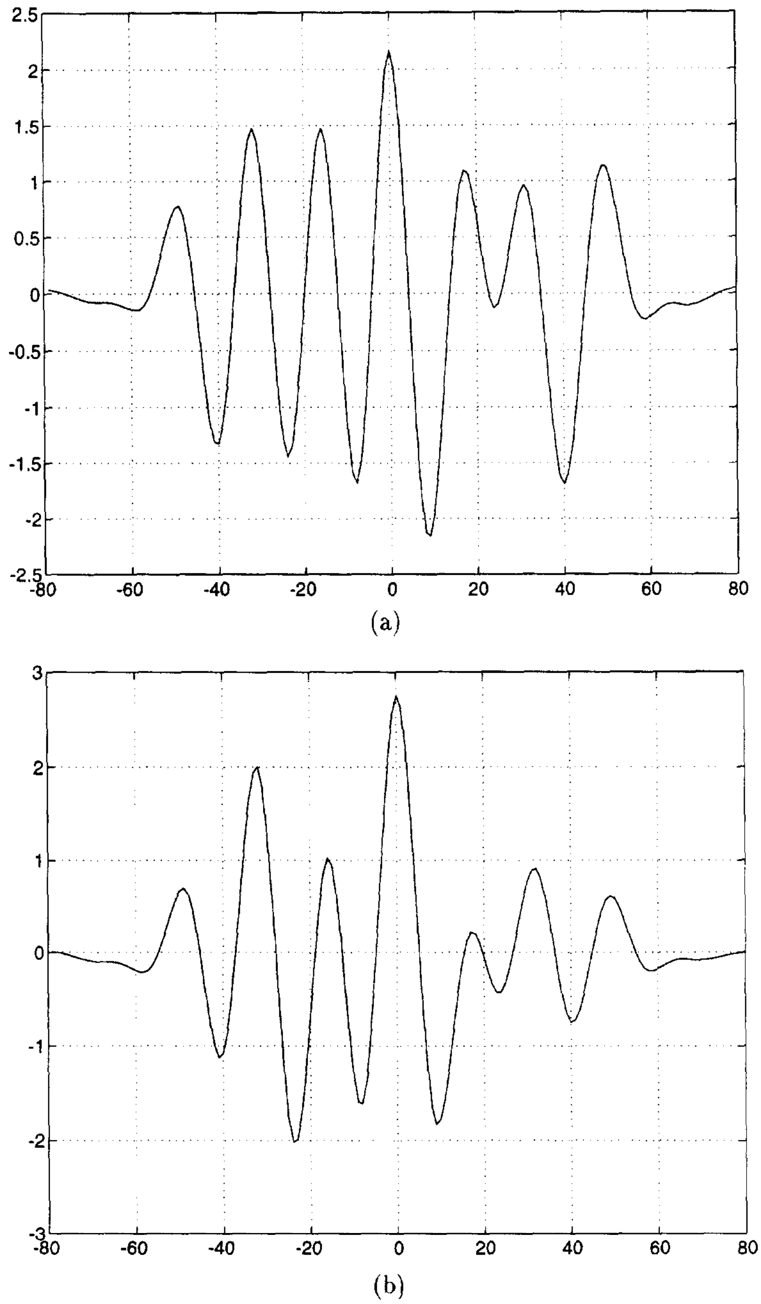
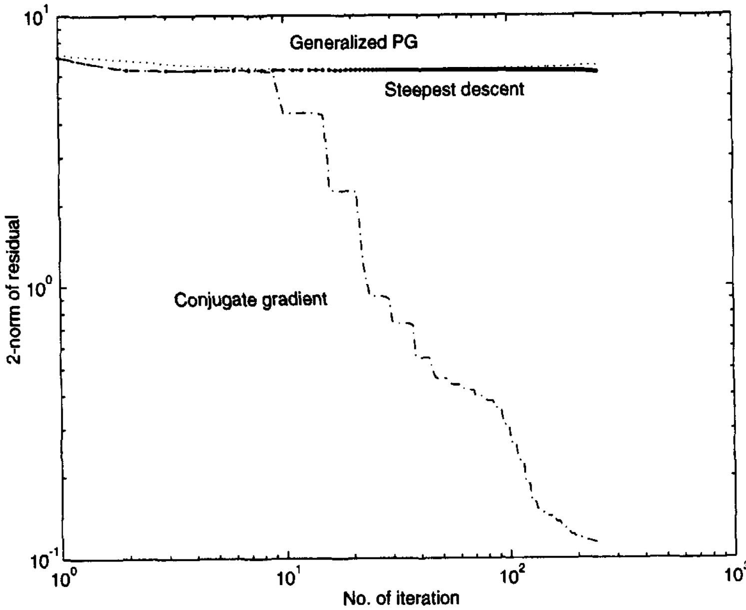
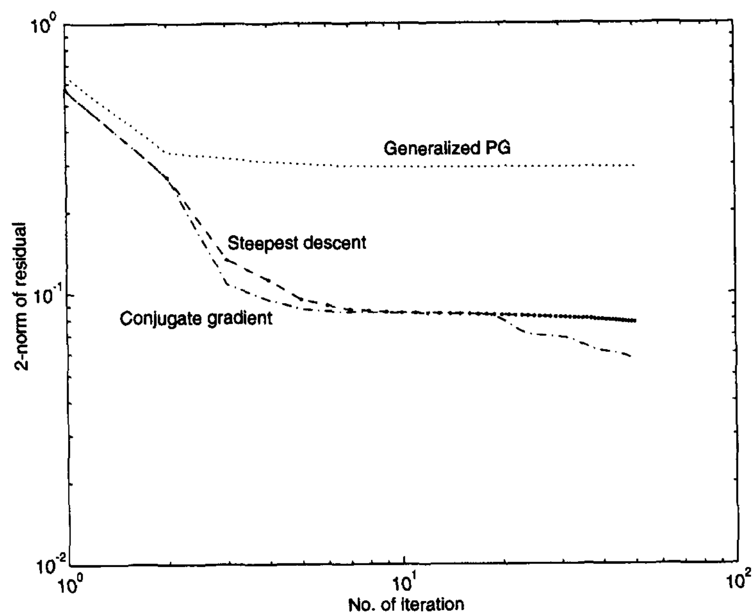


Fig. 2. Original signals in (a) Test problem 1 and (b) Test problem 2.



(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 $2M + 1$.

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 maximum eigenvalues of the matrix $Q = \boldsymbol{W}^{\mathrm{T}}\boldsymbol{T}_{\mathrm{t}}\boldsymbol{W}$, 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 $-3 \leqslant k \leqslant 4$ (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 $-3 \leqslant k \leqslant 4$ are non-zeros. Consequently, the degree of the freedom of the problem is $L = 8$.

We observe $(2M + 1)$ data points $x[n]$ with $|n| \leqslant M$, 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.

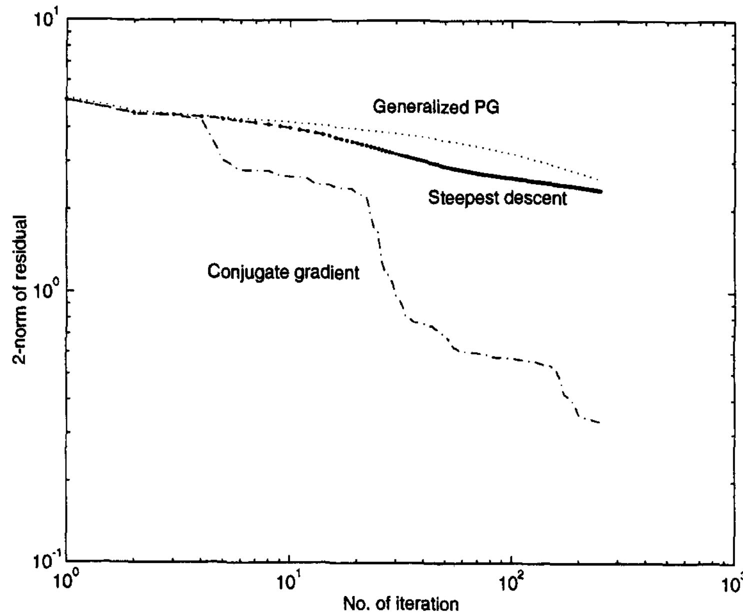
Test problem 2. In this problem, we use the same wavelet basis as in Test problem 1, but increase the number of non-zero wavelet coefficients so that the degree of freedom of this problem is $L = 12$. The test signal $x[n]$ is generated by randomly choosing the wavelet coefficients $b_{j,k}$ with $J_0 = 0$, $J = 1$ and $-3 \leqslant k \leqslant 4$ (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

Table 1
The maximum and minimum eigenvalues and the convergence rate bounds for Test problem 1

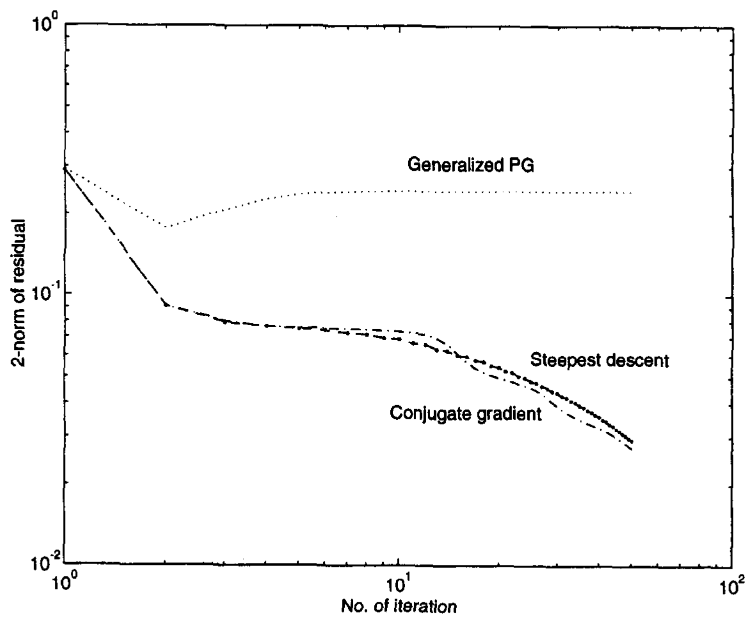
	$\lambda_{\min}$	$\lambda_{\max}$	r_{gpg}	r_{sd}
$M = 15$	0.0000081	0.6263491	0.9999838	0.9999482
$M = 50$	0.2212905	0.9985423	0.6063885	0.4059970

Table 2
The maximum and minimum eigenvalues and the convergence rate bounds for Test problem 2

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



(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 $2M + 1$.

same convergence rate of 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 observation is consistent with the theoretical derivation given in Section 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 orthogonal and semi-orthogonal wavelets in the continue-time case as well as for orthogonal wavelets in the discrete-time case, and its 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 maximum likelihood method", *SIAM J. Appl. Math.*, Vol. 44, No. 4, April 1984, pp. 425–442.

[2] S.D. Cabrera and T.W. Park, "Extrapolation and spectral estimation with iterative weighted norm modification", *IEEE Trans. Signal Processing*, Vol. 39, No. 4, April 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*, Vol. 2, October 1993.

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

[5] C.K. Chui, 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. Inform. Theory*, Vol. 38, March 1992, pp. 713–718.

[7] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[8] R.W. Gerchbert, "Super-resolution through error energy reduction", *Optica Acta*, Vol. 21, 1974, pp. 709–720.

[9] A.K. Jain and S. Ranganath, "Extrapolation algorithm for discrete signal with application in spectral estimation", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 29, August 1981, pp. 830–845.

[10] L.-C. Lin and C.-C.J. Kuo, "Signal extrapolation in noisy data with wavelet representation", in *SPIE's 1993 Annual Meeting*, San Diego, California, 11–16 July, 1993.

[11] D.G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, Reading, MA, 1984.

[12] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 11, 1989, pp. 674–693.

[13] A. Papoulis, "A new algorithm in spectral analysis and band limited extrapolation", *IEEE Trans. Circuits Systems*, Vol. 22, 1975, pp. 735–742.

[14] L.C. Potter and K.S. Arun, "Energy concentration in band-limited extrapolation", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 37, July 1989, pp. 1027–1041.

[15] J.L.C. Sanz and T.S. Huang, "Discrete and continuous bandlimited signal extrapolation", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. 31, 1993, pp. 1276–1285.

[16] D. Slepian, "Some comments on Fourier analysis, uncertainty and modeling", *SIAM Rev.*, Vol. 25, No. 3, July 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. Acoust. Speech Signal Process.*, Vol. 32, December 1984, pp. 1201–1212.

[18] J. von Neumann, *The Geometry of Orthogonal Spaces*, Princetion University Press, Princeton, NJ, Vol. II, 1950.

[19] X.G. Xia, C.-C.J. Kuo and Z. Zhang, "Signal extrapolation in wavelet subspaces", *SIAM J. Scientific Comput.*, Vol. 16, January 1995, pp. 50–73.