System Identification Using Chirp Signals and Time-Variant Filters in the Joint Time-Frequency Domain

Xiang-Gen Xia, *Member, IEEE*

***Abstract*—In this paper, we propose a novel method to identify an unknown linear time invariant (LTI) system in low signal-to-noise ratio (SNR) environment. The method is based on transmitting chirp signals for the transmitter and using linear time-variant filters in the joint time-frequency (TF) domain for the receiver to reduce noise before identification. Due to the TF localization property of chirp signals, a large amount of additive white noise can be reduced, and therefore, SNR before identification can be significantly increased. This, however, cannot be achieved in the conventional methods, where pseudo-random signals are used, and therefore, noise reduction techniques do not apply. Our simulation results indicate that the method proposed in this paper outperforms the conventional methods significantly in low SNR environment. This paper provides a good application of time-frequency analysis and synthesis.**

I. Introduction

THE SYSTEM identification problem is a classical and important problem in signal processing, which has applications in many fields including channel estimation in wireless communications. There have been extensive studies on this problem; see, for example, [2], [3], [28], [31], and [32]. The problem can be stated as

$$y[n] = \sum_k h[n-k]x[k] + v[n] \tag{1.1}$$

where

- $x[k]$ transmitted signal;
- $h[n]$ impulse response of a linear time invariant (LTI) system (or channel);
- $v[n]$ additive noise;
- $y[n]$ received signal.

The problem is to identify the LTI system transfer function $H(\omega)$ of $h[n]$ given the input and the output signals $x[n]$ and $y[n]$.

The conventional method for solving the above problem is the least-squared solution method that is equal to the cross-spectral method in stationary cases, i.e., the system transfer function $H(\omega)$ can be estimated by

$$H(\omega) = \frac{S_{xy}(\omega)}{S_{xx}(\omega)} \tag{1.2}$$

Manuscript received November 11, 1996; revised April 3, 1997. This work was supported in part by an initiative grant from the Department of Electrical Engineering, University of Delaware, the Air Force Office of Scientific Research (AFOSR) under Grant F49620-97-1-0253, and the National Science Foundation CAREER Program under Grant MIP-9703377. The associate editor coordinating the review of this paper and approving it for publication was Dr. Yingbo Hua.

The author is with the Department of Electrical Engineering, University of Delaware, Newark, DE 19716 (e-mail: xxia@ee.udel.edu).

Publisher Item Identifier S 1053-587X(97)05791-7.

where $S_{xy}(\omega)$ is the cross-spectrum of $x[n]$ and $y[n]$, and $S_{xx}(\omega)$ is the autospectrum of $x[n]$. When the additive noise $v[n]$ in (1.1) is a zero-mean Gaussian process and statistically independent of the input signal $x[n]$, the estimate in (1.2) is asymptotically unbiased, and its error variance approaches the Cramer–Rao lower bound that is proportional to the variance of the additive noise $v[n]$. Clearly, the performance is limited by this noise variance, or the signal-to-noise ratio (SNR). When this SNR is low, the performance of the estimate in (1.2) is poor. Since the autospectrum of the input signal $x[n]$ is in the denominator in the estimate (1.2), the input signal is, in general, chosen as a pseudo-random signal with flat spectrum [4]. With these kinds of input signals, noise reduction techniques before system identification do not apply. As a matter of fact, any traditional noise reduction technique, such as any Fourier transform technique, does not perform well for wideband signals. This implies that it is not possible to increase the SNR or the performance of the estimate (1.2) by transmitting a pseudo random signal and using the conventional Fourier noise reduction techniques. Several questions arise here:

i) Can we transmit other wideband signals, such as chirp signals, instead of pseudo random signals?
ii) If so, can we take the advantage of these wideband signals and reduce the noise $v[n]$ in (1.1)?
iii) If so, can we improve the performance of the estimate (1.2) after denoising?

The aim of this paper is to positively answer these questions. The main idea is the following. Chirp-type signals are transmitted, which have wideband characteristics in the frequency domain but concentrate in the joint time-frequency domain. Chirp-type signals are used quite often, such as in radar and in FM in communications systems. The TF concentration property usually holds after an LTI system (this will be seen later). Since a joint TF distribution usually spreads noises and localizes signals, in particular chirps, the receiver may use a TF analysis technique (see, for example, [5]–[27]) to map the received signal $y[n]$ from the time domain into the joint time-frequency domain. In this way, the SNR can be significantly increased in the joint TF domain, and the receiver may be able to see patterns in the joint TF plane and therefore reduce the noise by filtering in the joint TF domain. This filtering is basically a *time-variant* filtering. We use this name in the rest of this paper. The model (1.1) after a time-variant filter

becomes

$$\tilde{y}[n] = \sum_k h[n-k]x[k] + \tilde{v}[n] \tag{1.3}$$

where $\tilde{v}[n]$ is the new noise after the filtering.

The time-variant filter used in this paper is based on the discrete Gabor transforms, which was studied in [5]–[7]. For chirp-type signals, about 13 dB SNR is increased consistently with this filter in [6]. When the original SNR in (1.1) is not too low, say, for example, above -1 dB, the new SNR in (1.3) may reach a significant high level so that the estimate of $H(\omega)$ from $\tilde{y}[n]$ and $x[n]$ is accurate enough for many applications. In this paper, both denoising with several mask design methods and system identification simulations are performed. These simulations show that a much better performance over the conventional method can be achieved.

It should be pointed out that the optimal training signal design for dynamic system identification has a long history dating back over 20 years. The design methods are traditionally based on the minimization of the Cramer–Rao bound for the system parameter estimation in either the time or the frequency domain (see, for example, [28]–[32]) but not in the joint TF domain. The aim of this paper is, however, not focused on the optimal training signal design, although it is a very interesting topic. Denoising before identification using nonredundant discrete wavelet transform was studied in [33] for chemical process control applications.

This paper is organized as follows. In Section II, we briefly review discrete Gabor transforms and the iterative time-variant filtering studied in [5]–[7]. In Section III, we use the time-variant filter studied in Section II to reduce additive white Gaussian noise for a received signal. The filtering problem in this paper has its own characteristics due to the fact that the transmitter and the receiver know the transmitted chirp signal $x[n]$, and therefore, its TF information is known *a priori*. This TF information can be used in designing a mask in the time-variant filtering. In Section IV, we utilize the conventional system identification method, i.e., the cross-spectral method (1.2), after the denoising in Section III. In Section V, we conclude this paper by addressing some possibilities for further improvements.

II. Discrete Gabor Transform and Time-Variant Filtering

There have been many TF analysis techniques, such as Wigner–Ville distributions in the Cohen's class, spectrogram (short-time Fourier transform or Gabor transform or DFT filterbanks), and scalogram (wavelets) (see, for example, [5], [23]–[27]). Some of them, such as bilinear TF distributions, have high resolution but have crossterms for multicomponent signals. Some of them, such as linear techniques (for example, Gabor transforms and wavelet transforms), do not have crossterms for multicomponent signals but may not have very high resolutions. Since, in this paper, we deal with a linear combination (or a linear system) of various chirp signals, it is important for a TF analysis technique not to have crossterms while it should also have a good resolution. This leads us to consider Gabor transforms. In this section, we first review the discrete Gabor transforms (DGT).

Since oversampled DGT is more robust for noise, it is usually used in noise reduction applications. However, a disadvantage for oversampled DGT is that it is not an onto mapping. In other words, not every signal $S[k,l]$ in the DGT transform domain corresponds to a time domain signal $s[n]$ so that the DGT of $s[n]$ is exactly equal to $S[k,l]$. This causes problems in filtering in the DGT transform domain, which is that the filtered signal in the DGT transform domain may not correspond to any time domain signal as shown in Fig. 1. An intuitive solution for this problem is to take the least-squared error (LSE) solution in the time domain (see, for example, [8]–[13]). The LSE, however, usually does not have a desired TF characteristics in the DGT transform domain. When a signal is very long, the computational load for the LSE solution is significantly high because of the inverse matrix computation. Based on these observations, an iterative algorithm was proposed in [5]–[7]. Conditions on the convergence, properties of the limit signals, and the relationship between the LSE solutions and solutions from the iterative algorithms were obtained in [6] and [7], where a significant improvement over the LSE solution was also shown. The second part of this section is to briefly review some of these results.

A. Discrete Gabor Transform

We first review some basics on the DGT, which is necessary for this paper. For more about the discrete short-time Fourier transform, see [14], for more about DFT filterbanks, see [15], and for more about the DGT, see, for example, [16]–[22]. Let a signal $s[k]$, a synthesis window function $h[n]$, and an analysis window function $\gamma[n]$ be all periodic with same period L. Then

$$s[k] = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} C_{m,n}h_{m,n}[k] \tag{2.1}$$

$$C_{m,n} = \sum_{k=0}^{L-1} s[k]\gamma^*_{m,n}[k] \tag{2.2}$$

$$h_{m,n}[k] = h[k-m\Delta M]W_L^{n\Delta Nk} \tag{2.3}$$

$$\gamma_{m,n}[k] = \gamma[k-m\Delta M]W_L^{n\Delta Nk} \tag{2.4}$$

and $W_L = \exp(j2\pi/L)$, $j=\sqrt{-1}$. The coefficients $C_{m,n}$ are called the DGT of the signal $s[k]$, and the representation (2.1) is called the *inverse DGT* (IDGT) of the coefficients $C_{m,n}$. One condition on the analysis and synthesis window functions $\gamma[k]$ and $h[k]$ obtained by Wexler and Raz is the identity[1]

$$\sum_{k=0}^{L-1} h[k+mN]W_L^{-nMk}\gamma^*[k] = \delta[m]\delta[n], \quad 0\le m\le \Delta N-1, \quad 0\le n\le \Delta M-1 \tag{2.5}$$

[1]If we take the inverse discrete Fourier transform with respect to the parameter n at the both sides, the system (2.5) is the same as the one obtained in [14] when all convolutions are considered to be cyclic convolutions for finite length signals in [14].

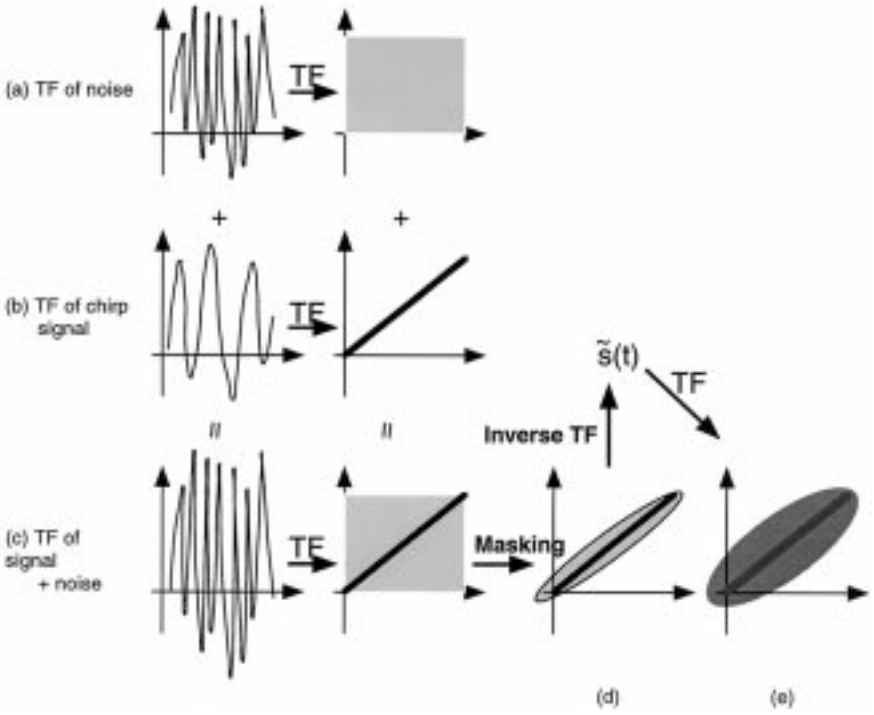

Fig. 1. TF transform illustration.

where ΔM and ΔN are the time and the frequency sampling interval lengthes, and M and N are the numbers of sampling points in the time and the frequency domains, respectively, $M \cdot \Delta M = N \cdot \Delta N = L$, $MN \geq L$ (or $\Delta M \Delta N \leq L$). The critical sampling case is when $M \cdot N = \Delta M \cdot \Delta N = L$. The condition (2.5) on window functions h and γ can be rewritten in matrix form as

$$H_{p\times L}\gamma^*_{L\times 1} = \mu_{p\times 1} \tag{2.6}$$

where the subscript $m \times n$ means the m by n matrix $p = \Delta M \cdot \Delta N, \gamma_{L\times 1} = (\gamma[0], \gamma[1], \cdots, \gamma[L-1])^T$, and $\mu_{p\times 1} = (1, 0, \cdots, 0)^T$ and the element at the $(m\Delta M + n)$th row and the kth column in the matrix $H_{p\times L}$ is

$$h[k + mN]W_L^{-nMk}, \qquad 0 \leq m \leq \Delta N - 1$$
$$0 \leq n \leq \Delta M - 1, \quad 0 \leq k \leq L - 1.$$

In the critical sampling case and when $H_{p\times L}$ has full rank, there is a unique solution for the analysis window function $\gamma[n]$. In the oversampling case and when $H_{p\times L}$ has full rank, there are infinite many solutions for the system (2.5). Among them, the minimum norm solution was given in [17]

$$\gamma^*_{L\times 1} = H^\dagger_{p\times L}(H_{p\times L}H^\dagger_{p\times L})^{-1}\mu_{p\times 1} \tag{2.7}$$

where † means the complex conjugate transpose. It was proved in [18]–[20] that the above minimum norm solution is also the most orthogonal-like solution, i.e., (a more general form was given in [22])

$$||\gamma_{L\times 1} - h_{L\times 1}|| = \min_{\hat{\gamma}_{L\times 1}:\, H_{p\times L}\hat{\gamma}^*_{L\times 1} = \mu_{p\times 1}} ||\hat{\gamma}_{L\times 1} - h_{L\times 1}||. \tag{2.8}$$

The DGT and IDGT can be also represented in matrix forms. Let

$$\boldsymbol{C} = (C_{0,0}, C_{0,1}, \cdots, C_{M-1,N-1})^T$$
$$\boldsymbol{s} = (s[0], s[1], \cdots, s[L-1])^T.$$

The DGT can be represented by the $MN \times L$ matrix $G_{MN\times L}$ with its $(mN + n)$th row and kth column element

$$\gamma^*_{m,n}[k] = \gamma^*[k - m\Delta M]W_L^{-n\Delta Nk}, \qquad 0 \leq m \leq M - 1$$
$$0 \leq n \leq N - 1, \quad 0 \leq k \leq L - 1.$$

The IDGT can be represented by the $L \times MN$ matrix $H_{L\times MN}$ with its kth row and $(mN + n)$th column element

$$h_{m,n}[k] = h[k - m\Delta M]W_L^{n\Delta Nk}, \qquad 0 \leq m \leq M - 1$$
$$0 \leq n \leq N - 1, \quad 0 \leq k \leq L - 1.$$

Thus

$$\boldsymbol{C} = G_{MN\times L}\boldsymbol{s} \quad \text{and} \quad \boldsymbol{s} = H_{L\times MN}\boldsymbol{C}. \tag{2.9}$$

The condition (2.5) implies that

$$H_{L\times MN}G_{MN\times L} = I_{L\times L} \tag{2.10}$$

where $I_{L\times L}$ is the $L \times L$ identity matrix.

B. Iterative Time-Variant Filtering Algorithm

We next want to briefly review the iterative time-variant filtering algorithm proposed in [5]–[7]. This algorithm is used later in the denoising for the system identification problem.

The oversampling of the DGT adds redundancy, which is usually preferred for noise reduction applications. From (2.1)–(2.5), (2.9), and (2.10), one can see that an L-dimensional signal $\boldsymbol{s}$ is transformed into an MN-dimensional

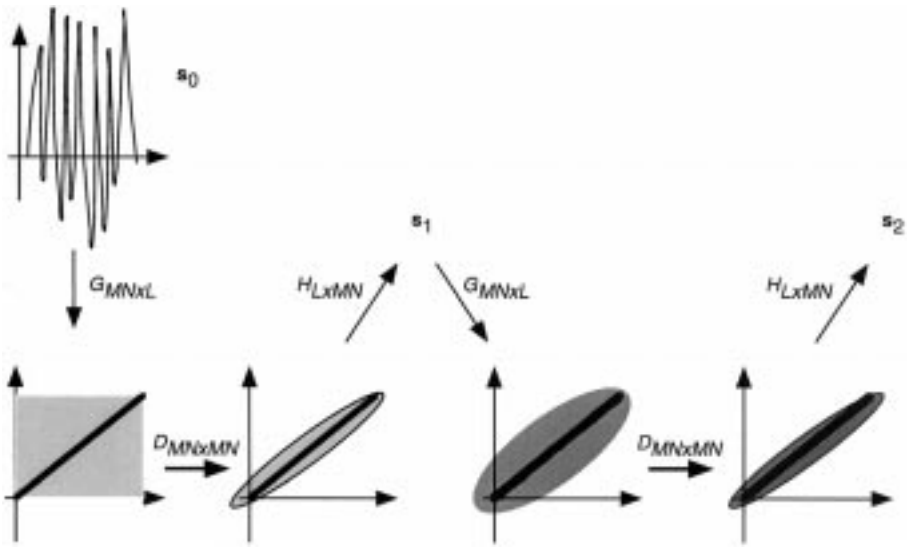

Fig. 2. Iterative time-varying filtering algorithm.

signal $\boldsymbol{C}$, and MN is greater than L due to the oversampling. Therefore, only a small set of MN-dimensional signals in the TF plane have their corresponding time waveforms with length L. Let $D_{MN\times MN}$ denote the mask transform, specifically, a diagonal matrix with diagonal elements either 0 or 1. Let $\boldsymbol{s}$ be a signal with length L in the time domain. The first step in the time-variant filtering is to mask the TF transform of $\boldsymbol{s}$

$$\boldsymbol{C}_1 = D_{MN\times MN} G_{MN\times L}\boldsymbol{s}$$

where $D_{MN\times MN}$ masks a desired domain in the TF plane. Since the DGT $G_{MN\times L}$ is a redundant transformation, the IDGT of $\boldsymbol{C}_1, H_{L\times MN}\boldsymbol{C}_1$ may not fall in the mask. In other words, in general

$$G_{MN\times L}H_{L\times MN}\boldsymbol{C}_1 \neq D_{MN\times MN}G_{MN\times L}H_{L\times MN}\boldsymbol{C}_1 \tag{2.11}$$

where $MN > L$, which is illustrated in Fig. 1(e). Notice that in the critical sampling case, i.e., $MN = L$, the inequality (2.11) becomes an equality. An intuitive method to reduce the difference between the right- and the left-hand sides of (2.11) is to mask the right-hand side of (2.11) again and repeat the procedure, which leads to the iterative algorithm

$$\boldsymbol{s}_0 = \boldsymbol{s} \tag{2.12}$$

$$C_{l+1} = D_{MN\times MN}G_{MN\times L}\boldsymbol{s}_l \tag{2.13}$$

$$S_{l+1} = H_{L\times MN}\boldsymbol{C}_{l+1}, \qquad l = 0, 1, 2, \cdots. \tag{2.14}$$

The above iterative algorithm is illustrated in Fig. 2.

Before going to the convergence, let us see what the LSE is. Based on the definition, the LSE solution is the $L \times 1$ vector $\overline{\boldsymbol{x}}$ that minimizes

$$\|G_{MN\times L}\overline{\boldsymbol{x}} - D_{MN\times MN}G_{MN\times L}\boldsymbol{s}\| = \min_{\boldsymbol{x}} \|G_{MN\times L}\boldsymbol{x} - D_{MN\times MN}G_{MN\times L}\boldsymbol{s}\|. \tag{2.15}$$

Then

$$\overline{\boldsymbol{x}} = (G^{\dagger}_{MN\times L}G_{MN\times L})^{-1}G^{\dagger}_{MN\times L}D_{MN\times MN}G_{MN\times L}\boldsymbol{s}. \tag{2.16}$$

Clearly, when the signal length L is large, the inverse matrix computation is expensive. Although the error in (2.15) is minimized, the DGT of the least-squared solution $\overline{\boldsymbol{x}}$ may not fall in the mask $D_{MN\times MN}$: $G_{MN\times L}\overline{\boldsymbol{x}} \neq D_{MN\times MN}G_{MN\times L}\overline{\boldsymbol{x}}$ when $MN > L$.

The complexity for the iterative algorithm (2.12)–(2.14) is, however, low, which does not need to compute inverses of large size matrices. By considering the DGT and IDGT in (2.1)–(2.4), the computational complexity in (2.12)–(2.14) is proportional to the signal length multiplied by the window length, i.e., LL_W. Notice that the complexity of directly computing the inverse matrices in (2.16) is proportional to L^3. Therefore, when the length of window functions h and γ is much shorter than the length of the signal $\boldsymbol{s}$, the computational complexity in the iterative algorithm (2.12)–(2.14) is much lower than the one for the least-squared solution in (2.16).

We next want to list several related results on the above iterative algorithm obtained in [6] and [7], such as the convergence, the properties of the limit signals, and the relationship between this algorithm and the LSE solution. These results are based on the condition on the window functions h and γ obtained in [6] and [7]:

$$\sum_{l=0}^{\Delta N-1} \gamma^*[lN+k]h[lN+k+m\Delta M] = \sum_{l=0}^{\Delta N-1} h^*[lN+k]\gamma[lN+k+m\Delta M] \tag{2.17}$$

for $k = 0, 1, \cdots, N-1$ and $m = 0, 1, \cdots, M-1$.

Theorem 1: When the synthesis and the analysis window functions $h[n]$ and $\gamma[n]$ satisfy condition (2.17), the iterative algorithm (2.12)–(2.14) converges.

There are two trivial cases where (2.17) holds. The first case is the orthogonal case $h[n] = \gamma[n]$ for all integer n. The second case is the critical sampling case $\Delta M = N$. Notice that the continuous Gabor transform is never orthogonal unless the window functions are badly localized in the frequency domain. This, however, is not the case for the DGT. The

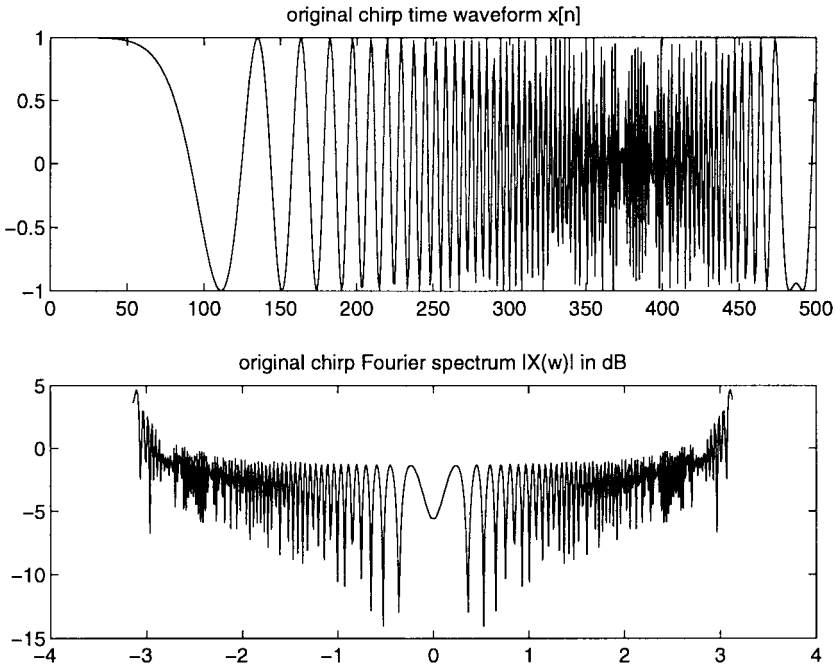

Fig. 3. Transmitted signal $x[n]$ and its Fourier spectrum $X(\omega)$.

most orthogonal-like solution was studied by Qian *et al.* in [18]–[20]. They showed that it is possible to have the analysis window function γ very close to the synthesis window function h when h is truncated Gaussian. The error between h and γ is less than 2×10^{-6} (see Fig. 4) while they are of unit energy, and therefore, the error is negligible. It was shown in [6] that the performance of the iterative algorithm strongly depends on (2.17). When this condition does not satisfy, the iterative algorithm may not converge.

Theorem 2: Under (2.17), the DGT of the limit $\overline{\boldsymbol{s}}$ of the iterative algorithm (2.12)–(2.14) falls in the mask $D_{MN\times MN}$, i.e.

$$G_{MN\times L}\overline{\boldsymbol{s}} = D_{MN\times MN}G_{MN\times L}\overline{\boldsymbol{s}}. \tag{2.18}$$

The above results say that as long as (2.17) on the analysis and synthesis window functions is satisfied, the iterative algorithm converges, and the limit signal has the desired TF characteristics, i.e., its DGT falls in the desired mask. One might ask whether it violates the known fact that an image of a TF transform of a signal in the TF plane cannot be compactly supported. This is because a signal cannot be time- and bandlimited simultaneously. To answer this question, we first need to know that the above known fact is true for continuous TF transforms. Moreover, the proof of the fact is based on the marginal properties of TF transforms. It may not be true for discrete TF transforms. In other words, discrete TF transforms may have compact support [5].

Theorem 3: Under (2.17), the first iteration $\boldsymbol{s}_1$ of the iterative algorithm (2.12)–(2.14) is equal to the least-squared solution in (2.16), i.e., $\boldsymbol{s}_1 = \overline{\boldsymbol{x}}$.

With this result, one will see later that the iterative algorithm (2.12)–(2.14) improves the least-squared solution when the number of iterations increases, and meanwhile, one does not need to compute the inverse matrix in (2.16).

III. Denoising for Received Signals Through a Noisy Channel

In this section, we want to do noise reduction with the time-variant filter studied in Section II for received signals in a noisy channel.

A. Some Parameters

The signal length is randomly chosen as 500. The signal $x[n]$ for the transmitter is

$$x[n] = \cos\left(\left[\frac{n+15}{150}\right]^4\right), \qquad n = 0, 1, \cdots, 499. \tag{3.1}$$

The waveform and its Fourier transform $X(\omega)$ of the above signal $x[n]$ are shown in Fig. 3. Notice that since the Fourier power spectrum $|X(\omega)|^2$ will be used in the denominator in the system identification, it should be as far away from zero as possible. Since the noise-reduction performance of the time-variant filtering in Section II depends on the localization of the signal in the TF plane, the transmitted signal $x[n]$ should be as concentrated in the joint time and frequency domain as possible. The synthesis and analysis window functions used in this paper are shown in Fig. 4, where their lengthes are 256. The synthesis window function is just the Gaussian function and its analysis window function is the most orthogonal-like solution given in (2.7). Their difference and the difference between the left-hand side and the right-hand side of (2.17), i.e., the condition error, are also shown in Fig. 4. One can see that they almost satisfy (2.17). The time sampling interval length $\Delta M = 16$ and the frequency sampling interval length $\Delta N = 2$ in the discrete Gabor transform and its inverse in Section II. These parameters are used throughout the rest of this paper. The DGT of $x[n]$ is shown in Fig. 5. The tail part

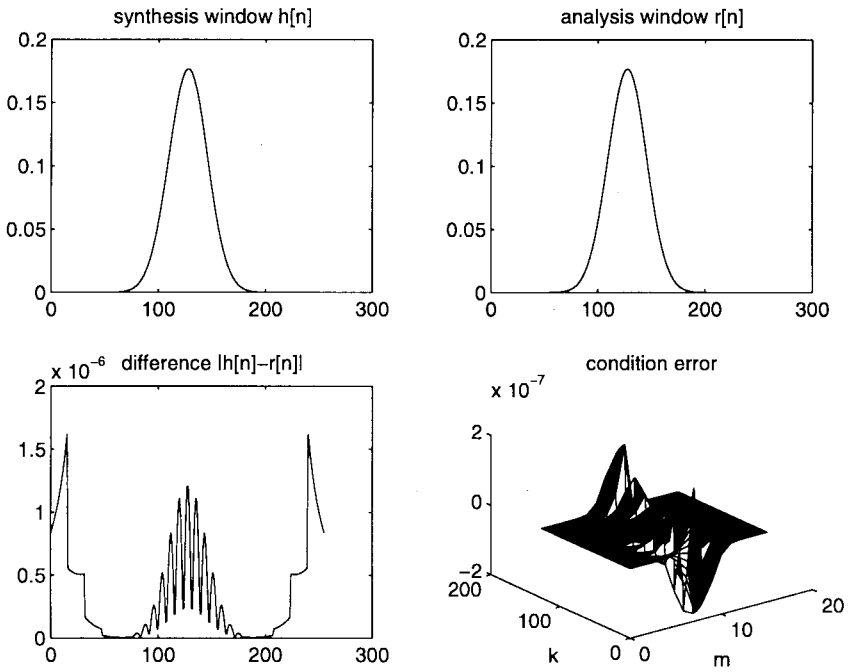

Fig. 4. Synthesis and analysis window functions and the condition (2.17) test.

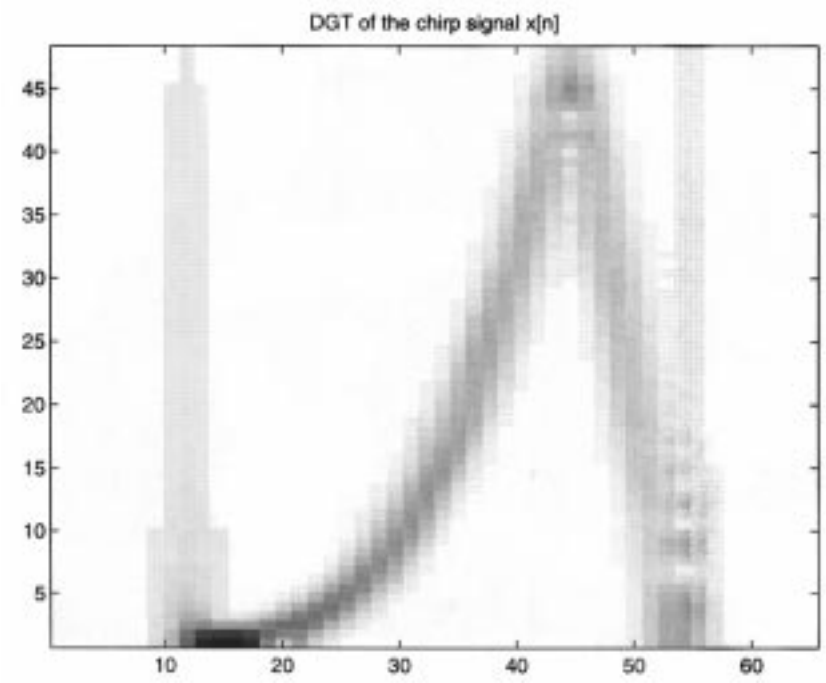

Fig. 5. Discrete Gabor transform of signal $x[n]$.

of the DGT in Fig. 5 is because of the discrete calculation aliasing.

In this paper, we use 20-tap LTI systems in our numerical examples, where the number 20 is just randomly chosen. The channel model is

$$y[n] = \sum_{k=0}^{N-1} h[k]x[n-k] + v[n] \tag{3.2}$$

where $N = 20$ in the following numerical examples, $v[n]$ is an additive white Gaussian noise and independent of the signal $x[n]$, and

$$s[n] = \sum_{k=0}^{N-1} h[k]x[n-k] \tag{3.3}$$

is considered to be the signal, $x[n]$ is the transmitted signal as in (3.1), $y[n]$ is the received signal, and $h[n]$ is an LTI system (or channel) impulse response. The original SNR for

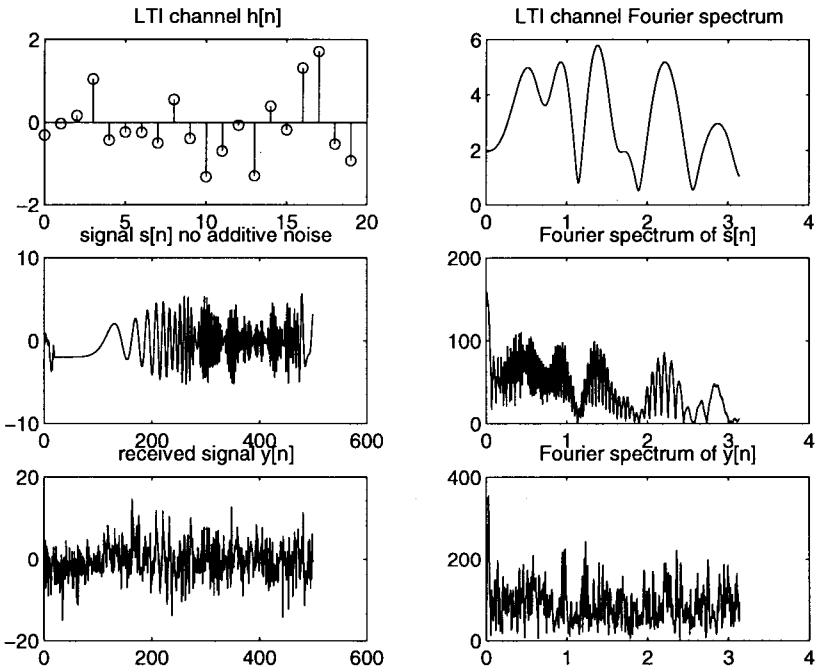

Fig. 6. Example of LTI channel $h[n]$, signal $s[n]$, and received signal $y[n]$ and their Fourier spectrum, where the SNR $= -4.5$ dB for the additive white Gaussian noise.

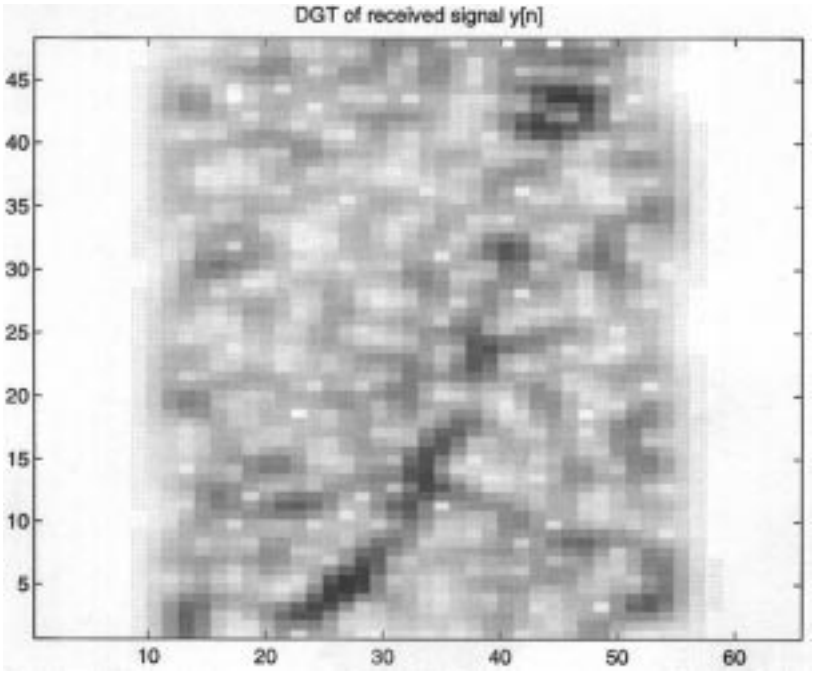

Fig. 7. Discrete Gabor transform of the received signal $y[n]$ in Fig. 6 with SNR $= -4.5$ dB.

the received signal is calculated by

$$10\log_{10}\left(\frac{\sum_{n=0}^{499}|s[n]|^2}{\sum_{n=0}^{499}|v[n]|^2}\right).$$

In the following, we randomly generate the channel $h[n]$. As an example, a channel Fourier spectrum and received signal time waveform $y[n]$ with SNR $= -4.5$ dB and the signal $s[n]$ without noise and their Fourier spectrum are shown in Fig. 6. The DGT of the received signal $y[n]$ with -4.5 dB SNR is shown in Fig. 7. In Fig. 7, one is still able to see the chirp pattern in the joint time and frequency plane, although it is impossible in the time or the frequency domain alone in Fig. 6.

B. Mask Design

The pattern in the DGT domain of the above signal $s[n]$ in (3.3) is similar to the one for the signal $x[n]$ in Fig. 5. This is

not only true for this particular example but is also the case for our numerous examples. The reason is due to the following analytic argument.

Assume the chirp signal $x[n] = \exp(jcn^r)$ for some constants $r \geq 2$ and $c \neq 0$. Then

$$\begin{aligned} s[n] &= \sum_k h[k]x[n-k] \\ &= \sum_k h[k]\exp(jc(n-k)^r) \\ &= \exp(jcn^r)\sum_k h[k]\exp\left(jc\sum_{l=0}^{r-1} c_l n^l k^{r-l}\right) \\ &= x[n]\sum_k h[k]\exp\left(jc\sum_{l=0}^{r-1} c_l n^l k^{r-l}\right) \end{aligned}$$

which is dominated by the original chirp $x[n]$ for finite tap LTI systems $h[k]$. It is because that the highest chirp order of $s[n], r$, and the corresponding chirp rate are the same as those of $x[n]$, whereas the chirp order for the above multiplier of $x[n]$ in $s[n]$

$$\sum_k h[k]\exp\left(jc\sum_{l=0}^{r-1} c_l n^l k^{r-l}\right)$$

is only $r-1$. As a special case, when $r = 2$

$$s[n] = x[n]G(2cn)$$

where $G(\omega)$ is the Fourier transform of the signal $h[n]x[n]$

$$G(\omega) = \sum_k h[k]x[k]\exp(-j2cnk).$$

When the channel $h[n]$ has only a finite tap, the function $G(\omega)$ is usually a smooth signal.

Since the transmitted signal $x[n]$ is known to both transmitter and the receiver, by the above property, its pattern in the DGT domain may help in designing a mask in the DGT domain for filtering noise. This is exactly the motivation for the following design method of a mask $D_{MN\times MN}$ in the iterative time-variant algorithm (2.12)–(2.14). The subscript $MN \times MN$ of the mask $D_{MN\times MN}$ will be dropped from now on without causing confusion in understanding.

1) Mask Design Procedure:

Step 1) Implement the DGT $C_{m,n}$ of the transmitted signal $x[k]$.

Step 2) Threshold the DGT coefficients $C_{m,n}$ and have a mask D_x from $C_{m,n}$

$$D_x(m,n) = \begin{cases} 1, & \text{if } |C(m,n)| > t_0 \\ 0, & \text{otherwise} \end{cases}$$

where t_0 is a predesigned positive number that is called *thresholding constant*.

Step 3) Implement Steps 1 and 2 for the received signal $y[k]$, and design a mask D_y with thresholding constant t_1 from the DGT coefficients of $y[n]$ with another predesigned constant $t_1 > 0$.

Step 4) The final mask is the product of D_x and D_y: $D = D_xD_y$.

Since the DGT of the signal $x[n]$ usually dominates the DGT of the signal $s[n]$, the pattern in the DGT domain of the signal $s[n]$ is usually in a close neighborhood of the pattern in the DGT domain of $x[n]$. Therefore, the mask D_x is usually designed so that it covers a relatively large area, i.e., the thresholding constant t_0 in Step 2 is usually chosen not too large. Since the received signal $y[n]$ is from a noisy channel, the resolution of its DGT pattern may be reduced, and therefore, the thresholding constant t_1 in Step 3 is usually chosen to be not too small. Otherwise, the mask D_y will cover too much unwanted area. Let us see an example. The mask D_x from $x[n]$, the mask D_y from $y[n]$, their product $D = D_xD_y$, and the mask D_s from the true signal $s[n]$ are shown in Fig. 8, respectively. The SNR in this case is SNR $= -1.4$ dB. The thresholding constants in Steps 1–3 are $t_0 = 0.12$ and $t_1 = 0.15 \cdot \max(\text{DGT}(y))$. It should be pointed out that the above mask design procedure may be improved by using more sophisticated designs. Possible improvements are

i) to find the optimal thresholding constants t_0 and t_1 by training a large number of signals and systems;

ii) to use more sophisticated statistical detection method in the DGT domain for the received signal $y[n]$ instead of a simple thresholding in Step 3;

iii) to smooth the mask $D = D_xD_y$ since the true mask D_s is usually smooth due to the nature of a chirp signal, but D_y from the noisy signal $y[n]$ may not be smooth. Some morphological operations, such as dilation, may be used to smooth the mask D.

Another observation from our various numerical examples is that the mask D_x is the mean of the true mask D_s in terms of different LTI systems $h[n]$.

C. Denoising Experiments

In this subsection, we want to implement the time-variant filtering algorithm in Section II with three masking techniques: using the mask $D = D_x$ from the transmitted signal, using the mask $D = D_yD_x$ as designed by Steps 1–4, using the true mask $D = D_s$. We run 100 tests in terms of different LTI systems $h[n]$ (randomly generated) and different additive white Gaussian noises $v[n]$ for each masking method and take their mean SNR. Nine iterative steps are used in the iterative algorithm (2.12)–(2.14). Fig. 9 shows the curves of the mean SNR versus iterative steps for the three masking methods.

First, we analyze the time-variant filter (2.12)–(2.14) with the mask $D = D_x$. From Fig. 9, the SNR drops after the second iteration. This is because the mask we used is $D = D_x$, which matches the transmitted signal $x[n]$ and not $s[n]$. Although there is a similarity (see Fig. 8) in the TF plane between the DGT of $x[n]$ and the DGT of $s[n]$, they are not equal. The similarity is exactly the reason why the SNR increases significantly in the first and the second iteration step. The difference between $x[n]$ and $s[n]$ causes the SNR to drop after the second iteration. Notice that the mask $D = D_x$ is known to the receiver, and it is a good candidate in the time-variant filtering if the iterative algorithm stops at the second iteration step.

We now analyze the performance of the mask $D = D_xD_y$. This mask rejects a lesser portion of the noise outside D_s than D_x alone does, when the first thresholding constant t_0 for D_x

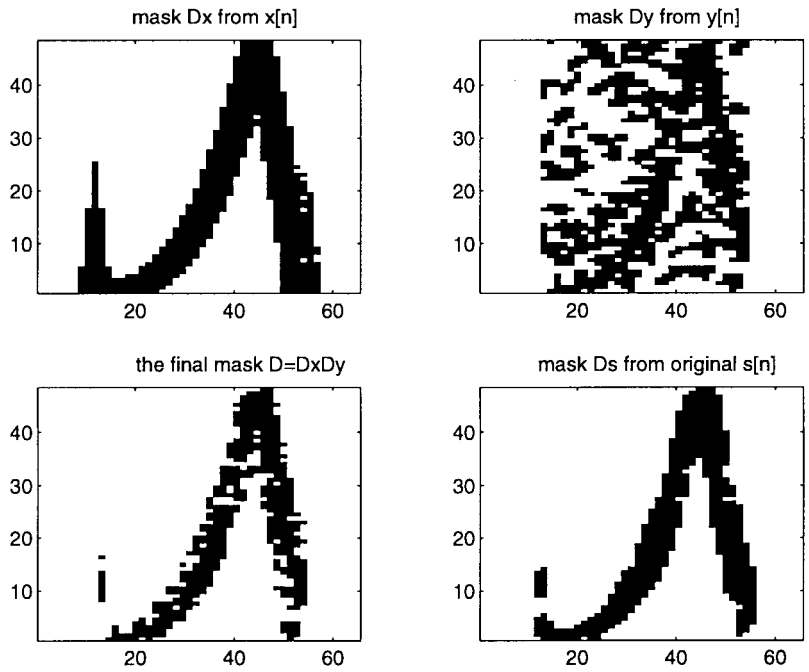

Fig. 8. Example of masks D_x from $x[n]$, D_y from $y[n]$, the final mask $D = D_x D_y$, and the true mask D_s from $s[n]$, where the SNR $= -1.4$ dB.

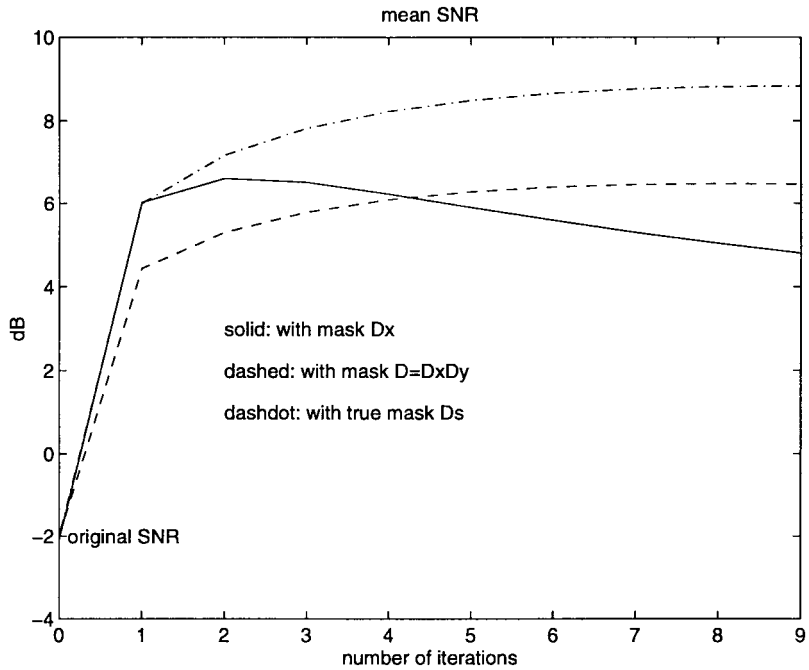

Fig. 9. Mean SNR curves of the iterative time-variant filtering with the following masks: $D = D_x$, $D = D_x D_y$, and $D = D_s$.

in Step 2 is less than the one in designing D_x alone. The reason why this t_0 for D should not be large is for the conservation because the mask D_x is multiplied by D_y in designing D. It, however, happens because the beginning SNR's are not as high as the ones in the time-variant filtering with the mask D_x, which is shown by the solid line in Fig. 9. Since, in general, $D = D_x D_y$ covers relatively more signal information than D_x alone does, the SNR increases when the iteration number increases.

The third masking $D = D_s$ method is the ideal case. With this ideal mask, about an 11 dB SNR increase with the iterative time-variant filtering over the original SNR is achieved consistently. Notice that by Theorem 3, the first iteration is equal to the conventional least squared solution. The iterative time-variant filtering outperforms the least squared solution by about 3 dB.

To improve the performance of the iterative time-variant filtering, what one can do further is to use more sophisticated methods to detect D_x and D_y, in particular D_y, so that their product $D = D_x D_y$ is as close to D_s as possible. Besides what has been mentioned in the previous subsections, directly minimizing the difference between $D = D_x D_y$ and

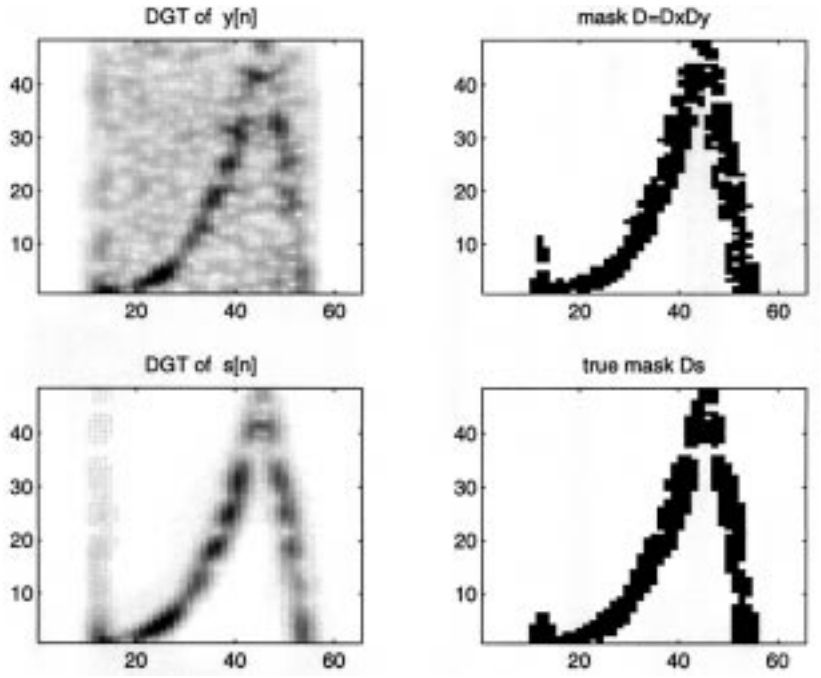

Fig. 10. DGT of $y[n]$ with noise and $s[n]$ without noise and their corresponding masks (original SNR $=$2.7 dB).

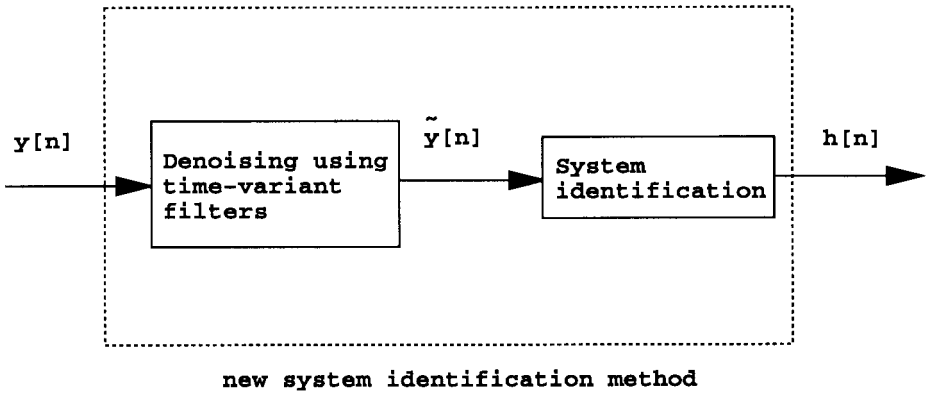

Fig. 11. New system identification method.

D_s with training signals is another potential approach. When the original SNR is not too low, the chirp pattern of $s[n]$ can usually be seen clearly in the DGT domain of the received signal $y[n]$. An example is shown in Fig. 10, where the original SNR $=$ 2.7 dB.

IV. System Identification

In this section, we first use the iterative time-variant filter (2.12)–(2.14) developed in the previous sections to reduce the additive white Gaussian noise $v[n]$ from the received signal $y[n]$. In the iterative time-variant filter, for calculation simplicity, we choose the first masking method studied in Section III-C, i.e., the mask $D = D_x$, for all calculations in this section. With this mask, two iterations are used in the time-variant filter in Section II-B. We then implement the conventional system identification method, as shown in Fig. 11.

The conventional system identification method used here is the cross-spectral method

$$H_{\text{new}}(\omega) = \frac{S_{\tilde{y}x}(\omega)}{S_{xx}(\omega)} \tag{4.1}$$

where $x[n]$ is the chirp signal defined in (3.1). It is compared with the conventional method without denoising, i.e.,

$$H_{\text{old}_1}(\omega) = \frac{S_{yx}(\omega)}{S_{xx}(\omega)} \tag{4.2}$$

where $x[n]$ is also the chirp signal. Since the system identification performance usually depends on the signal $x[n]$ transmitted, one might say that it is not fair to compare them using the chirp signal that is preferred here for denoising but might not be preferred for other methods. For this reason, we also compare our new method with the conventional method

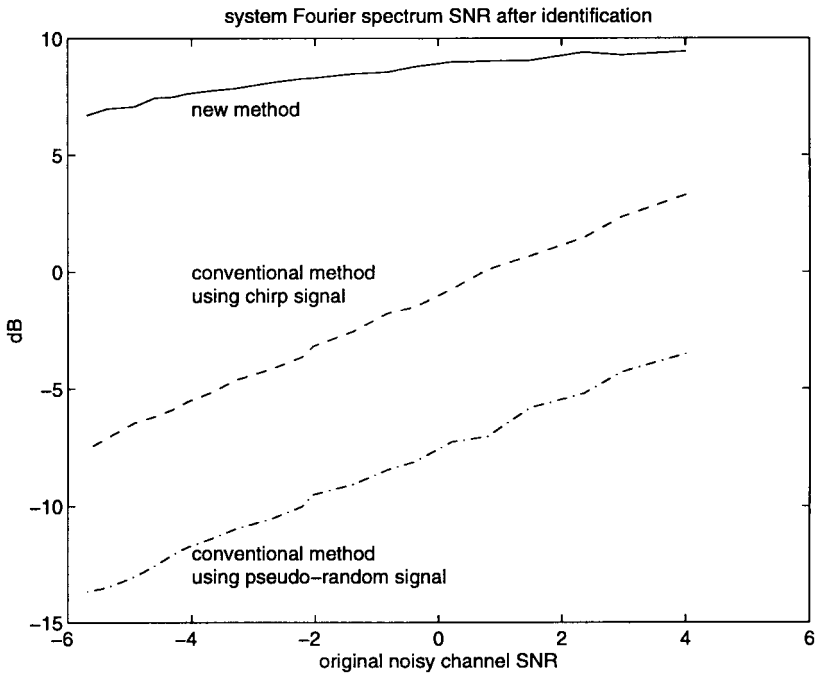

Fig. 12. Comparison of system identification methods. The conventional method using chirp signals; the conventional method using pseudo-random signals; new method using chirp signals, and time-variant filtering.

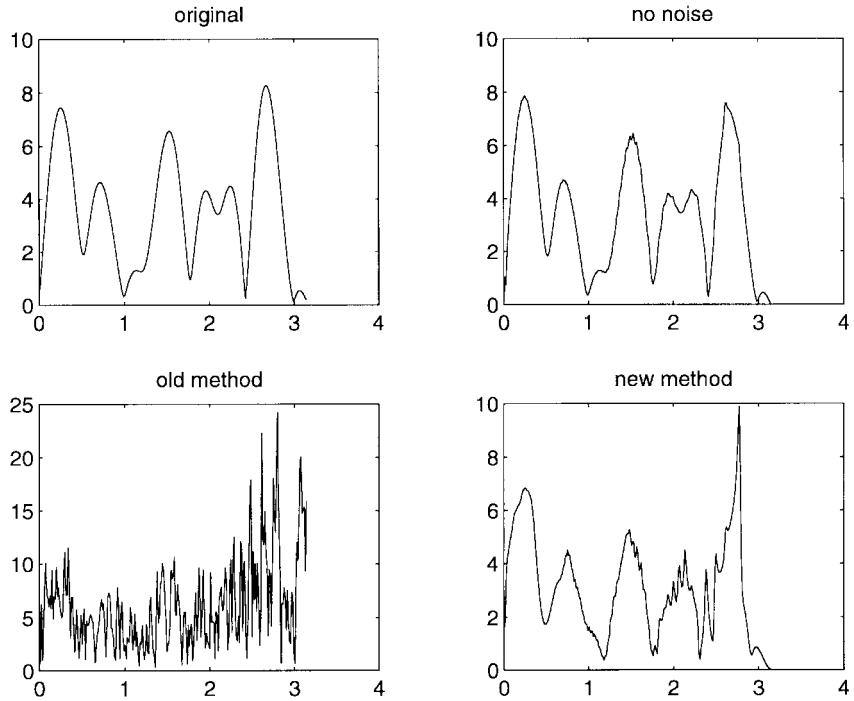

Fig. 13. System identification examples: Original spectrum $|H(\omega)|$; identified spectrum without additive noise using the chirp signal; conventional method with additive noise of SNR $= -0.4$ dB; new method with additive noise of SNR $= -0.4$ dB.

using pseudo-random sequences

$$H_{\text{old}_2}(\omega) = \frac{S_{y\tilde{x}}(\omega)}{S_{\tilde{x}\tilde{x}}(\omega)} \tag{4.3}$$

where $\tilde{x}[n]$ is a pseudo-random sequence.

Fig. 12 shows their performances, where 200 tests are used for the mean SNR curves for the system spectrum versus the original SNR. Our new method performs much better than others. Surprisingly, even for the conventional cross spectral method, the chirp signal in (3.1) outperforms pseudo-random signals by about 6 dB. In Fig. 13, some identification examples are shown, where the original SNR is -0.4 dB. As a remark, all system identification calculations used in this paper are based on the Matlab Signal Processing Toolbox.

V. Conclusion

In this paper, we proposed a system identification method. The proposed method is based on transmitting chirp signals

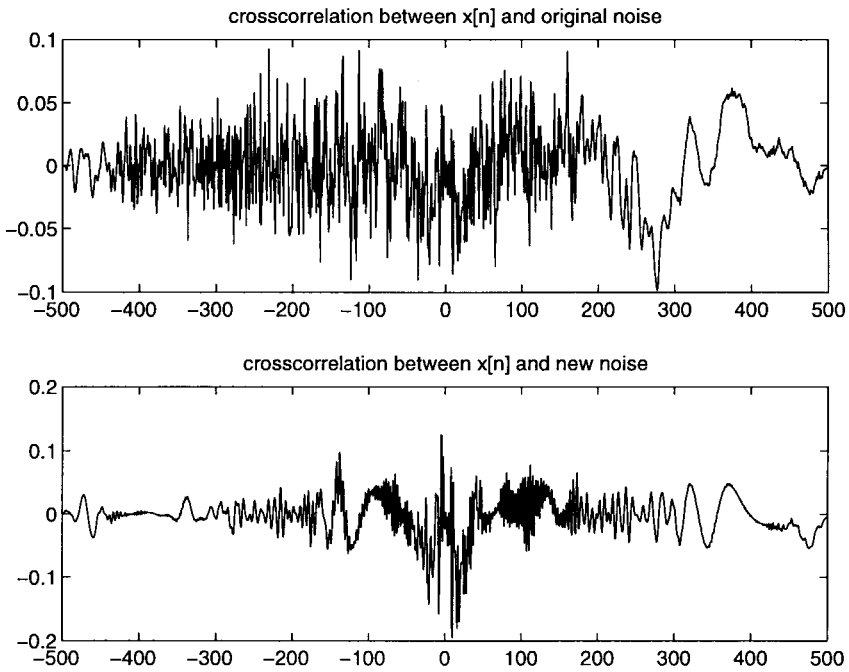

Fig. 14. Cross correlations between the new noise $\bar{v}[n]$ (SNR $= 0.74$ dB) and the signal $x[n]$ and the original noise $v[n]$ and the signal $x[n]$ (SNR $= -6.4$ dB).

and denoising followed by the conventional identification method. The denoising method is based on time-variant filtering in the joint time-frequency (TF) domain. Since transmitted signals are chirp-type signals, they are well-localized in the TF domain, and one is usually able to see their patterns in the TF domain, even in a very low SNR environment. Due to this property, a significant SNR increase after a time-variant filtering can be achieved. Our numerical simulations were performed to illustrate this theory. The simulations done in this paper were used simply for showing the potential performance of the new approach based on time-frequency analysis and synthesis techniques in very low SNR environment. Several further improvements are possible. They are

i) to use more sophisticated detection methods in designing masks D for the iterative time-variant filter;
ii) to search the optimal reference signal $x[n]$ so that its Fourier spectrum is as far away from 0 as possible and it localizes in the TF domain as much as possible;
iii) to use more sophisticated existing system identification methods, such as the method recently proposed in [1] by Shalvi and Weinstein, where the additive noise $v[n]$ in the system model is not necessarily independent of the signal $x[n]$.

The reason for mentioning iii) here is because of the following argument. Since a joint TF domain filter that usually depends on the signal $x[n]$ is used, the new noise $\tilde{v}[n]$ after denoising and the transmitted signal $x[n]$ may have similar TF characteristics, and therefore, they may be correlated, in particular, when the original SNR is too low. Such an example is shown in Fig. 14, where the original SNR $= -6.4$ dB and the SNR $= 0.74$ dB after the second iteration of the time-variant filtering. From Fig. 14, one can clearly see that the correlation between the new noise $\tilde{v}[n]$ after denoising and the signal $x[n]$ exists, whereas it does not exist between the original noise $v[n]$ and $x[n]$. It should be observed from our numerous numerical examples that this phenomenon only happens when the original SNR is very low.

ACKNOWLEDGMENT

The author would like to thank Prof. G. Arce and Prof. C. Boncelet for their encouragement and various discussions on the subject. He also wishes to thank the referees and the associate editor for bringing [28]–[33] to his attention.

REFERENCES

[1] O. Shalvi and E. Weinstein, "System identification using nonstationary signals," *IEEE Trans. Signal Processing*, vol. 44, pp. 2055–2063, Aug. 1996.
[2] L. Ljung, *System Identification Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
[3] T. Söderström and P. Stoica, *System Identification*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[4] K. Pahlavan and J. W. Matthews, "Performance of adaptive matched filter receivers over fading multipath channels," *IEEE Trans. Commun.*, vol. 38, pp. 2106–2113, Dec., 1990.
[5] S. Qian and D. Chen, *Joint Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[6] X.-G. Xia and S. Qian, "Discrete Gabor transform based time-variant filters," preprint, 1996.
[7] ——, "An iterative algorithm for time-variant filtering in the discrete Gabor transform domain," in *Proc. IEEE ICASSP'97*, Munich, Germany, Apr. 1997.
[8] C. Wilcox, "The synthesis problem for radar ambiguity functions," Tech. Summary Rep. 157, Math. Res. Cent., Univ. Wisconsin, Madison, Apr. 1960.
[9] G. F. Boudreaux-Bartels and T. W. Parks, "Time-varying filtering and signal estimation using Wigner distribution synthesis techniques," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 442–451, June 1986.
[10] S. Farkash and S. Raz, "Time-variant filtering via the Gabor expansion," in *Signal Processing V: Theories and Applications*. New York: Elsevier, 1990, pp. 509–512.
[11] F. Hlawatsch and W. Krattenthaler, "Bilinear signal synthesis," *IEEE Trans. Signal Processing*, vol. 40, pp. 352–363, Feb. 1992.

[12] W. Kozek and F. Hlawatsch, "A comparative study of linear and nonlinear time-frequency filters," in *Proc. IEEE Int. Symp. Time-Freq. Time Scale Anal.*, Victoria, B.C., Canada, Oct. 1992, pp. 163–166.
[13] F. Hlawatsch, A. H. Costa, and W. Krattenthaler, "Time-frequency signal synthesis with time-frequency extrapolation and don't-care regions," *IEEE Trans. Signal Processing*, vol. 42, pp. 2513–2520, Sept. 1994.
[14] M. R. Portnoff, "Time-frequency representation of digital signals and systems based on short-time Fourier analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 55–69, Feb. 1980.
[15] R. E. Crochiere and L. R. Rabiner, *Multirate Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1983.
[16] M. J. Bastiaans, "Gabor's expansion of a signal into Gaussian elementary signals," *Proc. IEEE*, vol. 68, pp. 594–598, Apr. 1980.
[17] J. Wexler and S. Raz, "Discrete Gabor expansions," *Signal Process.*, vol. 21, pp. 207–220, 1990.
[18] S. Qian and D. Chen, "Discrete Gabor transform," *IEEE Trans. Signal Processing*, vol. 41, pp. 2429–2438, July 1993.
[19] ——, "Optimal biorthogonal analysis window function for discrete Gabor transform," *IEEE Trans. Signal Processing*, vol. 42, pp. 694–697, Mar. 1994.
[20] S. Qian, K. Chen, and S. Li, "Optimal biorthogonal functions for finite discrete-time Gabor expansion," *Signal Process.*, vol. 27, pp. 177–185, 1992.
[21] A. J. E. M. Janssen, "Duality and biorthogonality for discrete-time Weyl-Heisenberg frames," RWR-518-RE-94001-ak unclassified rep. 002/94, Philips Res. Lab., Eindhoven, The Netherlands, 1994.
[22] X.-G. Xia, "On characterization of the optimal biorthogonal window functions for Gabor transforms," *IEEE Trans. Signal Processing*, vol. 44, pp. 133–136, Feb. 1996.
[23] L. Cohen, "Time-frequency distributions—A review," *Proc. IEEE*, vol. 77, pp. 941–981, July 1989.
[24] ——, *Time-Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.
[25] F. Hlawatsch and G. F. Bourdeaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Processing Mag.*, vol. 9, pp. 21–67, Apr. 1992.
[26] O. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Processing Mag.*, vol. 8, pp. 14–38, Oct. 1991.
[27] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, pp. 961–1005, Sept. 1990.
[28] C. C. Goodwin and L. Payne, *Dynamic System Identification Experiment Design and Data Analysis*. London, U.K.: Academic, 1977.
[29] R. K. Mehra, "Optimal input signals for parameter estimation in dynamic systems-survey and new results," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 753–768, 1974.
[30] E. Rafajlowic, "Unbounded power input signals in optimum experiment design for parameter estimation in linear system," *Int. J. Contr.*, vol. 40, pp. 383-391, 1984.
[31] B. Kuszta and N. K. Sinha, *Modeling and Identification of Dynamic Systems*. New York: Van Nostrand Reinhold, 1983.
[32] M. B. Zarrop, *Optimal Experiment Design for Dynamic System Identification*. Berlin-New York: Spinger-Verlag, 1979.
[33] S. Palavajjhala, R. L. Motard, and B. Joseph, "Process identification using discrete wavelet transform: design of prefilters," *AIChe J.*, vol. 42, pp. 777–790, Mar. 1996.

Xiang-Gen Xia (M'97) received the B.S. degree in mathematics from Nanjing Normal University, Nanjing, China, the M.S. degree in mathematics from Nankai University, Tianjin, China, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1983, 1986, and 1992, respectively.

He was a Lecturer at Nankai University from 1986 to 1988, a Teaching Assistant at the University of Cincinnati, Cincinnati, OH, from 1988 to 1990, a Research Assistant at the University of Southern California from 1990 to 1992, and a Research Scientist at the Air Force Institute of Technology, Wright-Patterson AFB, OH, from 1993 to 1994. He was a Senior/Research Staff Member at Hughes Research Laboratories, Malibu, CA, from 1995 to 1996. In September 1996, he joined the Department of Electrical Engineering, University of Delaware, Newark, where he is currently an Assistant Professor. His current research interests include communication systems including equalization and coding, wavelet transform and multirate filterbank theory and applications, time-frequency analysis and synthesis and numerical analysis and inverse problems in signal/image processing.

Dr. Xia received the National Science Foundation Faculty Early Career Development (CAREER) Program Award in 1997. He is currently an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is also a member of the American Mathematical Society.