

Article

A Steady-State Kalman Predictor-Based Filtering Strategy for Non-Overlapping Sub-Band Spectral Estimation

Zenghui Li ¹, Bin Xu ¹, Jian Yang ^{1,*} and Jianshe Song ²

¹ Department of Electronic Engineering, Tsinghua University, Beijing 100084, China; E-Mails: lizenghui11@mails.tsinghua.edu.cn (Z.L.); b-xu11@mails.tsinghua.edu.cn (B.X.)

² Xi'an Research Institute of Hi-Technology, Xi'an 710025, China; E-Mail: Songjianshe09@126.com

* Author to whom correspondence should be addressed; E-Mail: yangjian.ee@gmail.com; Tel.: +86-10-6279-4726.

Academic Editor: Vittorio M.N. Passaro

Received: 20 October 2014 / Accepted: 17 December 2014 / Published: 24 December 2014

Abstract: This paper focuses on suppressing spectral overlap for sub-band spectral estimation, with which we can greatly decrease the computational complexity of existing spectral estimation algorithms, such as nonlinear least squares spectral analysis and non-quadratic regularized sparse representation. Firstly, our study shows that the nominal ability of the high-order analysis filter to suppress spectral overlap is greatly weakened when filtering a finite-length sequence, because many meaningless zeros are used as samples in convolution operations. Next, an extrapolation-based filtering strategy is proposed to produce a series of estimates as the substitutions of the zeros and to recover the suppression ability. Meanwhile, a steady-state Kalman predictor is applied to perform a linearly-optimal extrapolation. Finally, several typical methods for spectral analysis are applied to demonstrate the effectiveness of the proposed strategy.

Keywords: AR model; equiripple FIR filter; linear prediction; spectral estimation; spectral overlap; sub-band decomposition

1. Introduction

As one of the most important tools, spectral estimation [1] has been extensively applied in radar, sonar and control systems, in the economics, meteorology and astronomy fields, speech, audio, seismic and

biomedical signal processing, and so on. In particular, sparse representation [2–4] opens an exciting new vision for spectral analysis. However, such methods are usually accompanied by high computational complexity, which makes their availability somewhat limited.

Sub-band decomposition-based spectral estimation (SDSE) [5] is an important research direction in spectral estimation, because it has several advantageous features, e.g., computational complexity decrease, model order reduction, spectral density whiteness, reduction of linear prediction error for autoregressive (AR) estimation and the increment of both frequency spacing and local signal-to-noise ratio (SNR) [6]. These features have been theoretically demonstrated under the hypothesis of the ideal infinitely-sharp bandpass filter bank [7]. Subsequent studies [8–10] indicate that these benefits aid complex frequency estimation in sub-bands, thereby enabling better estimation performance than that achieved in full-band. In addition, the computational complexity of most algorithms for spectral analysis has a superlinear relationship with the data size, and sub-band decomposition can considerably speed up these algorithms. Independently handling each sub-band enables parallel processing, which can further improve the computational efficiency. Both advantages are crucial for reducing the computational burden, especially when analyzing multi-dimensional big data, such as polarimetric and/or interferometric synthetic aperture radar images of large scenes.

Unfortunately, the ideal infinitely-sharp bandpass filter cannot be physically realized, and non-ideal (realizable) filters introduce energy leakage and/or frequency aliasing phenomena [11]. Due to these non-ideal frequency characteristics of analysis filters, spectral overlap between any two contiguous sub-bands occurs during the sub-band decomposition. Then, the performance of SDSE severely degrades.

In the relevant literature, several methods have been proposed to mitigate spectral overlap. We classify these methods into three categories. The first category is defined as ideal frequency domain filtering with a strict box-like spectrum, such as “ideal” Hilbert transform-based half-band filters [9] and harmonic wavelet transform-based filters [12,13]. Theoretically, sub-band decomposition with these filters is immune to spectral overlap. However, discrete Fourier transform will inevitably induce spectral energy leakage, which can likewise distort sub-band decomposition. The second category is known as convolution filtering with wavelet packet filters [8], Kaiser window-based prototype cosine modulated filters, discrete cosine transform (DCT) IV filters [10] and Comb filters [6,14]. It seems that increasing the filter order can improve the filtering performance and also the spectral overlap suppression capability. However, in the context of involving a finite-length sequence and performing convolution filtering, the nominal improvement of performance will lead to spectral energy leakage and inferior filtering accuracy [10]. Considering the compromise between suppressing spectral overlap and reducing spectral energy leakage, we have to restrict the filter order. The third category is frequency-selective filtering, and a representative method is SELF-SVD (singular value decomposition-based method in a selected frequency band) [15]. Essentially, SELF-SVD attempts to attenuate the interferences of the out-of-band components by the post-multiplication with an orthogonal projection matrix. Unfortunately, the attenuation is often insufficient when the out-of-band components are much stronger than the in-band components or the SNR is relatively low. In this case, the estimation of the in-band frequencies is seriously affected.

In this paper, a new filtering strategy is proposed to suppress spectral overlap for sub-band spectral estimation. First, we discuss the formation mechanism of spectral overlap. Nominally, a high-order finite impulse response (FIR) filter usually has a powerful ability in spectral overlap suppression. However, once we perform such a filter on a finite-length sequence with the convolution operation, the non-given samples at the forward and backward sampling periods of the sequence are assumed to be zeros. A certain filtering error therefore occurs and conversely disrupts the decomposed sub-bands. As a result, sub-band spectral analysis severely suffers from the mutual overlap of adjacent sub-band spectra. Second, we propose a filtering strategy to eliminate the filtering error and recover the suppression ability. This strategy intuitively takes the place of the artificial zeros with some extrapolated samples. Toward the problem of data extrapolation, many algorithms have been proposed based on various theories, such as linear prediction [16], Gerchberg–Papoulis [17], Slepian series [18], linear canonical transform [19] and sparse representation [20]. To establish an efficient method for the extrapolation in context and to evaluate the effectiveness of the proposed strategy, we preliminarily develop a linearly-optimal extrapolation based on the classical AR model identification and the Kalman prediction [21–23]. Third, we derive the formulas to estimate the residual filtering error and adapt two common information criteria with adaptive penalty terms for AR order determination. Moreover, equiripple FIR filters are applied as analysis filters in coordination with the proposed filtering, because of their advantageous features [11]. Finally, the entire algorithm and the computational complexity are summarized. Some details, such as the sub-band spectrum mosaicking procedure and parameter selection, are discussed in practice.

The remainder of the paper is organized as follows. In Section 2, the formation mechanism of spectral overlap is discussed. Based on this, a steady-state Kalman predictor-based filtering strategy is developed to suppress the overlapped spectra. In Section 3, the proposed filtering strategy is discussed for SDSE. In Section 4, experimental results with several typical algorithms for spectral analysis demonstrate the effectiveness of the proposed strategy. Finally, Section 5 concludes this paper.

2. Signal Filtering Based on AR Model Identification and Kalman Prediction

This section focuses on signal filtering. To reduce the filtering error induced by convolution filtering, we propose an extrapolation-based filtering strategy and apply a steady-state Kalman predictor for extrapolation. Two criteria with adaptive penalty terms for order determination are developed based on the estimation of the residual filtering error.

2.1. Problem Statement of Signal Filtering

FIR filters are typical linear time-invariant (LTI) systems. According to the linear system theory, the filter can be mathematically expressed as the convolution of its impulse response with the input. Suppose that $\{x_n\}$ is an input sequence and $\{h_n\}$ is the impulse response of a causal FIR filter; the filtered sequence $\{y_n\}$ can be derived as [11]:

$$y_n = h_n * x_n = \sum_{k=0}^{N_f-1} h_k x_{n-k} \quad (1)$$

where $*$ denotes the convolution operator and N_f is the filter length (*i.e.*, the length of the impulse response; the relationship between N_f and the filter order N_o can be written as $N_f = N_o + 1$). Alternatively, taking discrete-time Fourier transforms (DTFT), we can represent Equation (1) in the frequency domain as:

$$Y(e^{j\omega}) = H(e^{j\omega}) X(e^{j\omega}) \tag{2}$$

In addition, the filtered sequence length L , the input sequence length N and the filter length N_f satisfy:

$$L = N + N_f - 1 = N + N_o \tag{3}$$

Theoretically, given a large enough stop-band attenuation, spectral overlap can be thoroughly suppressed. Moreover, the spectral estimation error in sub-bands can be neglected, as long as the width of the transition band and the ripple of the passband are sufficiently small. Nonetheless, the pursuit of excellent filtering performance substantially increases both the filter order and the length of the filtered sequence (refer to Equation (3)). Such a high order is more likely to create error in part or even all of the filtered samples. This result is contrary to our original objective, and the resultant filtering quality is undesirable.

From the perspective of a discrete-time system, the output sequence of the convolution operation is equivalent to the zero-state response of the filter system, because the initial state of every delay cell is zero prior to the excitation of the input sequence. We take the example of the direct-type FIR system [24]. The value of the output sample at any time depends on all or part of the input samples and the system state at that time. The first $N_f - 1$ output samples suffer from biases, because a part of the delay cells do not yet become input-driven states; analogously, the last $N_f - 1$ output samples are invalid, because a part of the delay cells restore the initial zero-states. Thus, the length of the valid part of the output sequence, defined as L_v , satisfies:

$$L_v = L - 2(N_f - 1) = N - N_f + 1 = N - N_o \tag{4}$$

Actually, if we rewrite Equation (1) in the following matrix form:

$$\underbrace{\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_{L-1} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} x_0 & 0 & \cdots & 0 \\ x_1 & x_0 & \ddots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-1} & x_{N-2} & \ddots & x_0 \\ 0 & x_{N-1} & & x_1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{N-1} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{N_o} \end{bmatrix}}_{\mathbf{h}} \tag{5}$$

then we can find that the matrix \mathbf{X} possesses many zero elements, which probably makes the outputs $y_0, y_1, \dots, y_{N_o-1}; y_{L-N_o}, y_{L-N_o+1}, \dots, y_{L-1}$ not ideal. For example, $y_0 = x_0 h_0$, while the ideal output should be $\tilde{y}_0 = x_0 h_0 + x_{-1} h_1 + x_{-2} h_2 + \cdots + x_{-N_o} h_{N_o}$. This means that the unknown samples $x_{-N_o}, x_{-N_o+1}, \dots, x_{-1}$ are assumed to be zeros. The filtering error of y_0 is $y_0 - \tilde{y}_0 = -(x_{-1} h_1 + x_{-2} h_2 + \cdots + x_{-N_o} h_{N_o})$. Likewise, the outputs $y_1, y_2, \dots, y_{N_o-1}; y_{L-N_o}, y_{L-N_o+1}, \dots, y_{L-1}$

all suffer from errors under the zero assumption. Thus, we can conclude that the meaningless zeros are the error sources of the filtering.

Referring to Equation (4), we note that, if the filter order is not less than the length of the input sequence, the output samples are all invalid. Thus, improving filtering performance by means of unlimited increasing filter order is meaningless.

In the next subsection, we will identify an efficient way to resolve this problem.

2.2. Filtering Procedure Based on Signal Extrapolation

The desired output of the filtering process should have two characteristics:

- The original and filtered sequences should be of equal length;
- During the filtering process, the states of the delay cells in the filter system should always maintain input-driven states, *i.e.*, there are no artificial zeros, but authentic samples in \mathbf{X} .

As shown in Equation (5), the convolution filtering assumes the unknown samples $x_{-N_o}, x_{-N_o+1}, \dots, x_{-1}; x_N, x_{N+1}, \dots, x_{N+N_o-1}$ to be zeros, which leads to the filtering error. Thus, an intuitive thought is to extrapolate the sequence $\{x_n\}_{(n=0)}^{(N-1)}$ along two sides to provide a series of estimates for the unknown samples. Taking place of the zeros in the matrix \mathbf{X} with these estimates can mitigate the filtering error. The input sequence is extrapolated along both sides, yielding two extrapolated sequences, called Part A and Part B (see Figure 1). Suppose that L_A and L_B are the lengths of Part A and Part B, respectively; then, those $L_A + L_B$ extrapolated samples are used to replace zeros in \mathbf{X} . According to Equation (3), the length of the associated output sequence is $L_A + L_B + N + N_o$. From Equation (4), the effective length of the output can be given by $L_A + L_B + N - N_o$. To satisfy the requirement that the original and filtered sequence are of equal length, the extrapolated length can be derived as:

$$L_A + L_B + N - N_o = N \Rightarrow L_A + L_B = N_o \quad (6)$$

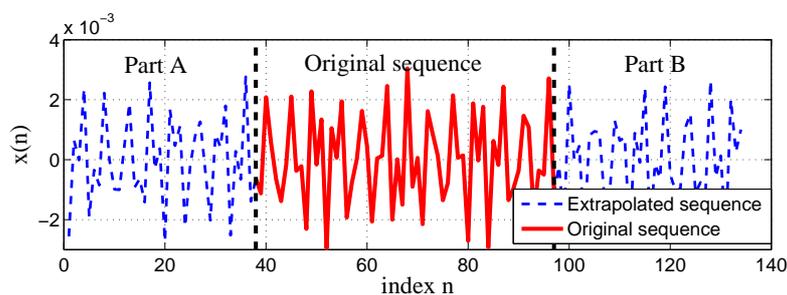


Figure 1. Original sequence and its extrapolated sequence.

Now, we can conclude that the extrapolated length should be equal to the filter order. We define L_G as the constant group delay of the filter. Between time N_o and time $N + N_o$, the output samples are valid. The output sample at time $N_o + n$ ($n = 1, 2, \dots, N$) corresponds to the input sample at time

$N_o - L_G + n$ ($n = 1, 2, \dots, N$), because of the group delay. Consequently, the input sample before time $N_o - L_G$ is merely used as a training sequence of the system state. Thus, we can obtain the relationships:

$$\begin{cases} L_A = N_o - L_G \\ L_B = L_G \end{cases} \tag{7}$$

Let \hat{x}_n and \hat{y}_n be the extrapolated sequence and associated filtered result, respectively. Then, they satisfy:

$$\begin{cases} \hat{x}_n : L_G - N_o \leq n \leq N + L_G - 1 \\ h_n : 0 \leq n \leq N_o \\ \hat{y}_n : L_G \leq n \leq L_G + N - 1 \end{cases} \tag{8}$$

$$\hat{x}_n = x_n \quad (0 \leq n \leq N - 1) \tag{9}$$

The filtering process can be rewritten in matrix form as:

$$\hat{\mathbf{y}} = \hat{\mathbf{X}}\mathbf{h} \tag{10}$$

where:

$$\hat{\mathbf{y}} = [\hat{y}_{L_G}, \hat{y}_{L_G+1}, \dots, \hat{y}_{L_G+N-1}]_{(N \times 1)}^T \tag{11}$$

and:

$$\hat{\mathbf{X}} = \begin{bmatrix} \hat{x}_{L_G} & \hat{x}_{L_G-1} & \cdots & \hat{x}_{L_G-N_o} \\ \hat{x}_{L_G+1} & \hat{x}_{L_G} & \cdots & \hat{x}_{L_G-N_o+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}_{L_G+N-1} & \hat{x}_{L_G+N-2} & \cdots & \hat{x}_{L_G+N-N_o-1} \end{bmatrix}_{(N \times N_f)} \tag{12}$$

2.3. Signal Extrapolation Based on AR Identification and Kalman Prediction

According to the linear prediction theory [25], the AR model is an all-pole model, whose output variable only linearly depends on its own previous values, that is,

$$\begin{cases} \Phi(q^{-1})x_n = \varepsilon_n \\ \Phi(q^{-1}) = \sum_{l=0}^p \phi_l q^{-l} \end{cases}, n = 0, 1, \dots, N - 1 \tag{13}$$

where q^{-1} denotes the unit delay, p is the model order, $\phi_0, \phi_1, \dots, \phi_p$ denote the coefficients of the model and $\phi_0 = 1$. The sequence $\{\varepsilon_n\}_{n=-\infty}^{\infty}$ is a white noise process, which satisfies:

$$\begin{cases} E(\varepsilon_n) = 0, & \forall n \\ E(\varepsilon_n^2) = \sigma^2, & \forall n \\ E(\varepsilon_n \varepsilon_{n'}) = 0, & n \neq n' \end{cases} \tag{14}$$

where $E(\cdot)$ denotes the expectation operator.

We choose the forward-backward approach [1] as the coefficient estimator for AR model, for its precision and robustness. Both criteria, including the Akaike information criterion (AIC) and Bayesian information criterion (BIC) [26] can be applied to determine the model order; whereas both criteria

sometimes suffer from overfitting. An alternative method of order determination will be discussed in Section 2.4.

A linearly-optimal prediction for AR sequences is derived in [21–23] under the minimum mean square error (MMSE) criterion. However, the prediction formula involves a polynomial long division and a coefficient polynomial recursion [23], making the calculation of the prediction somewhat inconvenient. Alternatively, the following steady-state Kalman predictor [27] provides an equivalent prediction with the MMSE predictor, while offering a simpler formula to facilitate the computation.

The AR model is regarded as a dynamic system. A specific state-space representation for a univariate AR(p) process can be written as [25]:

$$\begin{cases} \xi_{n+1} = \mathbf{F}\xi_n + \mathbf{\Gamma}\varepsilon_n \\ x_n = \mathbf{H}\xi_n + \varepsilon_n \end{cases} \quad (15)$$

where:

$$\mathbf{F} = \begin{bmatrix} -\phi_1 & 1 & 0 & \cdots & 0 \\ -\phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ -\phi_{p-1} & 0 & 0 & \cdots & 1 \\ -\phi_p & 0 & 0 & \cdots & 0 \end{bmatrix}_{(p \times p)} \quad (16)$$

$$\mathbf{\Gamma} = \begin{bmatrix} -\phi_1 & -\phi_2 & \cdots & -\phi_{p-1} & -\phi_p \end{bmatrix}_{(p \times 1)}^T \quad (17)$$

and:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}_{(1 \times p)} \quad (18)$$

The coefficient polynomials of x_n and ε_n are $\Phi(q^{-1})$ and one, respectively. Since they are relative prime polynomials (or coprime), *i.e.*, the transfer function is irreducible, the system of the AR model is a joint controllable and observable discrete linear stochastic system [28]. Thus, there exists a steady-state Kalman predictor:

$$\begin{cases} \hat{\xi}_{n+1|n} = F\hat{\xi}_{n|n-1} + \mathbf{K}e_n \\ x_n = \mathbf{H}\hat{\xi}_{n|n-1} + e_n \end{cases} \quad (19)$$

Since both ε_n and e_n are the innovation processes of x_n , they are equal [27]:

$$e_n = \varepsilon_n \quad (20)$$

By comparing Equation (15) with Equation (19), we have:

$$\begin{cases} \xi_n = \hat{\xi}_{n|n-1} \\ \mathbf{K} = \mathbf{\Gamma} \end{cases} \quad (21)$$

Therefore, the one-step steady-state Kalman predictor can be derived as [28]:

$$\begin{aligned} \hat{x}_{n+1|n} &= H\hat{\xi}_{n+1|n} = x_{n+1} - \varepsilon_{n+1} \\ &= x_{n+1} - \Phi(q^{-1})x_{n+1} = -\sum_{l=1}^p \phi_l x_{n-l+1} \end{aligned} \quad (22)$$

Similarly, the k -step steady-state Kalman predictor can be presented as:

$$\begin{cases} \hat{\xi}_{n+k|n} = F\hat{\xi}_{n+k-1|n-1} + F^{k-1}\mathbf{\Gamma}\varepsilon_n \\ \hat{x}_{n+k|n} = \mathbf{H}\hat{\xi}_{n+k|n} \end{cases} \quad (23)$$

Here, define:

$$\begin{cases} \mathbf{F}^{k-1}\mathbf{\Gamma} = [g_{k_0}, g_{k_1}, \dots, g_{k_{p-1}}]^T \\ G_k(q^{-1}) = g_{k_0} + g_{k_1}q^{-1} + \dots + g_{k_{p-1}}q^{-(p-1)} \end{cases} \quad (24)$$

Then, we can obtain:

$$\Phi(q^{-1})\hat{x}_{n+k|n} = G_k(q^{-1})\varepsilon_n \quad (25)$$

and that is,

$$\hat{x}_{n+k|n} = G_k(q^{-1})x_n = \sum_{l=0}^{p-1} g_{k_l}x_{n-l} \quad (26)$$

Analogously, the k -step backward extrapolation formula can be given by:

$$\hat{x}_{n-k|n} = \sum_{l=0}^{p-1} g_{k_l}^*x_{n+l} \quad (27)$$

where the superscript “*” denotes the complex conjugate operator. To guarantee reasonable and effective extrapolations, the step-size k should satisfy:

$$\begin{cases} L_G - N_o + k \geq 0 \\ N + L_G - 1 - k \leq N - 1 \end{cases} \Rightarrow L_G \leq k \leq N_o - L_G \quad (28)$$

In order to evaluate the residual filtering error of the proposed filtering strategy, we derive the mean square error (MSE) in Appendix A1.

2.4. Adaptive Information Criteria for AR Order Determination

Given the impulse response of an analysis filter and AR coefficients, we can directly calculate MSE by Equations (A2) and (A10). The precision of AR coefficient estimation is concerned with AR order. Consequently, the filtering error at different AR orders can be evaluated with the preceding formulas; conversely, the calculation of MSE can be used for order determination.

AIC and BIC are two common information criteria, whose purpose is to find a model with sufficient goodness of fit and a minimum number of free parameters. In terms of the maximum likelihood estimate $\hat{\sigma}_p^2$, we can denote AIC and BIC as [26]:

$$\min_p \text{AIC} = \log(\hat{\sigma}_p^2) + \frac{2(p+1)}{N} \quad (29)$$

$$\min_p \text{BIC} = \log(\hat{\sigma}_p^2) + \frac{p \log N}{N} \quad (30)$$

As explained in [29], due to the lack of samples, both criteria encounter the risk of overfitting, where the selected order will be larger than the truth order. In particular, AIC has the nonzero overfitting probability as the sample number tends to infinity. Theoretically, both criteria consist of two terms: the

first term involves MSE, and it decreases with the increment of the order p ; the other term is a penalty that is an increasing function of p . The preferred model order is the one with the lowest AIC or BIC value. As shown in Figure 2a, the objective function curve $\widetilde{S_1P_1E_1}$ reaches its minimum value at the point P_1 , which gives the correct order p . However, sometimes, both criteria may fail to determine available orders, and those failures are often related to inadequate penalties. Figure 2b illustrates a representative case. Since the change of the objective function instantly slows down as the order exceeds p , the point P_2 is the preferred point for order determination. However, the penalty strength is insufficient, so that the objective function is still falling after P_2 . To handle this situation, we propose an adaptive mechanism to adaptively adjust the penalty strength. A geometric interpretation is depicted in Figure 2b. We assume that the order interval for computation consists of the correct order. Then, the ray $\overrightarrow{S_2E_2}$ forms the X_2 axis, while the ray $\overrightarrow{O_2Y_2}$ forms the Y_2 axis perpendicular to the ray $\overrightarrow{S_2E_2}$ throughout the intersection O_2 of the ray $\overrightarrow{S_2E_2}$ and the objective function axis. Under the new coordination system $X_2O_2Y_2$, the minimum point P_2 of the curve $\widetilde{S_2P_2E_2}$ can help to determine the correct order. Meanwhile, this modification has no impact on the case that the criterion works well (see Figure 2a).

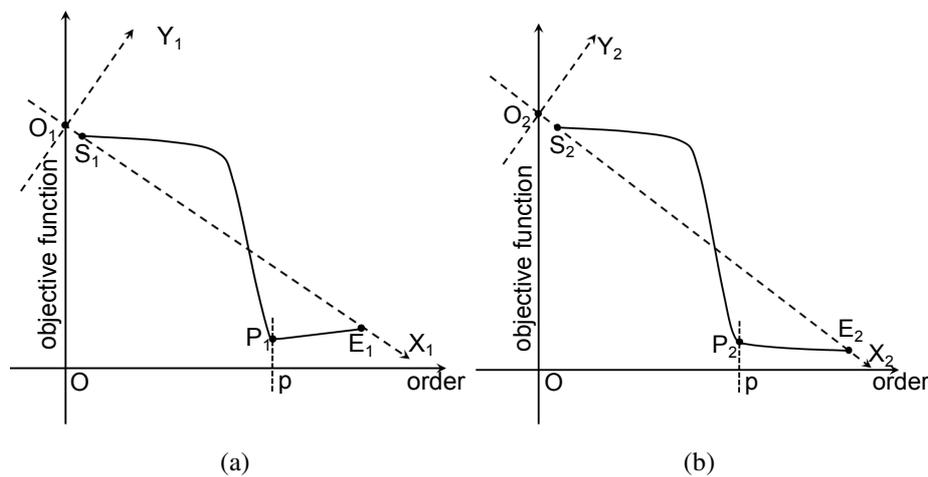


Figure 2. Geometric interpretation for adaptive Akaike information criterion (AAIC) and adaptive Bayesian information criterion (ABIC): the solid curves $\widetilde{S_1P_1E_1}$ and $\widetilde{S_2P_2E_2}$ draw objective function values for AIC or BIC. The preferred orders are located at the point P_1 and P_2 , respectively. (a) The case that the criterion (AIC or BIC) successfully determines the correct order, and (b) the case that the criterion fails due to the inadequate penalty strength. Under the new coordination system $X_2O_2Y_2$, the point P_2 becomes the minimum point of the curve, and the correct order is retrieved.

Therefore, the adaptive AIC (AAIC) based on MSE of the residual filtering error can be derived as:

$$\begin{cases} \min_p \text{AAIC} = \log(\hat{\sigma}_p^2) + \frac{2}{N}(p^\alpha + 1) \\ \text{s.t. AAIC}(p_s) = \text{AAIC}(p_e) \end{cases} \quad (31)$$

where p_s and p_e denote the start point and the end point of the computing order interval, respectively. If $p_s = 1$, the adaptive parameter α can be given by:

$$\alpha = \log_{p_e} \left[\frac{N}{2} \log \left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_{p_e}^2} \right) + 1 \right] \quad (32)$$

Analogously, the adaptive BIC (ABIC) can be represented as:

$$\begin{cases} \min_p \text{ABIC} = \log(\hat{\sigma}_p^2) + p^\beta \frac{\log N}{N} \\ \beta = \log_{p_e} \left[\frac{N}{\log N} \log \left(\frac{\hat{\sigma}_p^2}{\hat{\sigma}_{p_e}^2} \right) + 1 \right] \end{cases} \quad (33)$$

3. Implementation of SDSE

In this section, we discuss the implementation details of SDSE based on the proposed filtering strategy. In particular, equiripple FIR filters are used as the analysis filters for their advantageous features. To suppress spectral overlap and improve spectral precision in practice, we introduce a mosaicking operation for sub-band spectra and discuss the compensation of the residual error of the composite spectrum. After that, we summarize the entire algorithm and analyze the computational complexity.

3.1. Properties and Design of Equiripple FIR Filters

Besides the advantages of FIR filters, *i.e.*, exact linear phase response and inherent stabilization, equiripple FIR filters have an explicitly specified transition width and passband/stop-band ripples (see Figure 3). As analysis filters, equiripple FIR filters can bring some important benefits, such as stop-band attenuation with a fixed maximum, the explicitly specified width of the invalid part of the sub-band spectrum (which corresponds to the transition-band spectrum) and a limited maximum deviation of the valid part of the sub-band spectrum (which corresponds to the passband spectrum). As shown in Figure 3, the specifications of a typical equiripple FIR filter consist of the passband edge ω_p , stop-band edge ω_s and maximum error in passband and stop-band δ_p , δ_s , respectively. The approximate relationship between the optimal filter length and other parameters developed by Kaiser [11] is:

$$N_f \approx \frac{-20 \log_{10}(\sqrt{\delta_p \delta_s}) - 13}{14.6 \Delta f} + 1 \quad (34)$$

where Δf denotes the width of the transition-band,

$$\Delta f = \frac{\omega_s - \omega_p}{2\pi} \quad (35)$$

The maximum passband variation and the minimum stop-band attenuation in decibels are defined as:

$$A_p = 20 \log_{10} \left(\frac{1 + \delta_p}{1 - \delta_p} \right) \text{ dB} \quad (36)$$

and:

$$A_s = -20 \log_{10}(\delta_s) \text{ dB} \quad (37)$$

respectively.

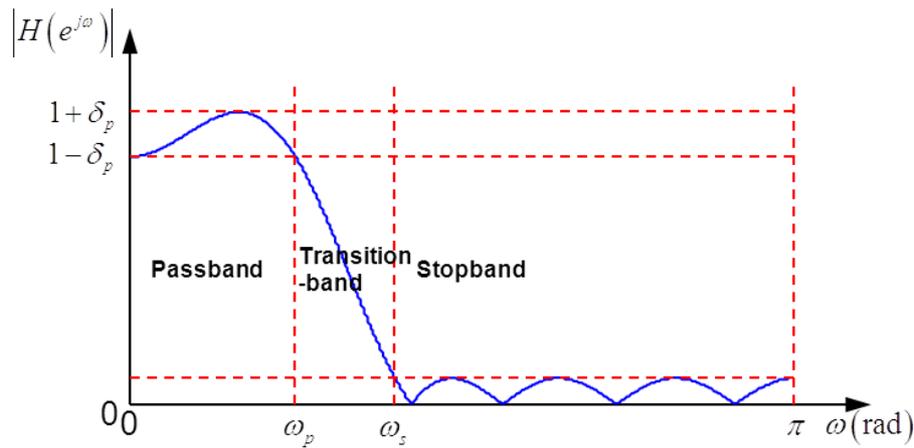


Figure 3. Magnitude response and design parameters of an equiripple low-pass FIR filter.

When the specification of a filter is explicitly specified, we can complete the design with the Parks–McClellan (PM) algorithm [30], since it is optimal with respect to the Chebyshev norm and results in about 5 dB more attenuation than the windowed design algorithm [11].

3.2. Practical Consideration of Equiripple FIR Filters

Firstly, the equiripple low-pass FIR filter is combined with a preprocessing step—complex frequency modulation—to form a passband filter for sub-band decomposition (see Figure 4).

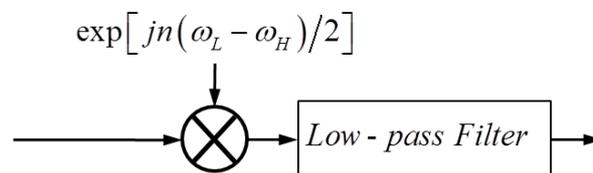


Figure 4. Block diagram of the analysis filter.

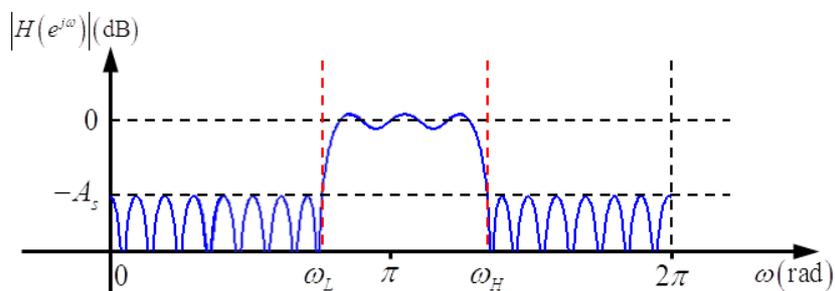


Figure 5. Magnitude response of the analysis filter.

The magnitude response of the analysis filter is shown in Figure 5, where ω_H and ω_L denote the high and the low edge of the stop-band, respectively. They satisfy:

$$\omega_H - \omega_L = 2\omega_s \tag{38}$$

As long as A_s is large enough and the downsample rate M meets the condition:

$$M \geq \frac{2\pi}{\omega_H - \omega_L} \Rightarrow M \geq \frac{\pi}{\omega_s} \quad (39)$$

frequency aliasing can be practically suppressed.

Secondly, due to the existence of the transition-band of each analysis filter, each sub-band spectrum contains two invalid parts. The spectral estimations of these invalid parts lead to errors. Consequently, according to [31], when mosaicking these sub-band spectral estimations into full-band, we should omit these invalid parts of spectral estimations. This procedure is illustrated in Figure 6. Thus, the composite full-band spectral estimation is practically immune to the spectral overlap.

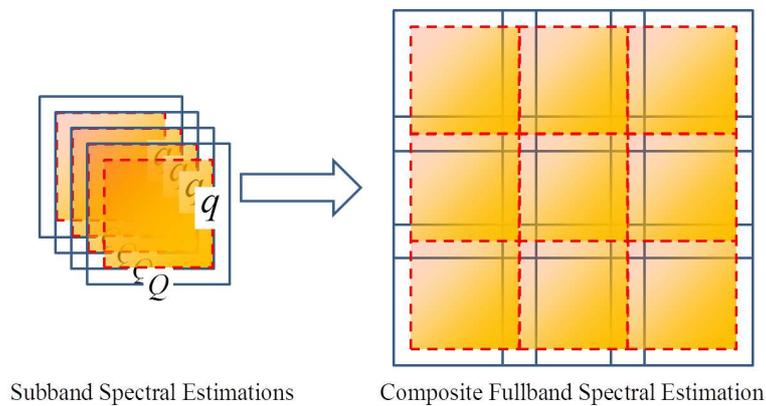


Figure 6. Illustration of mosaicking the sub-band spectral estimations into a composite spectrum. The sub-band spectral estimations are overlapped, while the composite full-band spectrum is without overlap (the boxes with solid lines cover the spectral estimation of the sub-bands; the boxes with dashed lines cover the valid spectral estimation).

Thirdly, due to the existence of passband ripples in equiripple FIR filters, there theoretically remains a small error in sub-band spectral estimations. Generally, by adjusting the maximum passband variation, we can limit the error to an allowable range. More precise spectral estimation necessitates compensation for the residual error. Since the ripple curve for any given equiripple FIR filter can be accurately measured, the compensation can be performed by weighting sub-band spectral estimations with the measured ripple curve.

Finally, we focus on selecting appropriate filter parameters in SDSE, which can improve the performance and reduce computational cost. The filter order should at least meet:

$$N \geq \max(L_G, N_o - L_G) \quad (40)$$

The maximum stop-band attenuation should exceed the dynamic range of the signal to be analyzed. Once the aforementioned conditions are satisfied, the shortest transition-width can be chosen by Equation (34). Moreover, specific requirement will help to set the maximum passband variation.

3.3. Computational Complexity of SDSE

Algorithm 1 Non-overlapping sub-band spectral estimation with the steady-state Kalman predictor-based filtering strategy.

Input: The sequence $\{x_n\}_{n=0}^{N-1}$.

Parameters: The maximum passband variation A_p and the minimum stop-band attenuation A_s in decibels; the sub-band number M .

Filter Design:

Set the stop-band edge ω_s by Equation (39);

Set the passband edge ω_p by Equations (34), (35), (36), (37) and (40);

Design the equiripple FIR filter by Parks–McClellan (PM) algorithm [30], and then, compute the impulse response $\{h_n\}_{n=0}^{N_o}$ and the group delay L_G .

AR Identification and Order Selection:

for $p_i = p_s$ to p_e (usually set $p_s = 1, p_e \leq N/2 - 1$) **do**

Estimate coefficients $\{\phi_i\}_{i=1}^{p_i}$ of AR model by the forward and backward estimator, with $\mathcal{O}(Np_i^2)$ flops;

Estimate the MSE $\hat{\sigma}_{p_i}^2$ by (A10) and (A11), with $\mathcal{O}\left(\frac{N_o^4}{48} + \frac{5}{24}N_o^3 + \frac{N_o^2}{6}\right)$ flops;

end for

Select an order p by Equation (31) or Equation (33), with $\mathcal{O}(N)$ flops.

Sequence Extrapolation:

Set the step-size k by Equation (28);

Calculate $\{g_{k_l}\}_{l=0}^{p-1}$ by Equations (16), (17) and (24), with $\mathcal{O}\left((2p)^{k-1}\right)$ flops;

Implement forward and backward extrapolations by Equations (26) and (27), and obtain $\{\hat{x}_n\}_{n=L_G-N_o}^{L_G+N-1}$, with $\mathcal{O}(N_o p)$ flops.

Sub-Band Spectral Estimation:

Set a rational factor $M_0 = \left\lceil \left\lceil \frac{\pi}{\omega_s} \right\rceil \right\rceil$, where $\lceil \cdot \rceil$ denotes a rational approximation;

for $i = 1$ to M **do**

Compute ω_H and ω_L by Equation (38) and $\omega_H + \omega_L = (2i - 1)\pi/M$;

Perform pre-modulation and filtering for $\{\hat{x}_n\}_{n=L_G-N_o}^{L_G+N-1}$ by Figure 4 and Equation (10), and the computational complexity is in $\mathcal{O}(2(N + N_o) \log(N + N_o))$ flops;

Decimate the sequence $\{\hat{x}_n\}_{n=L_G-N_o}^{L_G+N-1}$ by a factor of M_0 , and obtain the sub-band sequence $\{\hat{x}_n^{(i)}\}_{n=0}^{\lceil \frac{N}{M_0} \rceil - 1}$, where $\lceil \cdot \rceil$ denotes the ceiling function;

Perform spectral analysis for the sub-band sequence $\{\hat{x}_n^{(i)}\}_{n=0}^{\lceil \frac{N}{M_0} \rceil - 1}$, and denote the length of the sub-band spectrum as L_s ;

Compute the length of overlapped spectrum by $\left\lceil \left(\frac{1}{M_0} - \frac{\omega_p}{\pi} \right) \frac{L_s}{2} \right\rceil$ and omit the overlapped parts at both the left and the right side of the sub-band spectrum.

end for

Mosaic the residual sub-band spectrums into an entire spectrum.

Output: The entire spectrum.

As shown in Algorithm 1, we summarize SDSE with the proposed filtering strategy and give the computational complexity of the major steps. First, the proposed strategy can greatly reduce the computational burden. We take the commonly-used amplitude and phase estimation (APES) [32] algorithm as an example. The full-band APES needs $\mathcal{O}(N^2 \log N)$ flops [33], while the computation requirement is decreased to $\mathcal{O}\left(M\left(\left\lceil\frac{N}{M_0}\right\rceil\right)^2 \log\left(\left\lceil\frac{N}{M_0}\right\rceil\right)\right)$ flops by SDSE with the proposed strategy. Second, except the sub-band spectral estimation, the main computation requirement is induced by the AR identification and the order selection. The computational complexity of this step is generally much lower than that of the sub-band spectral estimation. In particular, if a proper order or a small enough order interval is preselected before the AR identification, the computation of this step can be negligible.

4. Simulations and Analysis

In this section, both the feasibility and the effectiveness of the proposed strategy are evaluated by typical numerical simulations, including FIR filtering and line spectral analysis of 1D or 2D sequences.

4.1. Filtering Analysis Using the Proposed Strategy

Suppose that the input sequence $\{x_n\}$ is a mixed complex exponential sequence:

$$\begin{cases} x_n = s_n^{(1)} + s_n^{(2)} + v_n \\ s_n^{(1)} = \exp(0.45j\pi n) \\ s_n^{(2)} = \sum_{l=0}^{25} 100 \exp\{(0.55 + 0.035l)j\pi n\} \\ n = 0, 1, \dots, 127 \end{cases} \quad (41)$$

where the measurement noise $\{v_n\}$ is a complex Gaussian process. All real parts and imaginary parts of $\{v_n\}$ are independent and identically distributed (i.i.d.) zero-mean Gaussian distributions with variance σ^2 , i.e., $\text{Re}(v_n), \text{Im}(v_n) \sim \mathcal{N}(0, \sigma^2)$. Our purpose is to non-distortedly extract the weak component $s_n^{(1)}$ from x_n or completely eliminate the strong component $s_n^{(2)}$.

The equiripple half-band low-pass FIR filter is chosen for the extraction. The specifications of the filter are:

$$A_p = 1.4295 \times 10^{-3} \text{dB}, A_s = 81.6852 \text{dB}, \Delta f = 0.08 \quad (42)$$

The length of the designed filter based on the given specifications is 119.

As shown in Figure 7a, the decreasing trend of the estimated residual error by the proposed strategy is consistent with the real error. When the order exceeds 57, the decrease of the estimated filtering error instantly slows down. Hence, the preferred order is 57. By comparison, due to the deficiency of the penalty strength, none of AIC and BIC can provide the right order; whereas, based on the adaptive penalty terms, both AAIC and ABIC get the right order 57 (see Figure 7b).

As shown in Figure 8b, the weak component $s_n^{(1)}$ is completely covered by the sidelobe of the out-of-band strong component $s_n^{(2)}$; thus, recognizing the existence of the weak component from the mixed spectrum is completely impossible. From the view of the magnitude response (see Figure 8a), the filter has the nominal ability of eliminating the interference of the out-of-band strong components for

the in-band weak component. Due to the existence of the convolution filtering error, we still cannot find out the weak component from the convolution spectrum, as shown in Figure 8b. By contrast, once the samples contaminated by the filtering error are omitted by Equation (4) from the filtered sequence, the weak component reappears in the spectrum of the remaining samples (refer to the truncated spectrum in Figure 8b). However, the truncated spectrum has a much wider main lobe than the original spectrum, which means the spectral resolution suffers from a severe decrease. In order to simultaneously maintain the resolution and filter out the interference, we apply the proposed filtering strategy to handle the case. As shown in Figure 8c, based on the proposed strategy, the restored spectrum for the noiseless sequence closely coincides with the truth weak spectrum in shape, especially retaining the spectral resolution. In addition, even when the signal-to-noise (SNR) of $s_n^{(1)}$ is low to -3 dB (when $\sigma^2 = 1$), the recovery is still effective (see the magnified details of Figure 8c).

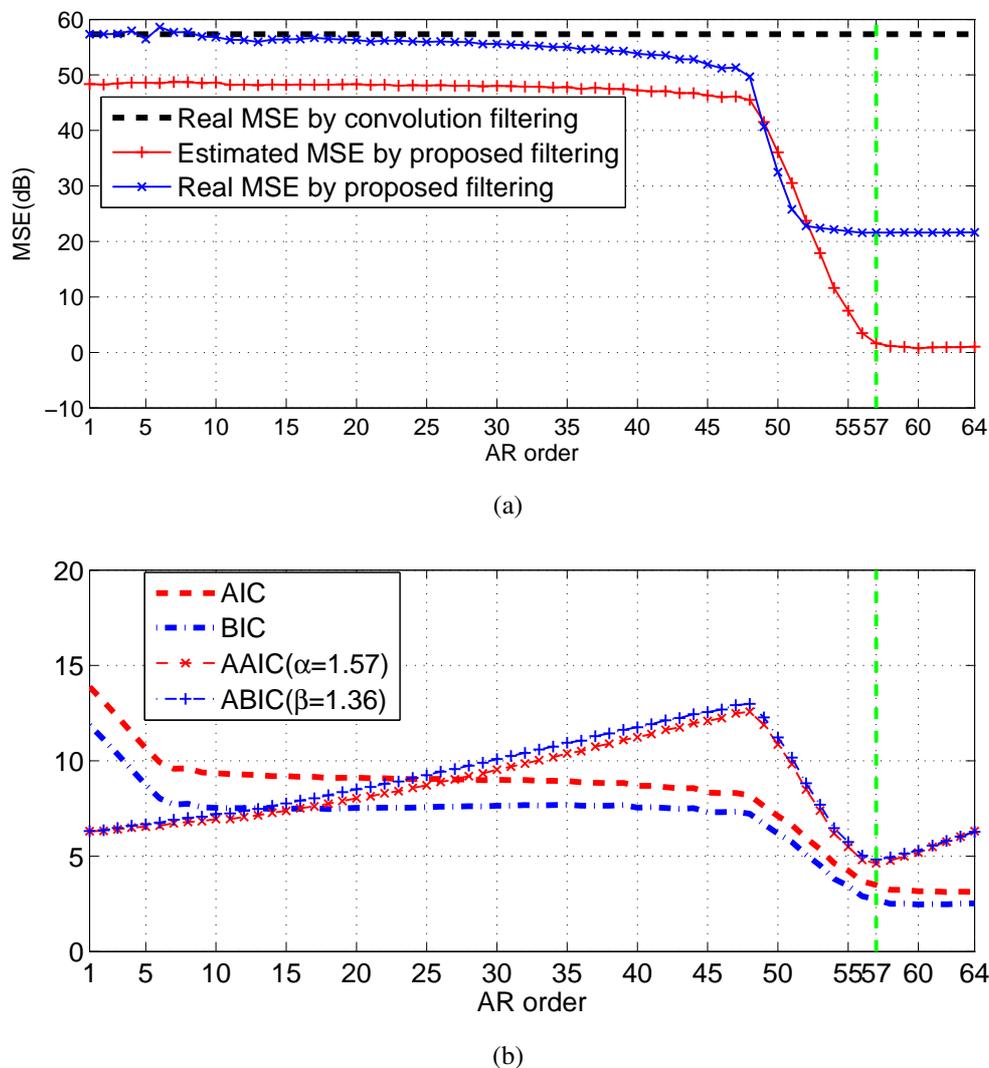


Figure 7. Mean square error of filtering and order selection: (a) quantitative comparison of the filtering error by convolution filtering and the proposed filtering; (b) comparison of the information criteria, including AIC, BIC, AAIC and ABIC.

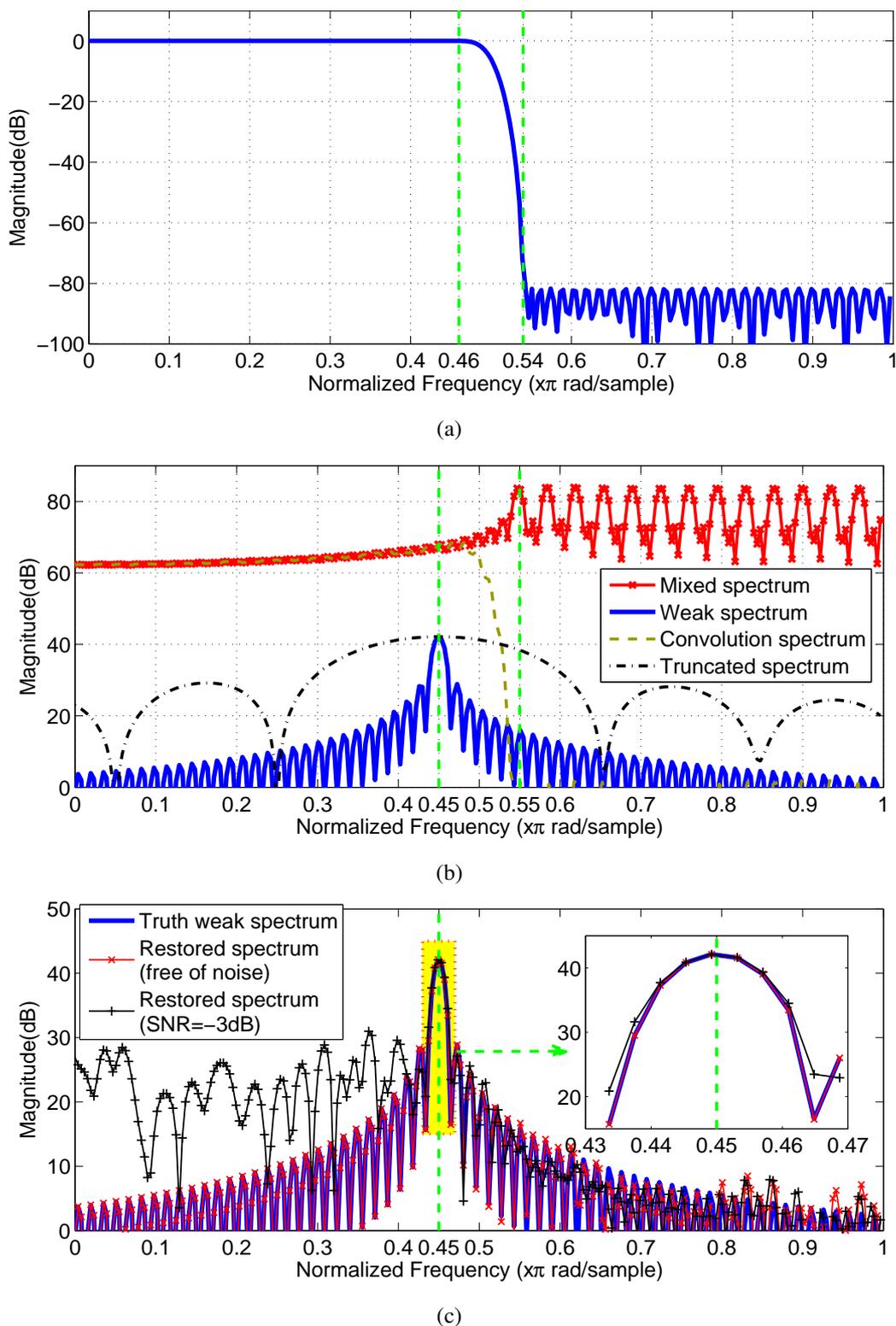


Figure 8. Quantitative comparison of filtering results: (a) magnitude response of the designed equiripple finite impulse response (FIR) half-band filter; (b) Fourier spectra of the mixed sequence, the truth weak component, the convolved sequence and the truncated sequence with 10 samples; (c) Fourier spectra of the truth weak component; the restored results by the proposed filtering strategy when the mixed sequence is free of noise or contaminated by noise (SNR = -3 dB).

4.2. Line Spectral Analysis Using 1-D Signals

A complex exponential model can be mathematically represented as:

$$\left\{ \begin{array}{l} x_n = s_n^{(1)} + s_n^{(2)} + v_n \\ s_n^{(1)} = \sum_{k=1}^5 \alpha_k \exp [j (\omega_k n + \phi_k)] \\ s_n^{(2)} = \sum_{i=0}^{16} 100 \left\{ \exp \left[j \left(\omega_i^{(+)} n + \phi_i^{(+)} \right) \right] + \exp \left[j \left(\omega_i^{(-)} n + \phi_i^{(-)} \right) \right] \right\} \end{array} \right. \quad (43)$$

where:

$$\begin{aligned} \alpha_1 = \alpha_3 = \alpha_5 = 5, \alpha_2 = \alpha_4 = 1 \\ \omega_1 = -0.075\pi, \omega_2 = -0.03125\pi, \omega_3 = 0.0125\pi, \omega_4 = 0.05625\pi \\ \omega_5 = 0.1\pi, \omega_i^{(+)} = (0.15 + 0.05i)\pi, \omega_i^{(-)} = -(0.15 + 0.05i)\pi \end{aligned}$$

and:

$$n = 0, 1, \dots, N - 1; N = 128$$

$\{v_n\}$ is a real-value sequence of i.i.d. zero-mean Gaussian random variables with variance $\sigma^2 = 1.5811$, i.e., $v_n \sim \mathcal{N}(0, \sigma^2)$. ϕ_k, ϕ_i^+ and ϕ_i^- are i.i.d. uniform random variables on the interval from zero to 2π , i.e., $\phi_k, \phi_i^{(+)}, \phi_i^{(-)} \sim \mathcal{U}[0, 2\pi)$.

In this case, we can get each component's SNR of $s_n^{(1)}$:

$$\text{SNR}_1 = \text{SNR}_3 = \text{SNR}_5 = 5\text{dB}, \text{SNR}_2 = \text{SNR}_4 = -2\text{dB}$$

We decompose the mixed-signal x_n into four sub-bands using the proposed method with the filter parameter set as:

$$A_p = 0.01\text{dB}, A_s = 60\text{dB}, \Delta f = 0.05 \quad (44)$$

The sub-band, whose radian frequency is within $[-0.125\pi, +0.125\pi)$, is used for frequency estimation. Furthermore, we estimate the frequencies of complex sinusoids of $s_n^{(1)}$ that are contained in both mixed-signal x_n and the decomposed sub-band signal, via MUSIC, ESPRIT [34,35] and SELF-SVD [15] algorithms (see Table 1). As shown in Table 1, we analyze the performance based on the Monte Carlo method. Compared with ESPRIT, SELF-SVD in full-band spectral estimation suffers from obvious performance degradations or even failures. Although SELF-SVD can theoretically attenuate the out-of-band components for the in-band frequency estimations, the ability of attenuation is not always sufficient, especially when the power of the out-of-band components are much stronger than that of the in-band components or the SNR is relatively low. Instead of performing the SVD method in the entire frequency domain as ESPRIT, SELF-SVD just performs it in the frequency interval of interest. Obviously, the remaining out-of-band interferences will be treated as in-band components, so that the frequency estimation with SELF-SVD sometimes fails. In the experiment, the power ratio of the out-of-band components to the in-band components ω_2 and ω_4 is up to 10,000 times. As a result, the corresponding frequency estimation with SELF-SVD fails to work. When we eliminate the out-of-band interferences with our method, the estimation of SELF-SVD for the residual signal exhibits similar performance as ESPRIT. In addition, MSEs of MUSIC and ESPRIT indicate that the frequency estimation in the sub-band is much more accurate than that in the full-band.

Table 1. Comparison of frequency estimation in full-band and sub-bands.

Means and MSEs of Estimating Frequencies	Full-band			Sub-band		
	MUSIC	ESPRIT	SELF-SVD	MUSIC	ESPRIT	SELF-SVD
$\bar{\omega}_1 (\omega_1 = -0.07500\pi)$	-0.0750π	-0.0750π	-0.0790π	-0.0749π	-0.0749π	-0.0750π
$\bar{\omega}_2 (\omega_2 = -0.03125\pi)$	-0.0309π	-0.0309π	\sim	-0.0308π	-0.0313π	-0.0308π
$\bar{\omega}_3 (\omega_3 = 0.01250\pi)$	0.0125π	0.0125π	0.0175π	0.0123π	0.0123π	0.0126π
$\bar{\omega}_4 (\omega_4 = 0.05625\pi)$	0.0572π	0.0572π	\sim	0.0555π	0.0557π	0.0555π
$\bar{\omega}_5 (\omega_5 = 0.10000\pi)$	0.1000π	0.1000π	0.1027π	0.1000π	0.1000π	0.1000π
$\hat{\sigma}_{\bar{\omega}_1}^2 \times 10^5$	0.0072	0.0072	1.8126	0.0057	0.0058	0.0044
$\hat{\sigma}_{\bar{\omega}_2}^2 \times 10^5$	1.6462	1.6462	\sim	0.1662	0.1259	0.1816
$\hat{\sigma}_{\bar{\omega}_3}^2 \times 10^5$	0.0125	0.0125	2.8847	0.0081	0.0074	0.0060
$\hat{\sigma}_{\bar{\omega}_4}^2 \times 10^5$	3.0443	3.0443	\sim	0.1729	0.1199	0.1775
$\hat{\sigma}_{\bar{\omega}_5}^2 \times 10^5$	0.0360	0.0360	0.7379	0.0049	0.0041	0.0046

“ \sim ” denotes meaningless estimates.

(Monte Carlo analysis: 100 runs). SELF-SVD, singular value decomposition-based method in a selected frequency band.

4.3. Line Spectral Analysis Using 2D signals

Let $C_k (k = 1, 2, \dots, K)$ be a series of random integers with unique values generated from a uniform discrete distribution on $[1, 1024 \times 1024]$. We define two sets of nonnegative integers as:

$$\begin{cases} p_k = \lfloor \frac{C_k}{1024} \rfloor \\ q_k = \text{mod}(C_k, 1024) \end{cases} \quad k = 1, 2, \dots, K \tag{45}$$

where $\lfloor \cdot \rfloor$ rounds a number to the nearest integer toward zero, and $\text{mod}(\cdot)$ is the modulo operator.

The 2D signal model can be expressed as:

$$\begin{cases} x_{n_1, n_2} = s_{n_1, n_2}^{(1)} + s_{n_1, n_2}^{(2)} + v_{n_1, n_2} \\ s_{n_1, n_2}^{(1)} = \sum_{k=1}^K \exp \left[j2\pi \left(\frac{n_1 p_k}{4N_1} + \frac{n_2 q_k}{4N_2} \right) + j\phi_k \right] \\ s_{n_1, n_2}^{(2)} = \sum_{l_1=0}^{16} 1,000 \left\{ \begin{aligned} &\exp \left[j2\pi \left(\frac{118.5n_1}{N_1} + \frac{(95.5+4l_1)n_2}{N_2} \right) + j\phi_{l_1}^{(1)} \right] \\ &+ \exp \left[j2\pi \left(\frac{54.5n_1}{N_1} + \frac{(95.5+4l_1)n_2}{N_2} \right) + j\phi_{l_1}^{(2)} \right] \end{aligned} \right\} \\ + \sum_{l_2=1}^{15} 1,000 \left\{ \begin{aligned} &\exp \left[j2\pi \left(\frac{(54.5+4l_2)n_1}{N_1} + \frac{95.5n_2}{N_2} \right) + j\phi_{l_2}^{(1)} \right] \\ &+ \exp \left[j2\pi \left(\frac{(54.5+4l_2)n_1}{N_1} + \frac{159.5n_2}{N_2} \right) + j\phi_{l_2}^{(2)} \right] \end{aligned} \right\} \end{cases} \tag{46}$$

where:

$$n_1 = 0, 1, \dots, N_1 - 1; n_2 = 0, 1, \dots, N_2 - 1$$

and:

$$N_1 = N_2 = 256, K = 8, 192$$

$\{v_{n_1, n_2}\}$ is a real-value sequence following $v_{n_1, n_2} \sim \mathcal{N}(0, \sigma^2)$ with $\sigma^2 = 0.005$. $\phi_k, \phi_{l_1}^{(1)}, \phi_{l_1}^{(2)}, \phi_{l_2}^{(1)}, \phi_{l_2}^{(2)}$ are uniform random variables on the interval from zero to 2π , i.e., $\phi_k, \phi_{l_1}^{(1)}, \phi_{l_1}^{(2)}, \phi_{l_2}^{(1)}, \phi_{l_2}^{(2)} \sim \mathcal{U}[0, 2\pi)$.

The spectrum of this 2D sequence is shown in Figure 9. Since the magnitude of $s_{n_1, n_2}^{(2)}$ is 60 dB greater than that of $s_{n_1, n_2}^{(1)}$, the sidelobe of the former significantly affects the spectral estimation of the latter.

This affect is especially more severe for the components around $s_{n_1, n_2}^{(1)}$. The region inside the red pane is used to verify the performance of the proposed method.

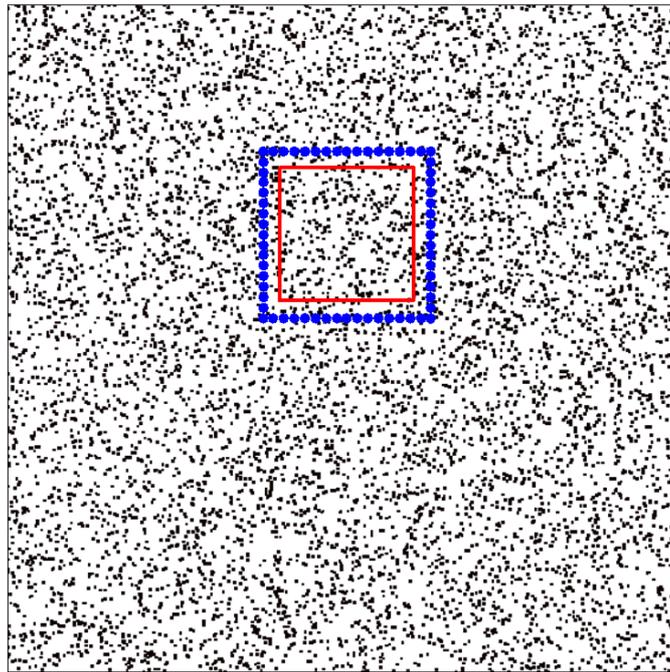


Figure 9. Actual magnitude spectrum of the 2D signal (the black dot corresponds to $s_{n_1, n_2}^{(1)}$: 0 dB; the rounded blue spot \bullet corresponds to $s_{n_1, n_2}^{(2)}$: 60 dB; the red pane covers the region to be analyzed).

The parameters of the analysis filter are selected as:

$$A_p = 0.2 \text{ dB}, A_s = 80 \text{ dB}, \Delta f = 0.05 \quad (47)$$

The comparison of Figure 10a and 10c indicates that the Fourier spectrum of $s_{n_1, n_2}^{(1)}$ is severely affected by $s_{n_1, n_2}^{(2)}$. By contrast, the result shown in Figure 10b seems to be almost exactly the same as the desired result shown in Figure 10c. This decomposition result verifies the effectiveness of the proposed method.

To further testify the performance of our method, we select the APES [32] and the iterative adaptive approach (IAA) [36,37] for spectral estimation. Since the ideal frequency domain filters suffer from energy leakage and/or frequency aliasing problems, the APES result shown in Figure 10c is somewhat blurred. By contrast, the APES result of the decomposed sub-band based on the proposed strategy (see Figure 10d) is quite similar to the actual spectrum (see Figure 10e). Theoretically, the IAA is superior to the APES. However, as shown in Figure 10g, it is even more likely than the APES to suffer from out-of-band interferences. From the view of the sub-band IAA spectrum (see Figure 10h), most of the interferences are eliminated, while the remaining filtering error still has impacts on the spectrum. Thus, the spectral estimation experiment reveals that the sub-band decomposition based on the proposed method can provide relatively ideal performance; whereas the developed method for extrapolation is imperfect, so it can affect the performance of the IAA algorithm.

In addition, a simulated single-polarized SAR image of an airplane based on the physical and optical model is processed *via* the APES. The computation time of full-band APES (refer to Figure 11a) is

26.85 h, while the time of sub-band APES (refer to Figure 11b) is just 0.84 h. Obviously, the two imaging results only have tiny differences, which are hardly recognized.

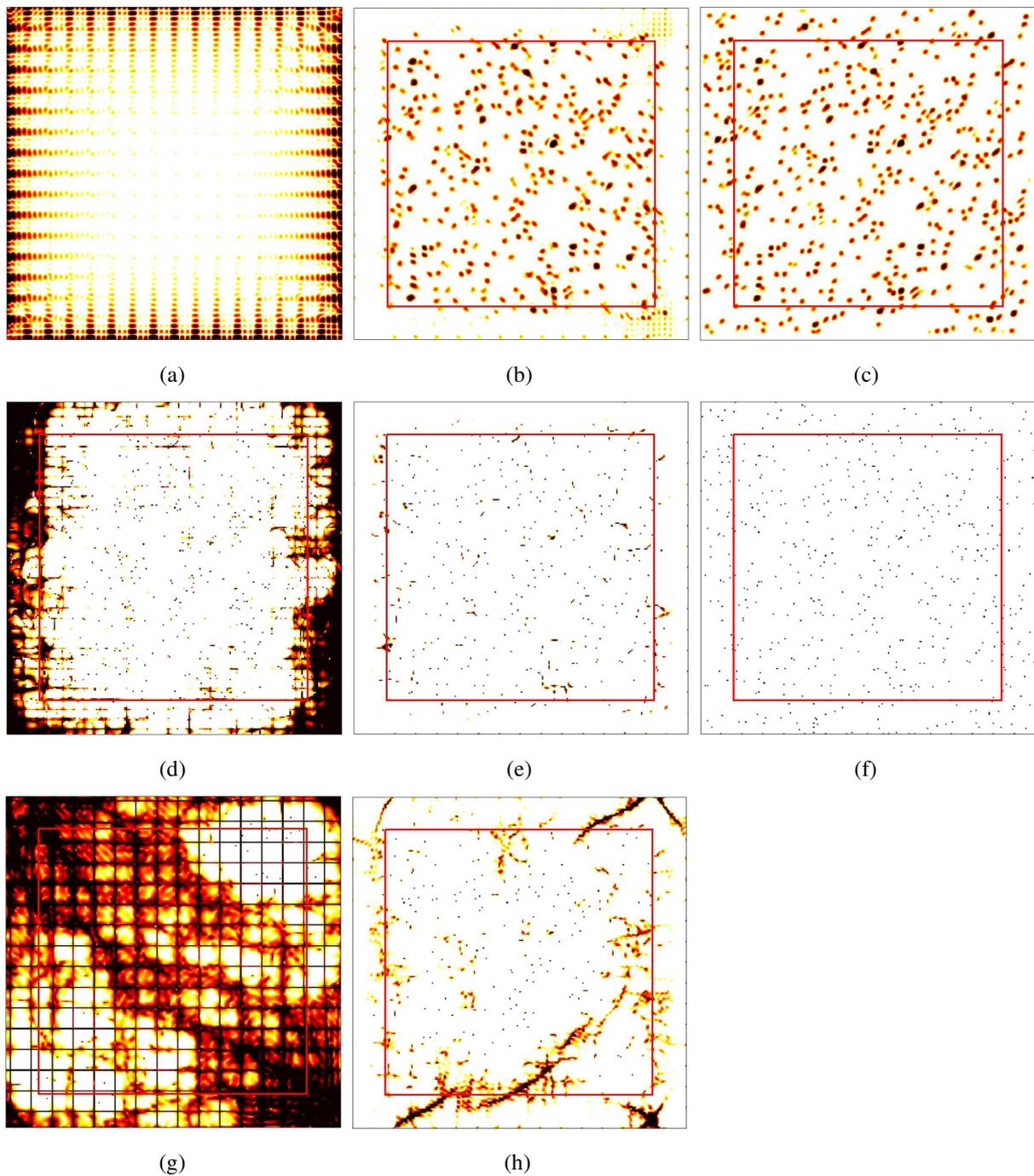


Figure 10. Sub-band decomposition and spectral estimation within the analyzed region: (a) the Fourier spectrum of x_{n_1, n_2} ; (b) the Fourier spectrum of decomposed sub-band signal based on our method; (c) the Fourier spectrum of $s_{n_1, n_2}^{(1)}$; (d) the amplitude and phase estimation (APES) result of (a) corresponding to the ideal frequency domain filters-based sub-band decomposition; (e) the APES result of (b); (f) the actual spectrum; (g) the iterative adaptive approach (IAA) result of (a); (h) the IAA result of (b).

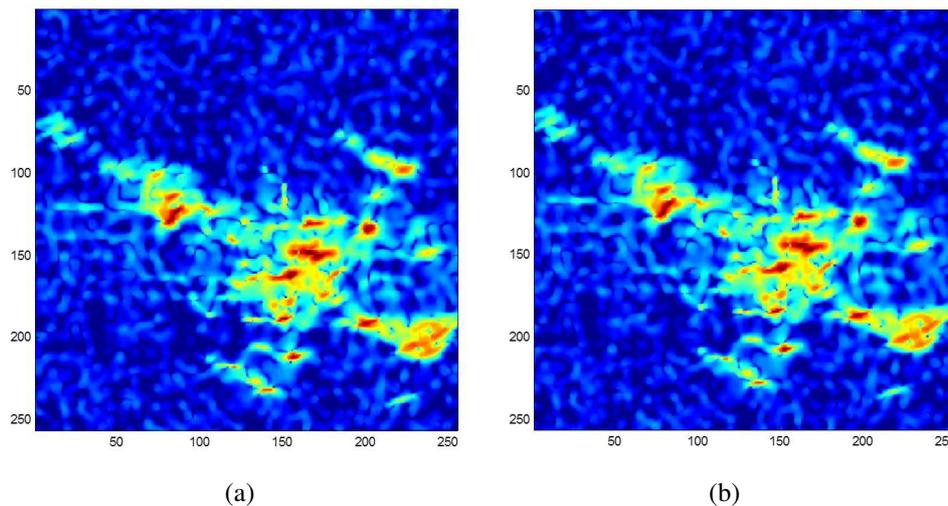


Figure 11. Comparison between full-band (a) and sub-band (b) APES images: the imaging for (a) costs 26.85 h, while for (b), it is 0.84 h.

5. Conclusion

This paper has investigated the problem of suppressing spectral overlap in sub-band spectral estimation. The spectral overlap phenomenon is originated from the non-ideal behavior of the analysis filtering, *i.e.*, the filtering error. The error formation in convolution filtering was therefore discussed, based on which an extrapolation-based filtering strategy was proposed to greatly suppress spectral overlap. Several classical theories, including AR identification, Kalman prediction and the equiripple FIR filtering technique, are integrated into the strategy for linearly-optimal extrapolation. To resolve the “overfitting” in order determination with AIC and BIC, we modified the penalty terms for both criteria. The improved criteria adaptively adjust the penalty strength and avoid “overfitting” to some extent. Both 1D and 2D complex exponential signals are utilized to validate the performance of the proposed method. Moreover, we employed SAR image formation for a single-polarized SAR data, simulated based on electromagnetic theory, to testify the efficiency of our method. Future research will focus on developing more sophisticated methods for the problem of extrapolation, with which we can avoid model order determination and further improve the extrapolation precision.

Acknowledgments

The work was in part supported by the NSFC (No. 41171317), in part supported by the key project of the NSFC (No. 61132008), in part supported by the major research plan of the NSFC (No. 61490693), in part supported by the Aeronautical Science Foundation of China (No. 20132058003) and in part supported by the Research Foundation of Tsinghua University. The authors would also like to thank the reviewers for their helpful comments.

Author Contributions

Zenghui Li contributed to the original idea, algorithm design and paper writing. Bin Xu contributed in part to data processing. Jian Yang and Jianshe Song supervised the work and helped with editing the paper.

Appendix

A1. Residual Filtering Error Analysis

A stationary AR(*p*) process can be represented as:

$$x_n = \sum_{l=0}^{\infty} \alpha_l \varepsilon_{n-l} \tag{A1}$$

where the coefficient series $\{\alpha_l\}$ meets mean square convergence and can be calculated recursively [28]:

$$\begin{cases} \alpha_l = -\phi_1 \alpha_{l-1} - \dots - \phi_p \alpha_{l-p} = - \sum_{i=1}^{\min(p,l)} \phi_i \alpha_{l-i} \\ \alpha_0=1, \alpha_l = 0 \ (l < 0) \end{cases} \tag{A2}$$

Then *k*-step prediction formula in an alternative form is:

$$\hat{x}_{n|n-k} = \sum_{l=k}^{\infty} \alpha_l \varepsilon_{n-l} \tag{A3}$$

The prediction error is $x_n - \hat{x}_{n|n-k} = \sum_{l=0}^{k-1} \alpha_l \varepsilon_{n-l}$, and this mean that the variance and covariance are:

$$E \{x_n - \hat{x}_{n|n-k}\} = 0 \tag{A4}$$

$$E \{|x_n - \hat{x}_{n|n-k}|^2\} = \sigma_{\varepsilon}^2 \sum_{l=0}^{k-1} |\alpha_l|^2 \tag{A5}$$

$$E \{(x_{n_1} - \hat{x}_{n_1|n_1-k}) (x_{n_2} - \hat{x}_{n_2|n_2-k})^*\} = \begin{cases} \sigma_{\varepsilon}^2 \sum_{(l_1, l_2) \in \Lambda} \alpha_{l_1} \alpha_{l_2}^* & |n_1 - n_2| \leq k - 1 \\ 0 & otherwise \end{cases} \tag{A6}$$

respectively, where the set is:

$$\Lambda = \{(l_1, l_2) | l_1 - l_2 = n_1 - n_2, 0 \leq l_1, l_2 \leq k - 1\}$$

We define the desired output as \tilde{y}_n and by analogy to Equation (11), we define its associated vector as:

$$\tilde{\mathbf{y}} = [\tilde{y}_{L_G}, \tilde{y}_{L_G+1}, \dots, \tilde{y}_{L_G+N-1}]_{(N \times 1)}^T \tag{A7}$$

where:

$$\tilde{\mathbf{y}} = \tilde{\mathbf{X}} \mathbf{h} \tag{A8}$$

and:

$$\tilde{\mathbf{X}} = \begin{bmatrix} x_{L_G} & x_{L_G-1} & \dots & x_{L_G-N_o} \\ x_{L_G+1} & x_{L_G} & \dots & x_{L_G-N_o+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{L_G+N-1} & x_{L_G+N-2} & \dots & x_{L_G+N-N_o-1} \end{bmatrix}_{(N \times (N_f))} \tag{A9}$$

All elements of $\tilde{\mathbf{X}}$ are truth samples, which are assumed to be known for derivation. Then, according to (A6), we can derive the MSE as:

$$\begin{aligned}
 \hat{\sigma}^2 &= \frac{1}{N} \sum_{n=L_G}^{N+L_G-1} \mathbb{E} \{ |\hat{y}_n - \tilde{y}_n|^2 \} \\
 &= \frac{1}{N} \sum_{n=L_G}^{N_o-1} \sum_{m_1=n+1}^{N_o} \sum_{m_2=n+1}^{N_o} h_{m_1} h_{m_2}^* \mathbb{E} \{ (\hat{x}_{n-m_1} - x_{n-m_1}) (\hat{x}_{n-m_2} - x_{n-m_2})^* \} \\
 &+ \frac{1}{N} \sum_{n=N}^{N+L_G-1} \sum_{m_1=0}^{n-N} \sum_{m_2=0}^{n-N} h_{m_1} h_{m_2}^* \mathbb{E} \{ (\hat{x}_{n-m_1} - x_{n-m_1}) (\hat{x}_{n-m_2} - x_{n-m_2})^* \} \\
 &= \frac{\sigma_\varepsilon^2}{N} \sum_{n=L_G}^{N_o-1} \sum_{(m_1, m_2) \in \Lambda_1} h_{m_1} h_{m_2}^* \sum_{(l_1, l_2) \in \Lambda^{(1)}} \alpha_{l_1} \alpha_{l_2}^* \\
 &+ \frac{\sigma_\varepsilon^2}{N} \sum_{n=N}^{N+L_G-1} \sum_{(m_1, m_2) \in \Lambda_2} h_{m_1} h_{m_2}^* \sum_{(l_1, l_2) \in \Lambda^{(2)}} \alpha_{l_1} \alpha_{l_2}^*
 \end{aligned} \tag{A10}$$

where:

$$\begin{aligned}
 \Lambda_1 &= \left\{ (m_1, m_2) \left| \begin{array}{l} |m_1 - m_2| \leq k - 1 \\ n + 1 \leq m_1, m_2 \leq N_o \\ L_G \leq n \leq N_o - 1 \end{array} \right. \right\} \\
 \Lambda^{(1)} &= \left\{ (l_1, l_2) \left| \begin{array}{l} l_1 - l_2 = m_2 - m_1 \\ 0 \leq l_1, l_2 \leq k - 1 \\ (m_1, m_2) \in \Lambda_1 \end{array} \right. \right\} \\
 \Lambda_2 &= \left\{ (m_1, m_2) \left| \begin{array}{l} |m_1 - m_2| \leq k - 1 \\ 0 \leq m_1, m_2 \leq n - N \\ N \leq n \leq N + L_G - 1 \end{array} \right. \right\} \\
 \Lambda^{(2)} &= \left\{ (l_1, l_2) \left| \begin{array}{l} l_1 - l_2 = m_2 - m_1 \\ 0 \leq l_1, l_2 \leq k - 1 \\ (m_1, m_2) \in \Lambda_2 \end{array} \right. \right\}
 \end{aligned} \tag{A11}$$

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Stoica, P.; Moses, R. *Spectral Analysis of Signals*; Robbins T., Ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2005; pp. 1–3, 244–247.
2. Çetin, M.; Stojanović, I.; Önhon, N.O.; Varshney, K.R.; Samadi S.; Karl W.C.; Willsky A.S. Sparsity-driven synthetic aperture radar imaging: Reconstruction, autofocusing, moving targets, and compressed sensing. *IEEE Signal Process. Mag.* **2014**, *31*, 27–40.
3. Sun, K.; Liu, Y.; Meng, H.; Wang X. Adaptive sparse representation for source localization with gain/phase errors. *Sensors* **2011**, *11*, 4780–4793.
4. Stoica, P.; Babu, P.; Li, J. New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data. *IEEE Trans. Signal Process.* **2011**, *59*, 35–47.

5. Rao, S.; Pearlman, W.A. Spectral estimation from sub-bands. In Proceedings of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, Victoria, BC, Canada, 4–6 October 1992; pp. 69–72.
6. Bonacci, D.; Mailhes, C.; Djuric, P.M. Improving frequency resolution for correlation-based spectral estimation methods using sub-band decomposition. In Proceedings of 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, China, 6–10 April 2003; pp. 329–332.
7. Rao, S.; Pearlman, W.A. Analysis of linear prediction, coding, and spectral estimation from sub-bands. *IEEE Trans. Inf. Theory* **1996**, *42*, 1160–1178.
8. Lambrecht, C.B.; Karrakchou, M. Wavelet packets-based high-resolution spectral estimation. *Signal Process.* **1995**, *47*, 135–144.
9. Rouquette, S.; Berthoumieu, Y.; Najim, M. An efficient sub-band decomposition based on the Hilbert transform for high-resolution spectral estimation. In Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, Paris, France, 18–21 June 1996; pp. 409–412.
10. Tkacenko, A.; Vaidyanathan, P.P. The role of filter banks in sinusoidal frequency estimation. *J. Franklin Inst.* **2001**, *338*, 517–547.
11. Madisetti, V.K. *The Digital Signal Processing Handbook*; 2nd ed.; CRC Press: Boca Raton, FL, USA, 2010.
12. Narasimhan, S.V.; Harish, M. Spectral estimation based on sub-band decomposition by harmonic wavelet transform and modified group delay. In Proceedings of 2004 International Conference on Signal Processing and Communications, Bangalore, India, 11–14 December 2004; pp. 349–353.
13. Narasimhan, S.V.; Harish, M.; Haripriya, A.R.; Basumallick, N. Discrete cosine harmonic wavelet transform and its application to signal compression and sub-band spectral estimation using modified group delay. *Signal Image Video Process.* **2009**, *3*, 85–99.
14. Bonacci, D.; Michel, P.; Mailhes, C. Sub-band decomposition and frequency warping for spectral estimation. In Proceedings of European Signal and Image Processing Conference (EUSIPCO), Toulouse, France, 3–6 September 2002; pp. 147–150.
15. Stoica, P.; Sandgren, N.; Selén, Y.; Vanhamme, L.; Huffel, S.V. Frequency-domain method based on the singular value decomposition for frequency-selective NMR spectroscopy. *J. Magn. Reson.* **2003**, *165*, 80–88.
16. Brito, A.E.; Chan, S.H.; Cabrera, S.D. SAR image superresolution via 2-D adaptive extrapolation. *Multidimens. Syst. Signal Process.* **2003**, *14*, 83–104.
17. Papoulis, A. A new algorithm in spectral analysis and band-limited extrapolation. *IEEE Trans. Circult Syst.* **1975**, *CAS-22*, 735–742.
18. Devasia, A.; IAENG; Cada, M. Extrapolation of bandlimited signals using slepian functions. In Proceedings of the World Congress on Engineering and Computer Science, San Francisco, CA, USA, 23–25 October 2013; pp. 1–6.
19. Shi, J.; Sha, X.; Zhang, Q.; Zhang, N. Extrapolation of bandlimited signals in linear canonical transform domain. *IEEE Trans. Signal Process.* **2012**, *60*, 1502–1508.

20. Gosse, Laurent. Effective band-limited extrapolation relying on Selpian series and ℓ_1 regularization. *Comput. Math. Appl.* **2010**, *60*, 1259–1279.
21. Ljung, L. *System Identification: Theory for the User*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999; pp. 18–62.
22. Hamilton, J.D. *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 1994; pp. 72–116.
23. Stoica, P. Multistep prediction of autoregressive signals. *Electron. Lett.* **1993**, *29*, 554–555.
24. Oppenheim, A.V.; Schafer, R.W.; Buck, J.R. *Discrete-Time Signal Processing*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 1999; pp. 366–367.
25. Peter, J.R.; Richard, A.D. *Introduction to Time Series and Forecasting*, 2nd ed.; Casella, G., Fienberg, S., Olkin, I., Eds.; Springer-Verlag: New York, NY, USA, 2002; pp. 137–174.
26. McQuarrie, A.D.R.; Tsai, C.L. *Regression and Time Series Model Selection*; World Scientific Publishing: Singapore, Singapore, 1998; pp. 15–24.
27. Deng, Z.L.; Gao, Y.; Mao, L. New approach to information fusion steady-state Kalman filtering. *Automatica* **2005**, *41*, 1695–1707.
28. Deng, Z.L. *Optimal Estimation Theory with Applications-Modeling, Filtering and Information Fusion Estimation*; Harbin Institute of Technology Press: Harbin, China, 2005; pp. 47–102.
29. Stoica, P.; Selen, Y. Model-order selection: A review of information criterion rules. *IEEE Signal Process. Mag.* **2004**, *21*, 36–37.
30. Parks, T.W.; McClellan, J.H. Chebyshev approximation for nonrecursive digital filters with linear phase. *IEEE Trans. Circuit Theory* **1972**, *CT-19*, 189–194.
31. DeGraaf, S.R. SAR Imaging via Modern 2-D Spectral Estimation Methods. *IEEE Trans. Image Process.* **1998**, *7*, 729–761.
32. Li, J.; Stoica, P. An adaptive filtering approach to spectral estimation and SAR imaging. *IEEE Trans. Signal Process.* **1996**, *44*, 1469–1484.
33. Glentis, G.O. Efficient algorithms for adaptive Capon and APES spectral estimation. *IEEE Trans. Signal Process.* **2010**, *58*, 84–96.
34. RAO, B.D.; HARI, K.V.S. Performance analysis of root-Music. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 1939–1949.
35. Roy, R.; Kailath, T. ESPRIT-Estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995.
36. Yardibi, T.; Li, J.; Stoica, P.; Xue, M.; Braggeroer, A. B. Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares. *IEEE Trans. Aerosp. Electron. Syst.* **2010**, *46*, 425–443.
37. Xue, M; Xu, L.; Li, J. IAA spectral estimation: Fast implementation using the Gohberg-Semencul factorization. *IEEE Trans. Signal Process.* **2011**, *59*, 3251–3261.