The HKU Scholars Hub The University of Hong Kong 香港大學學術庫

Title	A New Regularized Adaptive Windowed Lomb Periodogram for Time-Frequency Analysis of Nonstationary Signals With Impulsive Components
Author(s)	Zhang, Z; Chan, SC; Wang, C
Citation	IEEE Transactions on Instrumentation and Measurement, 2012, v. 61 n. 8, p. 2283-2304
Issued Date	2012
URL	http://hdl.handle.net/10722/155737
Rights	IEEE Transactions on Instrumentation and Measurement. Copyright © IEEE

A New Regularized Adaptive Windowed Lomb Periodogram for Time–Frequency Analysis of Nonstationary Signals With Impulsive Components

Zhiguo Zhang, *Member, IEEE*, Shing-Chow Chan, *Member, IEEE*, and Chong Wang

***Abstract*—This paper proposes a new class of windowed Lomb periodogram (WLP) for time–frequency analysis of nonstationary signals, which may contain impulsive components and may be nonuniformly sampled. The proposed methods significantly extend the conventional Lomb periodogram in two aspects: 1) The nonstationarity problem is addressed by employing the weighted least squares (WLS) to estimate locally the time-varying periodogram and an intersection of confidence interval technique to adaptively select the window sizes of WLS in the time–frequency domain. This yields an adaptive WLP (AWLP) having a better tradeoff between time resolution and frequency resolution. 2) A more general regularized maximum-likelihood-type (M-) estimator is used instead of the LS estimator in estimating the AWLP. This yields a novel M-estimation-based regularized AWLP method which is capable of reducing estimation variance, accentuating predominant time–frequency components, restraining adverse influence of impulsive components, and separating impulsive components. Simulation results were conducted to illustrate the advantages of the proposed method over the conventional Lomb periodogram in adaptive time–frequency resolution, sparse representation for sinusoids, robustness to impulsive components, and applicability to nonuniformly sampled data. Moreover, as the computation of the proposed method at each time sample and frequency is independent of others, parallel computing can be conveniently employed without much difficulty to significantly reduce the computational time of our proposed method for real-time applications. The proposed method is expected to find a wide range of applications in instrumentation and measurement and related areas. Its potential applications to power quality analysis and speech signal analysis are also discussed and demonstrated.**

***Index Terms*—Adaptive window selection, Lomb periodogram, M-estimation, nonuniformly sampled data, power quality analysis, regularization, time–frequency analysis (TFA), weighted least squares (WLS).**

I. Introduction

TIME–FREQUENCY ANALYSIS (TFA) techniques are widely used in a variety of fields of instrumentation and measurement, such as power quality analysis [1]–[4], fault detection [5], biometric authentication [6], etc. It aims to discover time-varying spectral features of nonstationary data for further processing. Numerous TFA methods have been proposed with different strengths/limitations and targeted applications, and most of them stem from two types of spectral estimation methods: nonparametric and parametric. For nonparametric spectral estimation, the signal is assumed to consist of sinusoidal components, and their magnitudes and phases are estimated by means of filtering as in periodogram and minimum variance spectral estimation. In parametric methods, the signal is assumed to be generated by a certain model, and the spectrum is estimated from the model parameters. For a thorough discussion of classical spectral estimation methods, see [7]; for an introduction to modern spectral estimation, see [8]. See [9] for a recent review on nonparametric and parametric TFA methods. Basically, a high-performance TFA method should possess two desirable properties: 1) able to achieve as high as possible the time and frequency resolution for nonstationary data and 2) able to deal with various data corruptions and degradations, such as noise, outliers, impulsive components, missing data, and nonuniform samples.

In the context of nonparametric TFA, which is the focus of this paper, the first property is usually realized by the selection of window or kernel function in TFA methods, which serves as a time–frequency filter to extract the desired time–frequency sinusoidal component. In short-time Fourier transform (STFT), a window with an appropriate size should be chosen to balance between time resolution and frequency resolution [10]. Cohen's class TFA methods use a kernel function to mitigate cross-term interference while keeping frequency features [11]. The wavelet transform (WT) addresses the window selection problem by using a short window at high frequencies for good time resolution and a long window at low frequencies for good frequency resolution [12]. However, WT has a degraded frequency resolution for high-frequency components and a degraded time resolution for low-frequency components. Intuitively, the window should vary in the time–frequency domain in order to adapt to local time–frequency characteristics of the data. Typically, a long time window is desirable for identifying slowly varying sinusoidal components, while a small window is preferred for estimating fast-varying frequency components.

The second property determines the applicability and robustness of TFA methods in practice. Real-world data are usually contaminated with different types of noise and interferences. A good TFA method should be able to effectively restrain unwanted noise components and accentuate useful signal components. Since most TFA methods implicitly employ least squares (LS) estimation, they are sensitive to impulsive

Manuscript received July 5, 2011; revised December 19, 2011; accepted December 29, 2011. Date of publication February 23, 2012; date of current version July 13, 2012. The Associate Editor coordinating the review process for this paper was Dr. Kurt Barbe.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2012.2186655

0018-9456/$31.00 © 2012 IEEE

components with large amplitudes in time domain. Consequently, the estimated time–frequency representation may be degraded severely. Another problem of most TFA methods is that they were devised for uniformly sampled data and cannot be directly adopted for analyzing nonuniformly sampled data. Nonuniform data are commonly encountered, which may be acquired naturally by random/irregular sampling system or obtained unexpectedly due to missing data [13]. Currently, TFA of nonuniformly sampled data is generally achieved by two steps: 1) to resample or interpolate nonuniform data into uniform grids and 2) to perform conventional TFA methods on the resampled or interpolated uniform data. Unfortunately, the resampling operation induces unwanted low-pass filtering and artifacts in the spectrum of nonuniform data [14], [15]. Spectral analysis and TFA of nonuniformly sampled data are of increasing interest in many related areas [16]–[19]. Thus, it is necessary to develop a TFA method suitable for dealing with nonuniformly sampled data.

In this paper, we propose a new class of windowed Lomb-periodogram (WLP) TFA methods for achieving the two desirable properties discussed earlier, i.e., adaptive and high time–frequency resolution, robustness to impulsive components, applicability to nonuniform data, etc. It is based on the Lomb periodogram, which is an LS-based spectral analysis method for possibly nonuniformly sampled data [20]. The basic idea of Lomb periodogram is to estimate the amplitude of a given sinusoid with a certain frequency based on an LS fitting of the data samples to the sinusoid. We use the Lomb periodogram in our work because of the following advantages. First, it is based on a linear model fitting, which is a convenient framework for further generalizations and theoretical analyses. Second, it reduces to the STFT if the signal is uniformly sampled and the window size is fixed. Moreover, if the frequency points to be evaluated are also uniformly spaced, discrete Fourier transform and, hence, fast Fourier transform (FFT) can be used. Third, it has unique advantages such as applicability to nonuniformly sampled data and parallel implementation, since estimation at selected frequency points can be done in parallel. Fourth, it has found a wide range of practical applications [16]–[19]. However, the studies in [16]–[18] only used Lomb periodogram (no analysis window is adopted to localize periodogram) for spectral analysis of stationary nonuniform data, so that they are not suitable for TFA of nonstationary data with time-varying frequency components. In [19], recursive estimation of time-varying Lomb periodogram based on recursive LS was proposed, but it can only calculate Lomb periodogram at sampling time points and it does not study the selection of window (i.e., forgetting factor in recursive LS) for optimized time–frequency resolution. In our preliminary work [21], we extended the Lomb periodogram to analyze time-varying spectral characteristics of nonstationary nonuniformly sampled data, and the results were encouraging. This motivates us to perform an in-depth study in this paper on the asymptotic performance of this locally adaptive Lomb-periodogram-based TFA and to further extend it with a more general regularized maximum-likelihood-type estimation [22], [23] to address the two desirable properties mentioned earlier. Moreover, a novel practical application of the proposed TFA on power quality monitoring is studied to illustrate its potential usefulness.

More precisely, the proposed class of WLP methods significantly extends the Lomb periodogram in the following two aspects.

1) To better estimate the time-varying periodogram of nonstationary data (the nonstationarity problem), we employ a sliding window to select short-time data segments for estimating the Lomb periodogram locally by a weighted LS (WLS) fitting of the data segment to a given sinusoid. The proper selection of a locally adaptive window size of this WLP is of great importance, since it affects the tradeoff between time resolution and frequency resolution. In this paper, this tradeoff problem is studied in-depth. In particular, we derive asymptotic expressions for the bias and variance of the WLS estimator. The asymptotic expressions reveal that both the bias and variance are functions of the window bandwidth and there exists an optimal local bandwidth which minimizes the mean squared error (MSE). Since the optimal bandwidth involves quantities that may not be available in practice, we propose to approximate the optimal bandwidth by means of an intersection of confidence interval (ICI) technique [24]–[31]. With the ICI technique, varying window size can be adaptively determined in the time–frequency domain for WLP, resulting in an adaptive WLP (AWLP) with a good tradeoff between time resolution and frequency resolution.
2) We adopt a more general regularized maximum-likelihood-type (M-) estimator [22], [23], instead of the LS estimator, in estimating the AWLP, and this yields an M-estimation-based regularized AWLP (ME-RAWLP) method with improved applicability and robustness. The regularized M-estimator is a generalization of the LS estimator by incorporating a regularization term and using an M-estimation cost function. The regularization technique can considerably reduce the estimation variance and accentuate sinusoidallike components for signals with predominant frequency contents at the expense of a small bias. Some regularization techniques, such as L_1 regularization [32] and smoothly clipped absolute deviation (SCAD) [22], can produce sparse representations by shrinking small coefficients toward zero. Thus, regularization is particularly suitable for estimating time–frequency representations of sinusoidallike data, where only a few predominant frequency components exist. Computer simulations to be presented in Section VI show that the SCAD regularization is more suitable than L_2 and L_1 regularizations because it is unbiased for large coefficients. On the other hand, the M-estimation cost function can suppress possible impulsive components in the data, which considerably degrade the time–frequency representations of sinusoidallike components when quadratic LS cost function is used. Furthermore, unlike relevant studies which simply restrain or remove impulsive components [33], this study assumes that impulsive components may also contain

meaningful information and that they should be separated from the sinusoidal data and be represented in time domain. Therefore, we propose to reconstruct the sinusoidal components from the ME-RAWLP and then estimate the impulsive components from the difference between the reconstructed sinusoidallike data and the observed data. Finally, the analysis of nonstationary data using the ME-RAWLP method leads to a time–frequency representation for sinusoidal components and a time-domain representation for impulsive components. Such signals are frequently encountered in power quality monitoring and recognition [34]. To demonstrate inherent parallelism of the proposed method, we have implemented the proposed method in parallel using graphics processing units (GPUs) and found that the computational time can be dramatically reduced, owing to the parallel implementation.

Simulation results show that the proposed RAWLP method offers superior performance in the following aspects:

1) achieving better time and frequency resolution than WLP with a fixed window support;
2) enhancing the predominant sinusoidallike components by suppressing trivial components and noise;
3) avoiding unwanted low-pass effects which are brought about by conventional resampling-based TFA methods for nonuniformly sampled data;
4) restraining the adverse effect of impulsive components on the TFA and allowing them to be reconstructed separately;
5) allowing parallel implementation and facilitating real-time applications.

In addition, compared with STFT, the proposed RAWLP method has two advantages: 1) It can achieve better time–frequency resolution by selecting adaptive windows, and 2) it can avoid artifacts induced by interpolation when dealing with nonuniformly sampled data. It should be noted that the proposed RAWLP method forms a useful framework, which can be reduced or simplified to a class of AWLP methods (such as LS-RAWLP, ME-AWLP, etc.) with different performance/complexity tradeoffs by adopting different cost function in fitting the linear model. The good performance of this class of AWLP methods is also illustrated through its potential application in power quality analysis. Simulation results show that ME-RAWLP can effectively identify and disentangle oscillatory transients and impulsive transients from fundamental 50-Hz power signals.

This paper is organized as follows. In Section II, the Lomb periodogram is briefly reviewed to motivate the development of the WLP. Section III is devoted to the bias-variance tradeoff of the WLP and the selection of locally adaptive window for WLP using the ICI technique. The incorporation of regularization and M-estimation techniques to obtain the proposed ME-RAWLP method is detailed in Section IV. Other practical issues in applying the proposed WLP methods are discussed in Section V. Simulation results and comparisons are described in Section VI. Finally, conclusions are drawn in Section VII.

II. WLS-Based WLP

Suppose that a set of discrete-time noisy samples x_n, $n = 1, 2, \ldots, N$, is acquired by sampling a continuous-time signal at sampling time instants t_n, which are possibly nonuniformly distributed. The Lomb periodogram $P(\omega)$ is computed by an LS fitting of the data x_n by a sinusoid with a given frequency ω. More precisely, the data x_n are described by the following linear regression model:

$$\begin{aligned} x_n &= a(\omega)\cos(\omega t_n) + b(\omega)\sin(\omega t_n) + e(t_n) \\ &= \boldsymbol{\phi}^{\mathrm{T}}(t_n, \omega)\boldsymbol{\beta}(\omega) + e(t_n) \end{aligned} \tag{1}$$

where x_n is the response variable, $\boldsymbol{\phi}(t_n, \omega) = [\cos(\omega t_n), \sin(\omega t_n)]^{\mathrm{T}}$ is the explanatory variable, and $\boldsymbol{\beta}(\omega) = [a(\omega), b(\omega)]^{\mathrm{T}}$ is the regression coefficient vector. Assuming that the additional components $e(t_n)$ are a zero-mean Gaussian process, the coefficient vector $\boldsymbol{\beta}(\omega)$ can be obtained by minimizing the MSE as

$$\begin{aligned} \hat{\boldsymbol{\beta}}(\omega) &= \arg\min_{\boldsymbol{\beta}} \sum_{n=1}^{N} e^2(t_n) \\ &= \arg\min_{\boldsymbol{\beta}} \sum_{n=1}^{N} \left[x_n - \phi^{\mathrm{T}}(t_n, \omega)\boldsymbol{\beta}(\omega)\right]^2 \\ &= \arg\min_{\boldsymbol{\beta}} \|\boldsymbol{X} - \boldsymbol{\Phi}(\omega)\boldsymbol{\beta}(\omega)\|_2^2 \end{aligned} \tag{2}$$

where $\boldsymbol{X} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ and $\boldsymbol{\Phi}(\omega) = \begin{bmatrix} \cos(\omega t_1), & \cos(\omega t_2), & \ldots, & \cos(\omega t_N) \\ \sin(\omega t_1), & \sin(\omega t_2), & \ldots, & \sin(\omega t_N) \end{bmatrix}^{\mathrm{T}}$. The LS solution to the aforementioned function is given by

$$\hat{\boldsymbol{\beta}}(\omega) = \left[\boldsymbol{\Phi}^{\mathrm{T}}(\omega)\boldsymbol{\Phi}(\omega)\right]^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(\omega)\boldsymbol{X}. \tag{3}$$

The Lomb periodogram at frequency ω can then be computed from the LS-based estimate of $\hat{\boldsymbol{\beta}}(\omega)$ as

$$P(\omega) = \frac{1}{N}\hat{\boldsymbol{\beta}}^{\mathrm{T}}(\omega)\boldsymbol{\Phi}^{\mathrm{T}}(\omega)\boldsymbol{\Phi}(\omega)\hat{\boldsymbol{\beta}}(\omega). \tag{4}$$

Usually, $P(\omega)$ is computed over a dense set of frequency points to provide a spectral distribution of the signal in the frequency domain. For a sinusoidallike signal which may contain multiple components, the number of sinusoidal components and their frequencies can be determined by detecting significantly dominant peaks of the Lomb periodogram using, for example, Bayesian information criterion [15]. At each dominant frequency ω, the estimated $\hat{\boldsymbol{\beta}}(\omega)$ can be used to reconstruct a sinusoidal waveform of frequency ω as $x_{n,\omega} = a(\omega)\cos(\omega t_n) + b(\omega)\sin(\omega t_n) = A(\omega)\sin(\omega t_n + \varphi(\omega))$, where $A(\omega) = \sqrt{a^2(\omega) + b^2(\omega)}$ and $\varphi(\omega) = \arctan[a(\omega)/b(\omega)]$ are the corresponding amplitude and phase, respectively. These sinusoidal components can be reconstructed separately at each dominant frequency, and their sum approximates the multicomponent sinusoidal signal. An important advantage of the Lomb periodogram is that it is able to detect frequency components larger than half of the average sampling frequency of nonuniformly sampled data (i.e., the

Nyquist frequency of uniformly sampled data) [14]. In fact, for nonuniform data, the highest frequency that can be identified by the Lomb periodogram is $\varpi = \pi/\Delta T_{\min}$, where $\Delta T_{\min}$ is the minimum sampling interval in the whole data.

The Lomb periodogram can be extended to perform TFA for nonstationary data with time-varying spectrum. At a given time instant τ, a time window around τ is applied to the data to localize the spectral information. The time window can be obtained by scaling a basis window function $w(\cdot)$ by a bandwidth parameter of h, i.e., $w_h(t-\tau) = (1/h)w_h((1/h)(t-\tau))$. The bandwidth h controls the effective size of the window and, hence, the number of neighboring data around τ used to estimate the periodogram. For a window $w_h(t-\tau)$, the cost function of (2) becomes

$$\hat{\boldsymbol{\beta}}(\tau,\omega) = \arg\min_{\boldsymbol{\beta}} \sum_{n=1}^{N} w_h(t_n-\tau)\left[x_n - \boldsymbol{\phi}^{\mathrm{T}}(t_n,\omega)\boldsymbol{\beta}(\tau,\omega)\right]^2 = \arg\min_{\boldsymbol{\beta}} \left\| \widetilde{\boldsymbol{W}}\left(\boldsymbol{X} - \boldsymbol{\Phi}(\omega)\boldsymbol{\beta}(\tau,\omega)\right)\right\|_2^2 \quad (5)$$

where $\boldsymbol{\beta}(\tau,\omega) = [a(\tau,\omega), b(\tau,\omega)]^{\mathrm{T}}$ is the local coefficient vector and $\widetilde{\boldsymbol{W}} = \mathrm{diag}\{[\sqrt{w_h(t_1-\tau)},\ldots,\sqrt{w_h(t_n-\tau)}]^{\mathrm{T}}\}$. The estimation of $\boldsymbol{\beta}(\tau,\omega)$ is a WLS fitting problem, and its solution is given by

$$\hat{\boldsymbol{\beta}}(\tau,\omega) = \left[\boldsymbol{\Phi}^{\mathrm{T}}(\omega)\boldsymbol{W}\boldsymbol{\Phi}(\omega)\right]^{-1}\boldsymbol{\Phi}^{\mathrm{T}}(\omega)\boldsymbol{W}\boldsymbol{X} \quad (6)$$

where $\boldsymbol{W} = \widetilde{\boldsymbol{W}}\widetilde{\boldsymbol{W}}$. If we evaluate the Lomb periodogram at a set of uniformly distributed evaluated time instants τ_m, $m = 1, 2, \ldots, M$, a WLP will be obtained, providing a time–frequency representation for the data. The WLP can be directly used for dealing with nonuniformly sampled data, which is an important advantage compared with conventional TFA methods. Another useful property of the WLP is that it is well suited for analyzing sinusoidal data, since it is based on a fitting to sinusoids. The reconstruction of a sinusoidal signal with time-varying spectrum from WLP can be achieved by a pointwise reconstruction. More precisely, at each evaluated time instant, we estimate the Lomb periodogram at a set of frequency points using windowed data, where the spectrum is assumed to be fixed, and then reconstruct the windowed data using the most dominant sinusoidal component(s). The reconstructed windowed data at the evaluated time instant are then divided by the window weight and retained as the reconstructed value for the whole data at this time. By conducting the reconstruction at each time instant, we can reconstruct a signal which may contain time-varying sinusoidal components.

A main problem of the WLP method is the selection of an appropriate window size. As mentioned before, the performance of WLP is greatly influenced by the bandwidth h, and a locally adaptive bandwidth is highly desirable. In the next section, we shall derive expressions for the asymptotic bias and variance of the WLS estimator in (6), mainly to illustrate theoretically how the window size determines the WLP and to validate the existence of a local optimal window size.

Finally, it should be noted that the Lomb periodogram focuses on the estimation of one frequency at a time using the linear model in (1). In general, the estimation may not be completely white, and a more complete linear model consisting of a dense set of sinusoidal components may be required. To reduce the estimation variance, regularization techniques have to be used. Consequently, the arithmetic complexity required is rather high, and it cannot be parallelized easily. In the AWLP to be introduced later, we adopt the simplified linear model in (1) and estimate the bandwidth and regularization parameter for each sinusoidal component independently. Therefore, the arithmetic complexity is considerably reduced, and the estimation at each frequency can be easily parallelized for implementation in hardware, digital signal processors (DSPs), and GPUs. Moreover, we found that the performance of the AWLP is satisfactory, particularly for time–frequency representations of sinusoidallike data, where only a few predominant frequency components exist.

III. Adaptive Window Selection in WLP

A. Asymptotic Bias and Variance of WLS

The underlying relationship between the coefficient estimate and the window size can be revealed by asymptotic analysis of the WLS estimator in (6). This falls into the more general local polynomial modeling estimator recently proposed in [28] and [29]. Based on these works, we have derived the asymptotic expressions for the bias and variance of (6) and have validated them in a supplementary material (http://www.eee.hku.hk/~zgzhang/publication/tim2011_supp.pdf). It is shown that the asymptotic bias and variance of the lth coefficient $\hat{\beta}_l(\tau,\omega)$, $l = 1$ or 2 ($\hat{\beta}_1 = \hat{a}$ and $\hat{\beta}_2 = \hat{b}$), are, respectively

$$Bias\left(\hat{\beta}_l(\tau,\omega)\right) = \frac{h\overline{\nu}_w}{\nu_w}\beta_l'(\tau,\omega) + o_P(h) \quad (7)$$

$$Var\left(\hat{\beta}_l(\tau,\omega)\right) = \frac{\widetilde{\nu}_w\sigma^2(\tau)}{hNf(\tau)\nu_w^2\phi_l^2(\tau,\omega)} + o_P\left(\frac{1}{Nh}\right) \quad (8)$$

where $\nu_w = \int_v w(v)dv$, $\overline{\nu}_w = \int_v v^2 w^2(v)dv$, $\widetilde{\nu}_w = \int_v w^2(v)dv$, β_l' is the first derivative of β_l, $f(\tau)$ is the sampling density at τ, and $\sigma^2(\tau)$ is the variance of the estimation residual of the LS fitting.

It can be seen from (7) and (8) that both the estimation bias and variance are functions of the bandwidth h. As h increases, the squared bias will increase while the variance will decrease. Since the MSE of the WLS estimator is the sum of squared bias and variance as

$$MSE\left(\hat{\beta}_l(\tau,\omega)\right) = Bias^2\left(\hat{\beta}_l(\tau,\omega)\right) + Var\left(\hat{\beta}_l(\tau,\omega)\right) \quad (9)$$

there exists a locally optimal bandwidth $h^+(\tau,\omega)$ to minimize the MSE criterion. By setting the derivative of (9) with respect to h to zero, the following optimal bandwidth at (τ,ω) is obtained:

$$h_l^+(\tau,\omega) = \left\{\frac{\widetilde{\nu}_w\sigma^2(\tau)}{2Nf(\tau)\phi_l^2(\tau,\omega)\left[\beta_l'(\tau,\omega)\right]^2\overline{\nu}_w^2}\right\}^{1/3}. \quad (10)$$

It can be seen that all the variables in the right side of (10) are inherent either in data or in the basis kernel so that the existence of an optimal asymptotic bandwidth is validated. However, this result, although useful in analytical work, cannot be used directly in practice because some quantities in (10), such as $\beta_l'(\tau,\omega)$, are difficult to be determined accurately. Moreover, in the context of spectral estimation, the input, which is the sinusoid with known frequency, is deterministic, and as $h \to 0$ in the asymptotic analysis, it can only capture the local characteristics of a sinusoid rather than its long-range behavior. This can be seen from (10) that $h_l^+(\tau,\omega)$ involves the local power of the sinusoidal function at location τ, $\phi_l^2(\tau,\omega)$. Therefore, in order to get a better result by capturing the long-range behavior, or in other words overcome the limitation of the requirement of the asymptotic analysis, a "data-driven" technique is essential to approximate the optimal bandwidth from a finite set of possible bandwidths with wider ranges of values.

B. Adaptive Varying Bandwidth Selection by ICI

In this paper, we employ an ICI technique to estimate the optimal bandwidth. The ICI technique is an empirical bandwidth selection technique first proposed in [25], and it has been successfully applied in many areas, such as data smoothing [26], image processing [27], and TFA [29]. Numerous adaptive window selection methods exist, and the ICI technique is chosen for its advantages of low arithmetic complexity and good performance in various conditions [26]. The ICI technique was initially proposed based on the assumption of Gaussianity of residuals, which may not hold when calculating the Lomb periodogram of multiple sinusoidal or nonsinusoidal signals. However, it has been shown in [30] and [31] that the ICI technique can still achieve decent results for non-Gaussian residuals, and thus, it is adopted in this study.

The ICI technique starts from a set of prescribed bandwidth parameters in an ascending order

$$\boldsymbol{H} = \{h_1 < h_2 < \cdots < h_J\} \tag{11}$$

where J is the number of candidate bandwidths. Then, the ICI technique determines a suboptimal bandwidth (the "optimal" one in the bandwidth set) by comparing the confidence intervals of the estimates with different bandwidths in the set. The confidence intervals $D_j = [L_j, U_j]$ are obtained from the estimated $\hat{\beta}_l(\tau,\omega;h_j)$ with different bandwidth h_j

$$U_j = \hat{\beta}_l(\tau,\omega;h_j) + \Gamma\sqrt{Var\left(\hat{\beta}_l(\tau,\omega;h_j)\right)} \tag{12}$$

$$L_j = \hat{\beta}_l(\tau,\omega;h_j) - \Gamma\sqrt{Var\left(\hat{\beta}_l(\tau,\omega;h_j)\right)} \tag{13}$$

where Γ is a threshold parameter used to adjust the width of the confidence interval. For example, $\Gamma = 2.58$ implies a 99% confidence interval. In practice, we select Γ from a set of threshold values $\{\Gamma_1, \Gamma_2, \ldots, \Gamma_L\}$ as the one that minimizes the sum of the weighted squared residuals [24]

$$\Gamma = \arg\min_{\Gamma}\left\{\sum_{n=1}^{N}\left(\frac{x_n - \boldsymbol{\phi}^{\mathrm{T}}(t_n,\omega)\hat{\boldsymbol{\beta}}_{(\Gamma)}(\tau,\omega)}{1 - w_h(t_n-\tau)}\right)^2\right\} \tag{14}$$

where $\hat{\boldsymbol{\beta}}_{(\Gamma_l)}$ is the coefficient estimated using the ICI technique with the threshold parameter Γ_l.

Because the squared bias increases and variance decreases with the increase of h, the length of confidence interval will gradually decrease with h while the center of the interval remains more or less fixed. When h is increased to a point where the observations cannot be satisfactorily modeled, a large bias is produced, and the confidence interval will no longer intersect those with smaller values of h. Motivated by this observation, the ICI technique examines the following quantities from the confidence intervals in order to detect this sudden change:

$$\begin{aligned}\overline{L}_j &= \begin{cases} 0, & \text{for } j = 1 \\ \max[\overline{L}_{j-1}, L_j], & \text{for } j = 2,3,\ldots,J \end{cases} \\ \underline{U}_j &= \begin{cases} 0, & \text{for } j = 1 \\ \min[\underline{U}_{j-1}, U_j], & \text{for } j = 2,3,\ldots,J \end{cases}\end{aligned} \tag{15}$$

where $\overline{L}_j$ is the largest lower bound of the confidence interval for bandwidth up to h_j while $\underline{U}_m$ is the corresponding upper bound. The largest value of these j's for which $\underline{U}_j \geq \overline{L}_j$ gives the desirable bandwidth $h_l^+(\tau,\omega)$ because the confidence intervals no longer intersect with each other above the bandwidth $h_l^+(\tau,\omega)$.

In the ICI technique, $Var(\hat{\beta}_l(\tau,\omega))$ has to be approximated to construct the confidence intervals in (12) and (13). Assuming local homoscedasticity (i.e., the noise variance is identical in the window), a finite-sample approximation of the covariance matrix of $\hat{\boldsymbol{\beta}}$ (τ and ω are omitted for notational simplicity) can be derived from (6) as

$$\begin{aligned} Cov(\hat{\boldsymbol{\beta}}) &= E\left\{\left[\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right]\left[\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}})\right]^{\mathrm{T}}\right\} \\ &= (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}E\left[(\boldsymbol{X}-\boldsymbol{m})(\boldsymbol{X}-\boldsymbol{m})^{\mathrm{T}}\right] \\ &\quad \times \boldsymbol{W}\boldsymbol{\Phi}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Phi})^{-1} \\ &= (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Sigma}_{\sigma}\boldsymbol{W}\boldsymbol{\Phi}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Phi})^{-1} \\ &= (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{W}\boldsymbol{\Phi}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Phi})^{-1}\sigma^2(\tau) \end{aligned} \tag{16}$$

where $\boldsymbol{\Sigma}_{\sigma} = \mathrm{diag}\{\sigma^2(t_1), \sigma^2(t_2), \ldots, \sigma^2(t_N)\}$. The residual variance $\sigma^2(\tau)$ is estimated as the normalized weighted residual sum of squares [35]

$$\hat{\sigma}^2(\tau) = \frac{\left\|\boldsymbol{W}(\boldsymbol{X} - \boldsymbol{\Phi}\hat{\boldsymbol{\beta}})\right\|_2^2}{tr\left\{\boldsymbol{W} - \boldsymbol{W}\boldsymbol{\Phi}(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\right\}}. \tag{17}$$

With (16) and (17), the variance of $\hat{\beta}_l(\tau,\omega;h_j)$ can be obtained, and the ICI technique can be implemented to select a suboptimal bandwidth from the set $\boldsymbol{H}$.

The ICI technique selects bandwidths $h_1^+(\tau,\omega)$ and $h_2^+(\tau,\omega)$ for $a(\tau,\omega)$ and $b(\tau,\omega)$ separately. However, because $a(\tau,\omega)$

and $b(\tau,\omega)$ are estimated jointly (using one common bandwidth), not separately (using different bandwidths) in one WLS estimation, $h_1^+(\tau,\omega)$ and $h_2^+(\tau,\omega)$ will be combined (generally, averaged [29]) to produce the suboptimal bandwidth for the coefficient vector $[a(\tau,\omega),b(\tau,\omega)]^{\mathrm{T}}$, i.e., $h^+(\tau,\omega)=[h_1^+(\tau,\omega)+h_2^+(\tau,\omega)]/2$.

As observed in [28] and [29], the varying suboptimal bandwidth estimated by the ICI technique usually exhibits considerable variability since it is based on a finite-sample approximation of the local noise variance. Hence, $h^+(\tau,\omega)$ can be further smoothed in the time–frequency domain to reduce the variance. Although there are a number of methods for smoothing a 2-D signal, it is still desirable to avoid over- or undersmoothing. This can be achieved by performing a steering-kernel-based local polynomial regression with ICI method (SK-LPR-ICI) [27], which was originally proposed for image smoothing. The SK-LPR-ICI method employs steering kernels with local orientation and adaptive scale to adapt better to local characteristics of 2-D signals. Therefore, the SK-LPR-ICI method can effectively suppress the noise while preserving well the local information, making it a suitable technique for smoothing the 2-D bandwidth function $h^+(\tau,\omega)$. If it is required to calculate the WLP at each frequency in parallel to reduce the computational time, then bandwidth can be smoothed only along the time axis.

We now summarize the proposed AWLP as follows.

- Step 1) At each evaluated time–frequency point (τ,ω), the coefficient $\hat{\boldsymbol{\beta}}(\tau,\omega;h_j)$ is calculated by (6) with each bandwidth h_j taken from $\boldsymbol{H}$ in (11).
- Step 2) An suboptimal bandwidth $h^+(\tau,\omega)$ is estimated using the ICI technique (12)–(15) based on the estimated $\hat{\boldsymbol{\beta}}(\tau,\omega;h_j)$, $j=1,\ldots,J$.
- Step 3) The 2-D bandwidth function $h^+(\tau,\omega)$ is further smoothed by the SK-LPR-ICI method to obtain a smoothed bandwidth function $\widetilde{h}^+(\tau,\omega)$.
- Step 4) The final coefficient estimate $\hat{\boldsymbol{\beta}}(\tau,\omega;\widetilde{h}^+(\tau,\omega))$ is obtained using WLS with $\widetilde{h}^+(\tau,\omega)$.

IV. ME-RAWLP

The selection of bandwidth in WLS to approach the minimum MSE has been addressed in the previous section. However, it is still desirable to further lower the variance and the MSE of the WLS estimator so as to accentuate predominant time–frequency components in WLP. This is also useful in situation when the number of data samples is limited. In addition, due to the use of LS criterion in the fitting process, impulsive components, which appear as outliers to the model, will significantly affect the bandwidth selection, and hence, the periodogram will be computed. These motivate us to adopt a more general regularized maximum-likelihood-type (M-) estimator instead of the LS estimator to compute the Lomb periodogram. The regularized M-estimator was first proposed in [22] as a generalization of the LS estimator by using an M-estimation cost function and incorporating a regularization term. In statistics, the M-estimation is well known as a more robust approach than the LS estimation to impulses or outliers. It is also well known that adding a regularization term can considerably reduce the variance at the expense of a small bias, resulting in an estimator with a smaller MSE. Note that the regularization techniques are more suitable and useful for dealing with high-dimensional variable selection problems, where estimation variance is large and variable selection is required. However, regularization techniques can also be used in the two-variable cases as well for the purpose of lowering variance and suppressing noise.

The regularized M-estimator is obtained by minimizing the cost function (τ and ω are omitted here for notational simplicity)

$$J_\rho(\tau,\omega)=\sum_{n=1}^{N} w_h(t_n-\tau)\rho\left(x_n-\boldsymbol{\phi}^{\mathrm{T}}(t_n,\omega)\boldsymbol{\beta}\right)+\sum_{l=1}^{2} p_\mu(\beta_l) \tag{18}$$

where $\rho(\cdot)$ is an M-estimate function and $p_\mu(\cdot)$ is the regularization function with μ being the regularization parameter. The first and second terms on the right-hand side of (18) represent the robust error measure of $\boldsymbol{\beta}$ and the regularization term, respectively. If $\rho(e)=e^2$, we obtain the regularized LS estimator. From a Bayesian perspective, the regularization is closely related to incorporating prior information on the random variable $\boldsymbol{\beta}$, while the WLS term represents the likelihood term. The regularization parameter μ is used to adjust the relative contributions from prior information and the likelihood function.

In this paper, we use the Huber function as the M-estimate function to deemphasize the "outliers" (impulsive components)

$$\rho(e)=\begin{cases} e^2/2, & 0<|e|<\xi \\ \xi|e|-\xi^2/2, & |e|\geq\xi \end{cases} \tag{19}$$

where ξ is a threshold used to restrain the outliers. Other M-estimate functions such as the Hampel three-part redescending function can also be used here. In [36], a systematic approach for estimating the threshold was proposed. The threshold ξ is generally determined from the standard deviation of the "impulse-free" estimation residual, σ_e, which can be approximated as [36]

$$\hat{\sigma}_e=\{median\,(|x_n-x_{n-1}|)\}/(\sqrt{2}\cdot 0.6745),\ n=2,\ldots,N. \tag{20}$$

For Huber function, the threshold can be set as $\xi=2.58\cdot\hat{\sigma}_e$, implying a 99% confidence to detect and reject the outliers. As a result, the effect of errors with large amplitudes will be reduced substantially beyond the threshold. For online applications, $\hat{\sigma}_e$ can be estimated recursively using the technique proposed in [36].

Commonly used regularization functions include the L_2 regularization function $p_\mu(\beta_l)=\mu\beta_l^2$, which leads to a ridge regression, and the L_1 regularization function $p_\mu(\beta_l)=\mu|\beta_l|$, which leads to a least absolute shrinkage and selection operator (lasso) [30]. We also test a SCAD regularization function because it yields an estimator with three desirable properties:

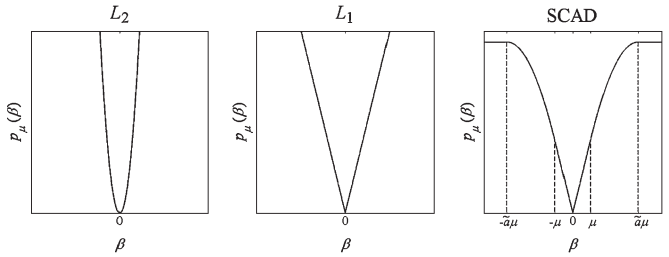


Fig. 1. L_2, L_1, and SCAD regularization functions.

unbiasedness, sparsity, and continuity [22]. The SCAD regularization function is given as follows:

$$p_\mu(\beta_l) = \begin{cases} \mu|\beta_l|, & \text{for } |\beta_l| \le \mu \\ -\frac{\left(|\beta_l|-\widetilde{a}\mu\right)^2}{2(\widetilde{a}-1)} + \frac{(\widetilde{a}+1)\mu^2}{2}, & \text{for } \mu < |\beta_l| \le \widetilde{a}\mu \\ \frac{(\widetilde{a}+1)\mu^2}{2}, & \text{for } |\beta_l| > \widetilde{a}\mu \end{cases} \tag{21}$$

where the parameter $\widetilde{a} > 2$ is, in general, selected as 3.7 [22]. The three regularization functions are shown in Fig. 1.

Ridge, lasso, and SCAD estimators all satisfy the condition of continuity, which means that $p_\mu(\beta_l)$ are continuous in the coefficient β_l. The lasso and SCAD estimators have the property of sparsity, which means that small coefficients can be set to zero, while the ridge estimator does not have. Therefore, if the data under study are sampled from sinusoidal signals, sparse estimators are preferable because only a few large coefficients are nonzero in this case. Unbiasedness implies that the modeling bias introduced by the regularization term should be zero when the true coefficients are large enough. Among the three regularization techniques, only SCAD has the property of unbiasedness. As a result, the SCAD estimator is desirable for estimating the amplitudes of sinusoids because the large coefficients will not be corrupted by extra bias.

The solution of (18) can be obtained by setting its derivative with respect to $\boldsymbol{\beta}$ to zero, which yields

$$-\sum_{n=1}^{N} w_h(t_n-\tau)\rho'\left(e(t_n)\right)\phi(t_n,\omega) + \boldsymbol{P}(\boldsymbol{\beta}) = \boldsymbol{0} \tag{22}$$

where $\boldsymbol{P}(\boldsymbol{\beta}) = [p'_\mu(\beta_1), p'_\mu(\beta_2)]^{\mathrm{T}}$. To solve the equation using iteratively reweight LS (IRLS) method, we first write $\rho'(e(t_n))$ as $\rho'(e(t_n)) = q(e(t_n))e(t_n)$ and then $p'_\mu(\beta_l)$ as $p'_\mu(\beta_l) = \psi_u(\beta_l)\beta_l$. Then, $q(e(t_n))$ and $\psi_u(\beta_l)$ are approximated iteratively from the previous coefficient estimate of $\boldsymbol{\beta}$. More precisely, given the estimate $\hat{\boldsymbol{\beta}}^{(i)}$ at the ith iteration, we have $q(e^{(i)}(t_n)) = \rho'(e^{(i)}(t_n))/e^{(i)}(t_n)$, where $e^{(i)}(t_n) = x_n - \phi^{\mathrm{T}}(t_n,\omega)\hat{\boldsymbol{\beta}}^{(i)}$ and $p'_\mu(\hat{\beta}_l^{(i)}) = \psi_u(\hat{\beta}_l^{(i)})\hat{\beta}_l^{(i)}$. For nondifferentiable L_1 and SCAD regularization functions $p_\mu(\beta_l)$, it is usual to approximate them locally at $\hat{\boldsymbol{\beta}}^{(i)}$ as a quadratic function as follows [22]:

$$p_\mu(\beta_l) \approx p_\mu\left(\hat{\beta}_l^{(i)}\right) + \frac{1}{2}\left[p'_\mu\left(\hat{\beta}_l^{(i)}\right)/\hat{\beta}_l^{(i)}\right]\left[\beta_l^2 - \left(\hat{\beta}_l^{(i)}\right)^2\right] \tag{23}$$

for $\hat{\beta}_l^{(i)} \neq 0$ and $p_\mu(\beta_l) = 0$ for $\hat{\beta}_l^{(i)} = 0$.

The aforementioned iteration and approximation lead to the following IRLS estimator of $\boldsymbol{\beta}$ [22], [23]:

$$\hat{\boldsymbol{\beta}}^{(i+1)} = \left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{V}^{(i)}\boldsymbol{\Phi} + \boldsymbol{\Psi}^{(i)}\right)^{-1}\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{V}^{(i)}\boldsymbol{X}\right) \tag{24}$$

where $i = 0, 1, \ldots$ is the number of iteration, $\boldsymbol{\Psi}^{(i)} = \mathrm{diag}\{\psi_\mu(\beta_1^{(i)}), \psi_\mu(\beta_2^{(i)})\}$, and $\boldsymbol{V}^{(i)} = \boldsymbol{W}\boldsymbol{Q}^{(i)}$ with $\boldsymbol{Q}^{(i)} = \mathrm{diag}\{[q(e^{(i)}(t_1)), \ldots, q(e^{(i)}(t_N))]^{\mathrm{T}}\}$. The IRLS estimator can start with $\boldsymbol{\Psi}^{(0)} = \boldsymbol{I}_2$ and $\boldsymbol{Q}^{(0)} = \boldsymbol{I}_N$ and will stop until the maximum number of iterations is reached or the difference between two successive iterations is small enough. The covariance matrix of the regularized WLS estimator of (24) is

$$Cov\left(\hat{\boldsymbol{\beta}}^{(i+1)}\right) = \left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{V}^{(i)}\boldsymbol{\Phi} + \boldsymbol{\Psi}^{(i)}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{V}^{(i)}\boldsymbol{V}^{(i)}\boldsymbol{\Phi} \times \left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{V}^{(i)}\boldsymbol{\Phi} + \boldsymbol{\Psi}V^{(i)}\right)^{-1}\sigma^2(\tau). \tag{25}$$

In practical implementations, the selection of regularization parameters μ and $\widetilde{a}$ is also an important issue. A universal value $\mu = \sqrt{2\ln(L)}$, where L is the number of variables, was suggested in [22], and thus, one gets $\mu = 1.2$ in this study. A recommend value of 3.7 is widely used in SCAD regularization [22]. We can also search the best values of μ and $\widetilde{a}$ over a grid as the one minimizing the generalized cross-validation (GCV) criterion as

$$[\mu, \widetilde{\alpha}] = \arg\min_{\mu,\widetilde{\alpha}}\left\{\frac{\frac{1}{N}\left\|\boldsymbol{X} - \boldsymbol{\Phi}\hat{\boldsymbol{\beta}}_{(\mu,\widetilde{\alpha})}\right\|_2^2}{\left(1 - \frac{1}{N}tr\left[\boldsymbol{\Phi}\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \boldsymbol{\Psi}^{(I)}_{(\mu,\widetilde{\alpha})}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\right]\right\}^2}\right\} \tag{26}$$

where $\hat{\boldsymbol{\beta}}_{(\mu,\tilde{\alpha})}$ is the coefficient estimate with μ and $\widetilde{a}$ while $\boldsymbol{\Psi}^{(I)}_{(\mu,\tilde{\alpha})}$ is the diagonal weighting matrix at the final iterative step.

Since it involves multiobjective optimization, it is difficult to derive analytical expressions to jointly optimize the bandwidth and regularization parameters. In addition, it is also computationally expensive to search for their optimal combination in a 2-D or even 3-D grid. Therefore, we propose to select the two types of parameters in two steps: 1) h is selected using the ICI technique, and it can approach the optimal solution of the WLS estimator without regularization; 2) μ and $\widetilde{a}$ are selected based on the "optimal" h. Although such parameter selection is not optimal, it works well and achieves satisfactory results in our simulations. As mentioned before, the bandwidth and regularization parameters may show some variability because they are determined by the local variance estimation from finite number of samples. Such estimation variability can be suppressed by a 2-D smoothing technique in the time–frequency domain, either on 2-D parameters or on the amplitude estimates. To preserve the time–frequency resolution, the SK-LPR-ICI method [27] can be used to smooth bandwidth and regularization parameters.

The regularized M-estimator can be implemented in Step 4 of the proposed AWLP method, instead of the conventional WLS, yielding the ME-RAWLP. The proposed ME-RAWLP method

forms a useful framework, which can be reduced or simplified to a class of AWLP methods (such as LS-RAWLP, ME-AWLP, etc.) with different performance/complexity tradeoffs by adopting a different cost function in fitting the linear model.

The ME-RAWLP method is capable of providing a robust time–frequency representation free from the adverse influence of impulsive components. Furthermore, using the ME-RAWLP estimates of the sinusoidal components, we can reconstruct all the sinusoidallike components, and the impulsive components can be estimated as the difference between the reconstructed signal and the original observation. Such separation of impulsive components offers representations of these impulsive components in time domain, which is very helpful to disclose important information conveyed by the impulses. This is important, for instance, in power quality analysis where the impulsive components may be due to various faults to be analyzed and recognized in the future.

V. Practical Issues

A. Selection of Method

Up to now, we have proposed a class of WLP methods. The ME-RAWLP method, which integrates ICI window selection and regularized M-estimation, is well suited for general spectral analysis. In real applications, some of the features in ME-RAWLP may not be necessary and may be omitted to meet tight requirement in computational complexity. The problem of methodology selection is detailed in the last paragraph of Section VI after the presentation of extensive simulation results and comparisons.

B. Selection of Evaluated Time and Frequency Points

The proposed method is readily applicable to uniformly and nonuniformly sampled data. To achieve a good time–frequency representation of uniformly and nonuniformly sampled data, the WLP and its extensions can be evaluated at a set of uniformly distributed time points. Considering the compromise between precision and complexity, the sampling interval of evaluated time points can be equal to the mean sampling interval of the nonuniform data samples.

In Lomb periodogram, the evaluated frequency points are usually selected as a uniform grid with a step $\Delta\omega$ much smaller than the frequency resolution $\omega_R = 2\pi/(t_N - t_1)$, such as $\Delta\omega = \omega_R/10$, so that no frequency information could be omitted [15]. In WLP, since a window with bandwidth h is introduced, the frequency resolution of a windowed data segment at time τ is $\omega_R^{(w)}(\tau, h) = 2\pi/(t_N^{(w)}(\tau, h) - t_1^{(w)}(\tau, h))$, where $t_N^{(w)}$ and $t_1^{(w)}$ are, respectively, the last and the first sampling time in the windowed data. We can see that the frequency resolution varies at different time and with different bandwidth. Thus, the frequency step can be set as $\Delta\omega = \max(\omega_R^{(w)}(\tau, h))/10$ in WLP.

The maximum evaluated frequency in the Lomb periodogram could be the Nyquist frequency, $\varpi = \pi/\Delta T_{\min}$, where $\Delta T_{\min}$ is the minimum sampling interval of the whole data. However, in WLP, because the minimum sampling interval of the windowed data varies with the evaluated time and bandwidth, the Nyquist frequency of each windowed data is also a varying function, i.e., $\varpi^{(w)}(\tau, h)$. If we select the overall Nyquist frequency as $\varpi = \max(\varpi^{(w)}(\tau, h))$, aliasing will occur at some evaluated time points, where the Nyquist frequency is smaller than the overall Nyquist frequency. On the other hand, if the overall Nyquist frequency is set as $\varpi = \min(\varpi^{(w)}(\tau, h))$, spectral information higher than ϖ cannot be identified at some evaluated time points. Therefore, in this study, we select the maximum evaluated frequency as the largest Nyquist frequency of windowed data, although there may be aliasing effects in the WLP. In practice, the maximum and minimum evaluated frequencies can also be determined according to known or interested frequency range.

C. Computational Complexity

We first consider the arithmetic complexity of the WLS estimator in (3). Since $\boldsymbol{\Phi}$ is an $N \times 2$ matrix, $\boldsymbol{X}$ is an $N \times 1$ vector, and $\boldsymbol{W}$ is an $N \times N$ diagonal matrix, the complexities of computing $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{W}X$ are $O\{4N\}$ and $O\{2N\}$, respectively. Thus, the complexity of the WLS estimator at each evaluated time–frequency point is $O\{N\}$. The computational complexity of the WLS algorithm can be further simplified. If the kernel has a limited support, such as the Epanechnikov window used in this paper, then the number of actual samples (denoted as N_w) included in the window is finite. Hence, the corresponding complexity of WLS will decrease considerably to $O\{N_w\}$. Since N_w increases with the bandwidth parameter h, a large bandwidth will increase the computational complexity. As for the regularized WLS estimators and M-estimators, since they are to be estimated in an iterative manner and the regularization parameter is tuned through GCV criterion, their complexity is much higher. In addition, since the WLP is computed at each evaluated time–frequency point, its overall complexity will be determined by the number of evaluated time–frequency points. Similarly, the number of bandwidths in the set $\boldsymbol{H}$ will also influence the complexity of AWLP.

Table I lists the arithmetic complexity required by various TFA methods proposed in this paper and other conventional methods. We can see that, compared with fast algorithms like STFT, the proposed methods do require higher complexity. However, STFT with fixed window size, in general, has a lower time–frequency resolution compared with AWLP, as demonstrated in the next section and the literature [21]. The window size for estimation has to be adaptively chosen, and hence, the arithmetic complexity will have to be increased. Therefore, there exists a performance-complexity tradeoff in selecting proper TFA methods. With the advent of microelectronics, low-cost DSPs, FPGA, and GPUs are now commonly available. Therefore, a high-performance spectral analysis method is, in general, preferred. Moreover, to facilitate implementation in DSP, FPGA, and GPUs, a highly modular algorithm with functionally identical elements is highly desirable.

In real-time applications, the complexity problem of the proposed methods can be addressed without much difficulty by three approaches.

1) Selecting an appropriate WLP method: In this paper, a class of WLP methods under a general framework is

TABLE I
COMPARISON OF ARITHMETIC COMPLEXITY OF VARIOUS TFA METHODS AT ONE EVALUATED TIME INSTANT

Method	Complexity	Remark
STFT	$O\{N \log N\}$ for N frequency points	N : the number of frequency points
WLP	$O\{N_w\}$ for one frequency point	N_w : the number of data samples in the calculating window
AWLP	$\sum_{j=1}^{J} O\{N_w^{(j)}\} + O\{N_w^{(+)}\}$ for one frequency point - the 1st item is the total complexity to calculate WLP with each bandwidth h_j in the bandwidth set $\boldsymbol{H}$ - the 2nd item is the complexity to calculate WLP with sub-optimal bandwidth h^+	$N_w^{(j)}$: the number of data samples in the calculating window with bandwidth h_j; $N_w^{(+)}$: the number of data samples in the calculating window with sub-optimal bandwidth h^+ (for simplicity, it is assumed to be identical at all frequency points); J : the number of candidate bandwidths in the bandwidth set $\boldsymbol{H}$
ME-RAWLP	$\mathrm{I} \cdot [\sum_{j=1}^{J} O\{N_w^{(j)}\} + O\{N_w^{(+)}\}]$ for one frequency point	I : the number of iterations (for simplicity, it is assumed to be identical at all frequency points and for all bandwidths)

developed to cater for the diverse requirements (different tradeoffs between performance and complexity) in practical situations. Therefore, one can select an appropriate WLP method to meet the desired tradeoff between performance and complexity. For example, in the application of voltage dip detection (see Section VI-C for details), where only the amplitude at one single frequency is to be estimated and thus regularization is not necessary, AWLP can be chosen so that the computational complexity can be lowered significantly because iterations involved in RAWLP are unnecessary. In the analysis of speech signals (see Section VI-D for details), where impulsive components may not be encountered, LS-based AWLP/RAWLP with a lower arithmetic complexity is sufficient to obtain good performance.

2) Using prior information to compute spectrum at the required frequency points: An advantage of the WLP-based methods is that a set of selected frequency points can be computed independently. Therefore, the WLP may be computed at a small set of frequency points of interest based on prior information of the signals to reduce the arithmetic complexity. For example, in the problem of voltage dip tracking, it is only necessary to estimate the spectrum at the fundamental frequency of 50 Hz (see Section VI-C for details).
3) Parallel implementation: The simplified linear model of the adaptive Lomb periodogram allows the spectral estimates at each time–frequency point to be computed separately, which greatly facilitates parallel implementation. Therefore, by exploring the inherent parallelism of the method in hardware and GPU implementation, the computational time can be significantly reduced. In particular, GPUs have emerged as a powerful and yet relatively inexpensive desktop parallel computing platform in recent years. They are high-performance many-core processors that are suitable for parallel computing in single instruction and multiple data. There are also many languages supporting programming in GPUs such as Nvidia's Compute Unified Device Architecture, Khronos Group's Open Computing Language (OpenCL), and Microsoft's DirectCompute. In Section VI-E, a parallel implementation of the proposed RAWLP method in GPU using OpenCL will be described to illustrate the inherent parallelism and other complexity-reducing methods mentioned earlier.

D. Selection of Window Type

A kernel with compact support is preferable because it can reduce the complexity significantly. In this paper, following the recommendation in [35], the following basis Epanechnikov window is employed:

$$w(u) = \begin{cases} \frac{3}{4}\left(1 - |u|^2\right), & |u| < 1 \\ 0, & |u| \geq 1. \end{cases} \tag{27}$$

Consequently, for an Epanechnikov window with bandwidth h, $w_h(u) = (1/h)w(u/h)$, only the data samples included in interval $(\tau - h, \tau + h)$ are used for local estimation, and thus, the effective size of the window is $2h$. Other types of windows, such as Gaussian window and Hanning window, can also be used in the WLP. Different types of windows have different frequency response characteristics and will influence the WLP results to some extent, but the selection of window type is beyond the scope of this paper.

For the purpose of online tracking, where only past data are available, one-sided window should be used. For example, a one-sided Epanechnikov window is given by

$$w(u) = \begin{cases} \frac{3}{4}\left(1 - |u|^2\right), & -1 < u \leq 0 \\ 0, & u \leq -1 \text{ or } u > 0. \end{cases} \tag{28}$$

In online implementation of the proposed WLP methods, other relevant operations making use of future data, like the smoothing of bandwidth in the time–frequency domain, should also be modified slightly. See [28] for details about the online implementation of local polynomial modeling, which can be used in the WLP problems as well.

E. Selection of Bandwidth Set

The next problem is how to select the bandwidth set $\boldsymbol{H}$ used in the empirical ICI bandwidth selection technique. The minimum bandwidth h_1 should be selected to make the WLS

TABLE II
PARAMETERS INVOLVED IN PROPOSED METHODS AND THEIR SELECTION APPROACHES

Method	Parameter	Selection approaches	Key references
AWLP	Bandwidth set $\boldsymbol{H}$	- Follow the guidance provided in Section V-E - Ad hoc selection based on practical applications (see the example of power quality analysis in Section V-E)	- Section IV-D-3 in [28] - Chapter 4.5.4 in [35]
	Threshold parameter Γ used in ICI	- Search the best value in a grid to minimize the SWSR (14) - Typical grid: {0.67 0.84 1.04 1.28 1.44 1.65 1.96 2.58 2.81 3.29} - Typical value: $\Gamma = 2.58$	- Section II in [24] - Section IV in [30] (a refinement)
ME-RAWLP - Regularization	Regularization parameters μ and $\widetilde{a}$	- Search the best value in a grid to minimize the GCV (27) - Typical grid for μ : 1.2×{0.01, 0.1, 1, 10 100} - Typical grid for $\widetilde{a}$: 3.7×{0.01, 0.1, 1, 10 100} - Typical value: $\mu = 1.2$, $\widetilde{a} = 3.7$	- Section 2.1 and Section 4.2 in [22]
ME-RAWLP - M-estimation	M-estimation threshold parameter ξ	- Typical values (for Huber function): {1.96, 2.24, or 2.58}, which are associated with {95, 97.5, or 99}% confidence to detect and reject the impulse.	- Section II in [36]

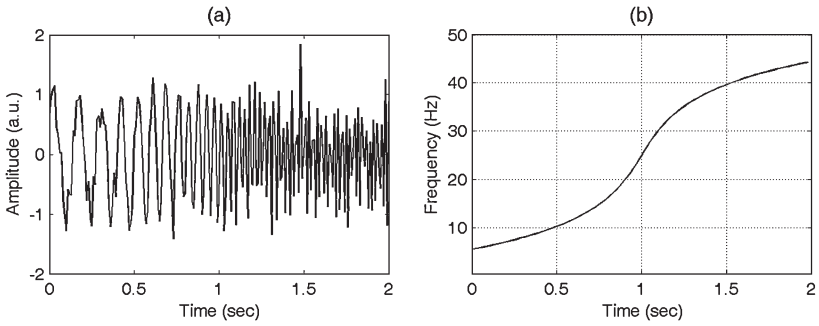


Fig. 2. One realization of the uniformly sampled data with one sinusoidal component. (a) Data samples. (b) Instantaneous frequency.

estimators solvable. More precisely, the number of data samples included in the interval $(\tau - h_1, \tau + h_1)$, N_w, should be equal to or larger than two. On the other hand, the largest element in the bandwidth set, h_J, that is generally selected can be large enough so that the corresponding local window includes all data points. However, as mentioned before, too long a window will result in high computational complexity. As for other bandwidths in the set $\boldsymbol{H}$, more bandwidth parameters will lead to more refined estimation results but will also increase the complexity. To achieve a good tradeoff between performance and complexity, we generally select 3–5 bandwidth parameters uniformly distributed in between the minimum bandwidth h_1 and the maximum bandwidth h_J. From simulation results, it was found that the proposed bandwidth setting gave satisfactory results.

In practical implementation, the bandwidth set can be determined based on known or interested time–frequency properties of the data under study or conventional setting in specific problems. For example, in power quality analysis, the bandwidth set can be selected as 1/4, 1/2, and 1 cycles, which follow the routine analysis windows in these problems.

Finally, to provide a clear and complete guidance for selecting parameters contained in this paper, a summary of the parameter selection procedures is given in Table II.

It can be seen from Table II that most parameters used in the proposed methods are automatically determined so that it is not difficult to be applied in practice. In fact, most of the parameters listed in Table II are selected automatically from a grid of candidate parameters. The grid of parameters should cover possible ranges of parameters to make sure the "optimal" parameter in the range defined by the grid. As a result, the parameter selection problem is reduced to how to determine a grid of parameters. For the threshold parameter Γ used in ICI, the values of the grid can be {0.67 0.84 1.04 1.28 1.44 1.65 1.96 2.58 2.81 3.29}, which are associated with confidence levels of {50.0 60.0 70.0 80.0 85.0 90.0 95.0 99.0 99.5 99.9}%. For regularization parameters μ, the grid can be $1.2 \times \{0.01, 0.1, 1, 10100\}$, where 1.2 is the typical value recommended in [22] for regularization parameters $\widetilde{a}$, and the grid could be $3.7 \times \{0.01, 0.1, 1, 10100]$, where 3.7 is the typical value recommended in [22]. It is possible to use a denser or wider grid, but it will also increase the complexity.

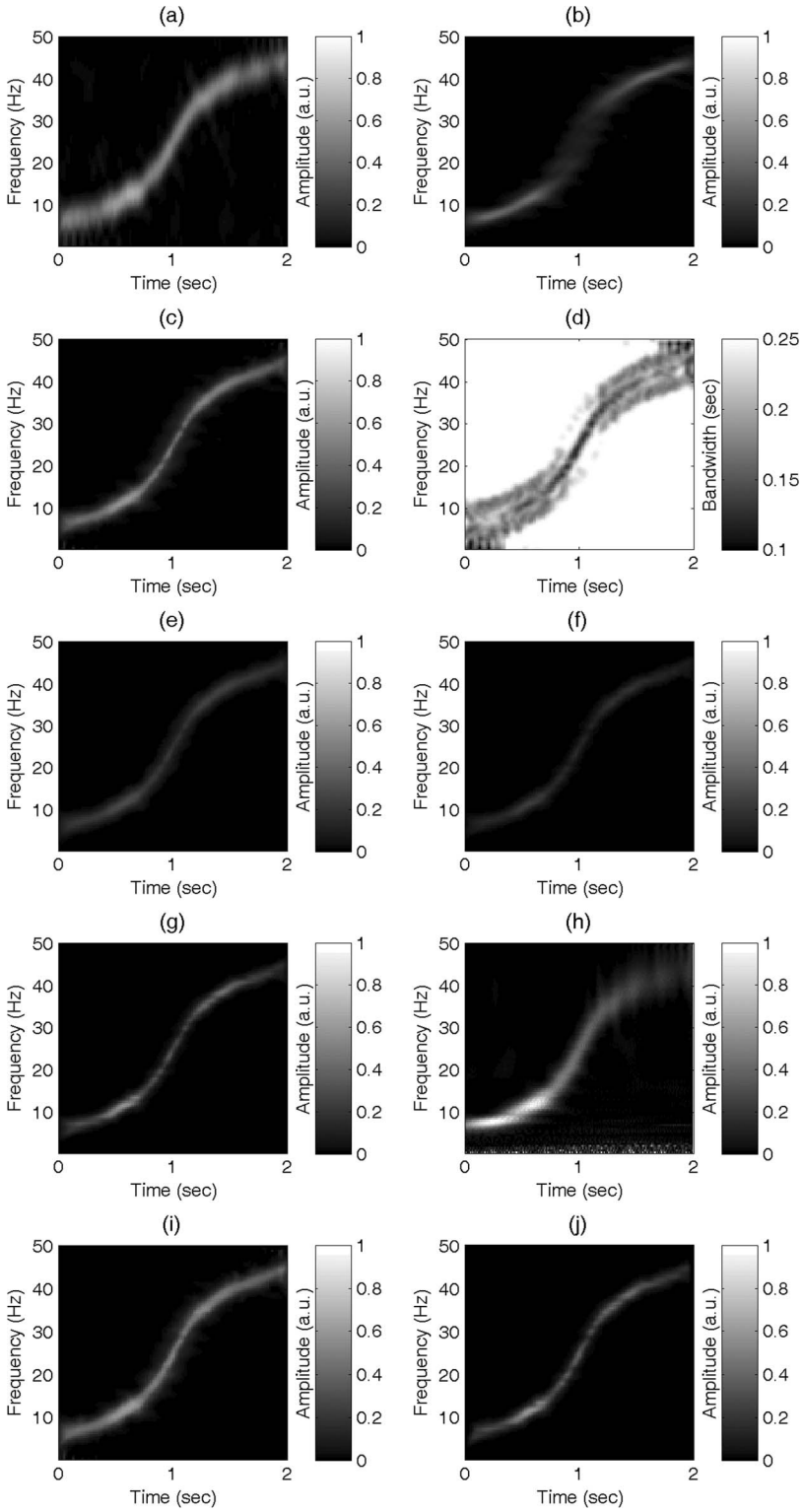


Fig. 3. Time–frequency representations of one example of uniformly sampled data in Fig. 2(a) using different TFA methods. (a) LS-WLP ($h = 0.1$ s). (b) LS-WLP ($h = 0.25$ s). (c) LS-AWLP (varying h). (d) ICI-selected h. (e) LS-Ridge-RAWLP (varying h). (f) LS-Lasso-RAWLP (varying h). (g) LS-SCAD-RAWLP (varying h). (h) WT. (i) ME-AWLP (varying h). (j) ME-SCAD-RAWLP (varying h).

Our simulations in this study and other studies show that these parameters can usually obtain satisfactory results in different applications. As to the bandwidth selection, the selection of a grid of bandwidth is more dependent on the application problems. If no prior knowledge is available, following the recommended setting for bandwidth selection can generally get good results. In Table II, the only one nonautomatically selected parameter is the M-estimation threshold parameter ξ because this parameter can offer flexibility for users to control the extent of impulsive noise suppression. If one wants to suppress noise to a larger extent, a smaller ξ should be selected, and vice versa [36]. Fortunately, the effect of ξ is easy to comprehend, and it is

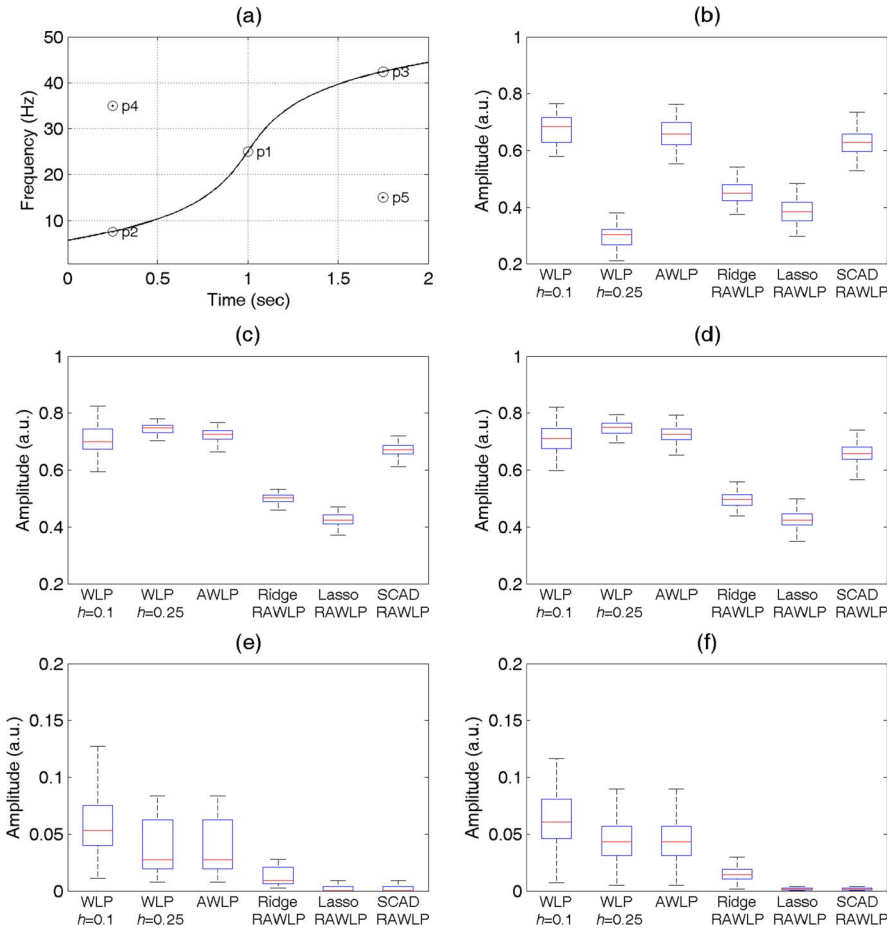


Fig. 4. Comparison of amplitude estimates at selected representative time–frequency points using different WLP-based TFA methods. The box has lines at the lower quartile, median, and upper quartile values. (a) Instantaneous frequency. (b) p1. (c) p2. (d) p3. (e) p4. (f) p5.

usually not difficult to determine this parameter to control the extent of impulsive noise suppression.

Finally, we remark that, except for the proposed bandwidth parameter section, the other parameters are selected according to the established and well-recognized techniques in literature. Following the guidelines as suggested in the literatures and the present paper, the window selection procedure is usually not sensitive to these parameters. For instances, good results in related applications, including data smoothing [26], image denoising [27], power quality analysis [28], biomedical data analysis [29], and speech data analysis (Section VI-D), have been obtained.

VI. Simulation Results

A. Uniformly Sampled Data (Single-Component Sinusoidal Signal)

We now compare various TFA methods using simulated uniform data sampled from a chirp signal

$$m(t) = A(t)\sin\left(\Omega(t) + \varphi\right) \tag{29}$$

where $A(t) = 1$ is the amplitude, $\varphi = \pi/3$ is the phase, and $\Omega(t) = 2\pi\int_0^t f(\upsilon)d\upsilon$ is the instantaneous phase calculated from an instantaneous frequency $f(t)$. Here, $f(t)$ is given as

$$f(t) = 25 + 7\text{asinh}\left(8(t-1)\right) \tag{30}$$

where $\text{asinh}(\cdot)$ is the inverse hyperbolic sine function. The data duration is 2 s, and the sampling frequency is 100 Hz, resulting in 200 data samples. An additive Gaussian white noise with a signal-to-noise ratio (SNR) of 10 dB is added. One realization of the uniformly sampled data is shown in Fig. 2(a), and the instantaneous frequency is shown in Fig. 2(b). We can see that this signal contains both slowly varying sinusoidal components (in the beginning and end periods) and fast-varying frequency components (in the middle period), so it can effectively illustrate the bandwidth selection results.

In the WLP, Epanechnikov windows are employed, and the bandwidth set used in the ICI technique is chosen as $\boldsymbol{H} = \{0.1, 0.15, 0.2, 0.25\}$ s. The evaluated time instants are set as the sampling time instants, and the evaluated frequency instants are from 0.5 to 50 Hz with a step of 0.5 Hz. Since STFT and WLP are equivalent for uniformly sampled real-valued data, we do not show the STFT results here.

The LS-based AWLP and RAWLP (accompanied with ridge, lasso, and SCAD regularization) are compared to LS-based WLP with a constant bandwidth. We also test the ME-based WLP methods and compare them with LS-based WLP methods. The WT with Morlet wavelet is also tested for comparison.

One representative simulation is shown in Fig. 3, where the Lomb-periodogram-based time–frequency representations are shown in the same scale (the maximum possible amplitude

TABLE III
SELECTED TYPICAL TIME–FREQUENCY POINTS: TIME–FREQUENCY FEATURES AND ADAPTIVE WINDOW SIZES SELECTED BY ICI (SNR = 10 dB)

Point	Location	Time-frequency features	Ideal Amplitude	h (unit: sec) mean±SD
p1	1.00 sec 25.00 Hz	A sinusoid with a fast-varying frequency	1	0.235±0.017
p2	0.25 sec 7.56 Hz	A sinusoid with a slowly-varying frequency (low)	1	0.301±0.018
p3	1.75 sec 42.44 Hz	A sinusoid with a slowly-varying frequency (high)	1	0.291±0.010
p4	0.25 sec 35.00 Hz	No significant time-frequency features	0	0.500±0.000
p5	1.75 sec 15.00 Hz	No significant time-frequency features	0	0.500±0.000

of these time–frequency representations is $\max(P(\tau,\omega)) = 1$). The following can be observed from Fig. 3.

1) As shown in Fig. 3(a), if a small bandwidth is used, the WLP can identify the fast-varying frequency components clearly, which implies a good time resolution. However, the spectral density of slowly varying components is dispersed along the frequency domain, leading to a bad frequency resolution. On the other hand, as shown in Fig. 3(b), if a large bandwidth is used, the frequency resolution for slowly varying bandwidth is increased. However, the fast-varying frequency component is blurred seriously due to a decreased time resolution.
2) The ICI technique can adaptively select the bandwidth based on the specific time–frequency features of the data. We can see from Fig. 3(d) that the bandwidths for fast-varying components are quite small while relatively large bandwidths are assigned to slowly varying components. For those time–frequency areas where there are no meaningful frequency components, the bandwidths are given the largest value to restrain the noise effects. With the adaptive varying bandwidth, the AWLP achieves a better time–frequency representation than WLP with a constant bandwidth. The AWLP also outperforms the result of WT, which has a bad time resolution for low-frequency components and a bad frequency resolution for high-frequency components.
3) Regularization techniques can further lower the estimation variance so that the three RWLS-based RAWLPs have better time–frequency resolution than the WLS-based AWLP. However, it can also be seen that the ridge and lasso estimators produce considerable bias to the large coefficients, so that the resultant ridge- and lasso-based RAWLPs have lower peak values than the WLS-based AWLP. The SCAD-based RAWLP has comparable peak amplitude values with the WLS-based AWLP because SCAD is an unbiased estimator.
4) When analyzing data without impulsive components, ME-based WLPs achieve almost the same results as corresponding LS-based WLPs. For brevity, only ME-AWLP and ME-SCAD-AWLP are shown in Fig. 3. Because M-estimation involves iterative operations and estimation of noise variance, it has a higher complexity than LS estimator. Therefore, when no impulsive components could occur in the data, the LS-based WLP methods are sufficient.

The simulation results are further compared in a quantitative manner. Since the AWLP and RWALP can select bandwidth adaptively according to the local time–frequency features of the data, we choose a set of time–frequency points with representative time–frequency features for evaluation. The locations of these time–frequency points are shown in Fig. 4(a), and their time–frequency features and true amplitude are listed in Table III. Their ICI-selected window sizes (mean and standard deviation), which are obtained from 100 Monte Carlo simulations, are also given in Table III.

To give a quantitative measure for comparison, we estimate the amplitude $\hat{A}(\tau,\omega)$ of sinusoids at these points. Box plots are used to depict the distribution patterns of the amplitude estimates, as shown in Fig. 4. To save space, the quantitative results of ME-based WLPs were not shown, and they are almost identical to the results of the corresponding LS-based WLPs. The statistical properties, including the bias, variance and MSE, of the estimated amplitudes can be approximated by means of Monte Carlo simulations. Suppose that we perform Q independent Monte Carlo realizations of the simulated sampled data, and denote the average of the estimates as $\overline{A}(\tau,\omega) = (1/Q)\sum_{q=1}^{Q}\hat{A}(\tau,\omega;q)$, where $\hat{A}(\tau,\omega;q)$ was the estimated coefficient of the qth realization. The bias, variance, and MSE are computed analogously as

$$Bias\left[\hat{A}(\tau,\omega)\right] = \overline{A}(\tau,\omega) - A(\tau,\omega) \tag{31}$$

$$Var\left[\hat{A}(\tau,\omega)\right] = \frac{1}{Q}\sum_{q=1}^{Q}\left[\hat{A}(\tau,\omega;q) - \overline{A}(\tau,\omega)\right]^2 \tag{32}$$

$$MSE\left[\hat{A}(\tau,\omega)\right] = \frac{1}{Q}\sum_{q=1}^{Q}\left[\hat{A}(\tau,\omega;q) - A(\tau,\omega)\right]^2. \tag{33}$$

TABLE IV
COMPARISONS OF BIAS, VARIANCE, AND MSE (UNIT: dB) BETWEEN DIFFERENT LS-WLP-BASED TFA METHODS (SNR = 10 dB)

Point	Methods	$Bias^2$	Variance	MSE
p1	WLP (h=0.10 sec)	-9.74	-24.91	-9.64
	WLP (h=0.25 sec)	-2.84	-25.11	-2.82
	AWLP	-8.87	-24.81	-8.78
	Ridge-RAWLP	-5.31	-27.78	-5.29
	Lasso-RAWLP	-4.29	-25.84	-4.27
	SCAD-RAWLP	-8.29	-26.36	-8.23
p2	WLP (h=0.10 sec)	-9.12	-27.32	-9.09
	WLP (h=0.25 sec)	-10.05	-29.25	-9.99
	AWLP	-9.97	-29.53	-9.94
	Ridge-RAWLP	-5.81	-31.47	-5.80
	Lasso-RAWLP	-4.62	-30.73	-4.61
	SCAD-RAWLP	-8.85	-31.22	-8.69
p3	WLP (h=0.10 sec)	-9.35	-26.57	-9.31
	WLP (h=0.25 sec)	-11.09	-28.21	-11.00
	AWLP	-10.23	-27.66	-10.15
	Ridge-RAWLP	-5.93	-29.23	-5.91
	Lasso-RAWLP	-4.65	-28.07	-4.63
	SCAD-RAWLP	-9.03	-28.63	-8.95
p4	WLP (h=0.10 sec)	-25.39	-31.82	-24.70
	WLP (h=0.25 sec)	-28.52	-34.60	-27.78
	AWLP	-28.52	-34.60	-27.78
	Ridge-RAWLP	-38.05	-44.12	-37.32
	Lasso-RAWLP	-51.97	-47.61	-47.13
	SCAD-RAWLP	-51.97	-47.61	-47.13
P5	WLP (h=0.10 sec)	-24.06	-27.10	-22.68
	WLP (h=0.25 sec)	-30.23	-34.60	-29.17
	AWLP	-30.23	-34.60	-29.17
	Ridge-RAWLP	-39.76	-44.14	-38.71
	Lasso-RAWLP	-59.96	-62.22	-58.36
	SCAD-RAWLP	-59.96	-62.22	-58.36

TABLE V
ADAPTIVE BANDWIDTH (MEAN ± SD; UNIT: SECONDS) SELECTED BY ICI IN LS-AWLP AT SELECTED TIME–FREQUENCY POINTS UNDER DIFFERENT SNRS

Point	SNR=0dB	SNR=5dB	SNR=10dB	SNR=20dB
p1	0.350±0.084	0.237±0.026	0.235±0.017	0.211±0.010
p2	0.366±0.075	0.348±0.037	0.301±0.018	0.265±0.007
p3	0.382±0.031	0.341±0.019	0.291±0.010	0.274±0.000
p4	0.497±0.005	0.500±0.000	0.500±0.000	0.500±0.000
p5	0.495±0.010	0.500±0.000	0.500±0.000	0.500±0.000

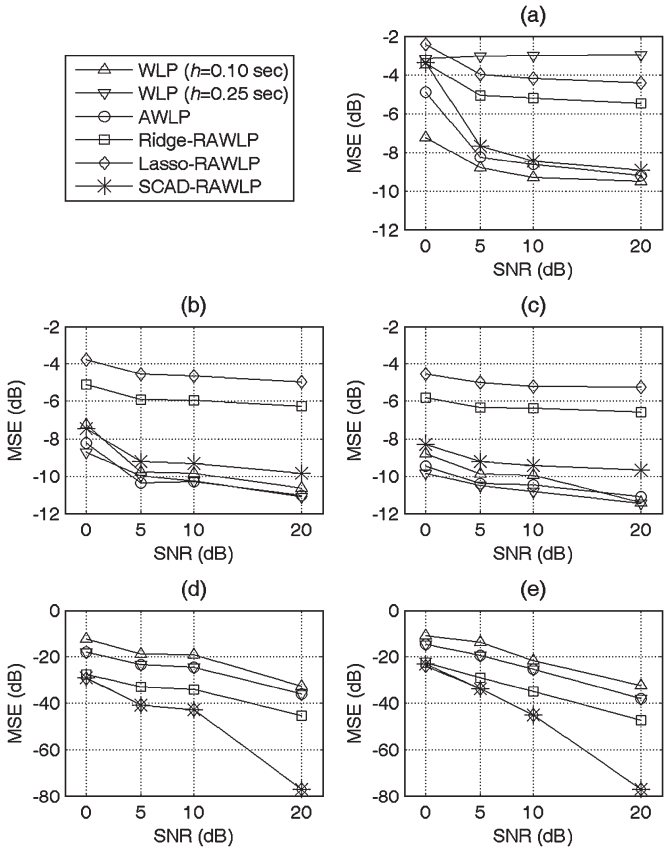


Fig. 5. MSE comparison of amplitude estimates using different LS-WLP-based TFA methods under different extents of noise. (a) p1. (b) p2. (c) p3. (d) p4. (e) p5.

The bias, variance, and MSE values approximated from 100 Monte Carlo simulations are listed in Table IV. The bias of amplitude estimates is induced by the time-varying properties of frequency components, and the variance is caused by finite-sample estimation of variance of local residual. The following can be concluded from Fig. 4 and Table IV.

1) A small bandwidth results in a large variance and a small bias, while a large bandwidth leads to a large bias and a small variance. The ICI-selected bandwidth can achieve a compromise between bias and variance.
2) Compared with the WLS-based AWLP, ridge-, lasso-, and SCAD-based RAWLPs all exhibit extra bias for large amplitudes (p1, p2, and p3). Among the three regularized estimators, lasso has the largest bias while SCAD has the smallest bias because SCAD is asymptotically unbiased for large coefficients.
3) As for the variance at p1, p2, and p3, RAWLPs usually have smaller variance than the WLS-based AWLP.
4) For small coefficients caused by noise (p4 and p5), lasso and SCAD can shrink them to zero, and thus, they have very small bias and variance. Ridge can also reduce the variance to some extent.

We now test the performances of the ICI technique used in AWLP under different extents of additive Gaussian noise. Zero-mean white Gaussian noises with different SNRs of 0, 5, 10, and 20 dB are added in the simulated signal model of (29). The suboptimal bandwidths adaptively selected by the ICI technique in different SNR conditions are listed in Table V.

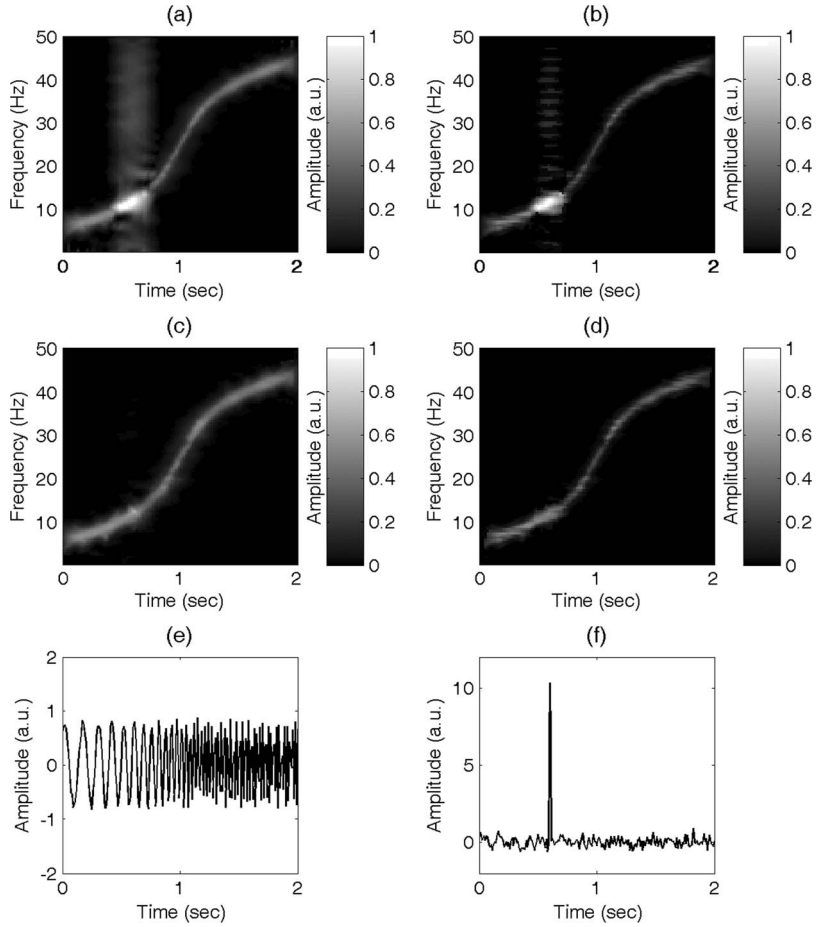


Fig. 6. Time–frequency representations of one example of uniformly sampled data containing an impulsive component at 0.6 s, and reconstructed sinusoidal and impulsive components from ME-SCAD-RAWLP. (a) LS-AWLP (varying h). (b) LS-SCAD-RAWLP (varying h). (c) ME-AWLP (varying h). (d) ME-SCAD-RAWLP (varying h). (e) Reconstructed sinusoid. (f) Reconstructed impulse.

We can conclude that, in general, the suboptimal bandwidth is decreased with the increase of SNR.

Next, we compare the MSE performances of the Lomb-periodogram-based TFA methods at different noise levels, which are shown graphically in Fig. 5. As expected, all the TFA methods under test have degraded performance with the decrease of SNR. By comparing the noise suppression results at points p4 and p5, the superiority of RWALP (particularly the lasso- and SCAD-based RWALPs) is more obvious when the SNR is higher.

Furthermore, we add an impulsive component with an amplitude of 10 to the data of Fig. 2(a) at 0.60 s, so as to compare the performances of LS-based and ME-based WLPs for data containing impulsive components. Fig. 6 clearly shows that the LS-based AWLP and SCAD-RAWLP time–frequency representations of the chirp signal are severely degraded around the time when the impulsive component occurs. On the contrary, the impulsive component has little influence on the ME-based AWLP and RAWLP. Moreover, the sinusoidallike chirp signal can be reconstructed from ME-SCAD-RAWLP, as shown in Fig. 6(e), and the impulsive component can be estimated from the difference between the reconstructed chirp signal and the original data, as shown in Fig. 6(f).

B. Nonuniformly Sampled Data

One major advantage of the proposed Lomb-periodogram-based TFA methods is that they can be employed for nonuniformly sampled data without any preprocessing (resampling) to the data. The nonuniform data are sampled from the chirp signal of (29) as well. One hundred samples are randomly selected from 200 uniformly distributed samples in 2 s, resulting in a mean sampling frequency of 50 Hz. In order to test conventional resampling-based TFA methods for the nonuniformly sampled data, a cubic spline interpolation is used to generate uniformly distributed data at a sampling frequency of 100 Hz. One realization of the nonuniformly sampled data and the resampled data are shown in Fig. 7. Their Lomb periodograms are also shown in Fig. 7, and we can see that the resampled data lose important spectral information over 25 Hz, which is half of the mean sampling frequency of the original nonuniformly sampled data.

Testing parameters used in the TFA methods are the same as those in the uniform cases. We also test the STFT on the interpolated data, and the STFTs employ the same window settings (Epanechnikov windows with bandwidths of {0. 1, 0. 15, 0. 2, 0. 25} s) as the WLP. For uniformly sampled real-valued data, the Lomb periodogram and the FFT-based periodogram have exactly the identical results, so do the WLP and

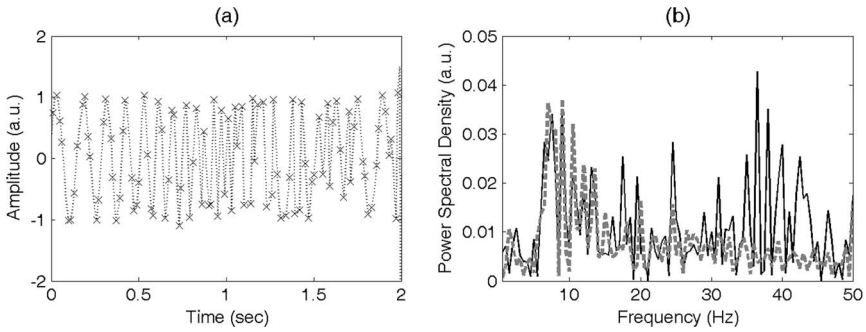


Fig. 7. One realization of the nonuniformly sampled data (denoted as cross) and their Lomb periodogram (solid line). The interpolated data with cubic spline are plotted as dashed line in the left panel, and their Lomb periodogram is plotted as dashed line in the right panel. (a) Data samples. (b) Lomb periodogram.

STFT-based spectrograms if they use the same windows. Here, we are interested to compare the performances of STFT and WLP on nonuniformly sampled data. Since the signal under test does not contain impulsive components, M-estimation-based methods have similar results to the LS-based methods, and they are not shown here. The time–frequency representations of one realization ($SNR = 10$ dB) are shown in Fig. 8. Quantitative analysis is also conducted, and numerical results are obtained for nonuniformly sampled data. Because the implication and interpretation of the results obtained from nonuniformly sampled data are similar to those from uniformly sampled data, detailed results are not presented here for brevity. We only summarize key observations.

First, it is obvious that the conventional resampling-based WT and STFT methods fail to reveal the time–frequency features over 25 Hz but induce considerable low-frequency artifacts, which shows the unique merit of the WLP-based TFA methods for nonuniformly sampled data. Second, time–frequency representations of nonuniformly sampled data exhibit larger variability than those of uniformly sampled data. This is due to the fact that, when data are nonuniformly sampled, it is quite possible that several windowed data segments have fewer samples than others and thus have large estimation variance. As a result, the WLP (particularly when the bandwidth is small) and AWLP in Fig. 8 are contaminated by considerable artifacts. Third, because of the relatively large variability existed in WLP of nonuniformly sampled data, regularization is particularly useful under this condition. The performance improvement by regularization, particularly by lasso and SCAD, is noticeable in Fig. 8.

C. Applications to Power Quality Analysis

We now evaluate the potential application of the proposed WLP-based TFA methods to power quality analysis. The first application is the detection of a voltage dip (i.e., a short duration of voltage decrease). Fig. 9(a) shows a simulated 50-Hz voltage waveform of a voltage dip event from 100 to 180 ms. A zero-mean Gaussian white noise with an SNR of 30 dB is added to the voltage waveform. Conventionally, time-dependent root-mean-square (RMS) values calculated from data in short-time windows are used to detect the dip event. Fig. 9(b) shows the RMS values computed using 1/8 cycle and one cycle (1 cycle $=$ 20 ms), and we can see that the window size has a considerable influence on the RMS results. We use WLP and AWLP methods to estimate the amplitude $A(\omega)$ of the voltage at 50 Hz. Fig. 9(c) shows the WLP results with sliding windows of 1/8 cycle and 1 cycle and the AWLP results using window sizes of [1/8, 1/4, 1/2, 1] cycles. One-sided windows are used in WLP and AWLP for the purpose of online tracking. We can see that a short window can detect the rise and drop of the voltages promptly but also has a large variability (as shown from the WLP result with 1/8 cycle). On the other hand, a long window has a much smoother result but leads to a longer detection lag (as shown from the WLP result with 1 cycle). The proposed AWLP, which uses adaptive window sizes, detects the sharp changes in the voltage dig quickly while achieving smooth estimates when the voltage amplitude is stable.

A real voltage dip is shown in Fig. 10. The voltage dip signal was recorded from phase A of a 220-kV power line of a 500-kV station of Guangdong Power Grid, China, on September 2, 2009, by a ZH-2X series oscilloscope provided by Wuhan Zhongyuan Huadian Science & Technology Co. Ltd., China. The power line frequency is 50 Hz, and the sampling rate of the oscilloscope is 2000 Hz. We compare the results from AWLP and the RMS (1/4 cycle and1/2 cycle). The duration of the dip is determined as the period when the voltage amplitude is smaller than 90% of the normal values. We can see that the AWLP can detect the starting of end points of the voltage dip faster (around 4.5 ms) than the 1/2 cycle RMS, which is very useful for further handling the dip.

In the second example, we applied the ME-RAWLP method in the analysis of power transients. There are generally two types of power transients: impulsive and oscillatory transients [34]. An impulsive transient has a very short duration (typically $<$ 1 ms) and a large amplitude, and thus, it is more suitable to be presented in the time domain. An oscillatory transient usually has a primary frequency (up to 5 M Hz) and a damping amplitude lasting several cycles, so a time–frequency representation is suitable to describe such oscillatory transients. Fig. 11(a) shows a 50-Hz power waveform contaminated with an impulsive transient at 40 ms and a 300-Hz oscillatory transient from 95 to 125 ms. A zero-mean Gaussian noise with an SNR of 20 dB is also added to the power waveform. A high-pass filtering at 100 Hz is first used to separate the 50-Hz component and the high-frequency oscillatory transients for subsequent independent time–frequency analyses. We use the proposed ME-RAWLP method with SCAD

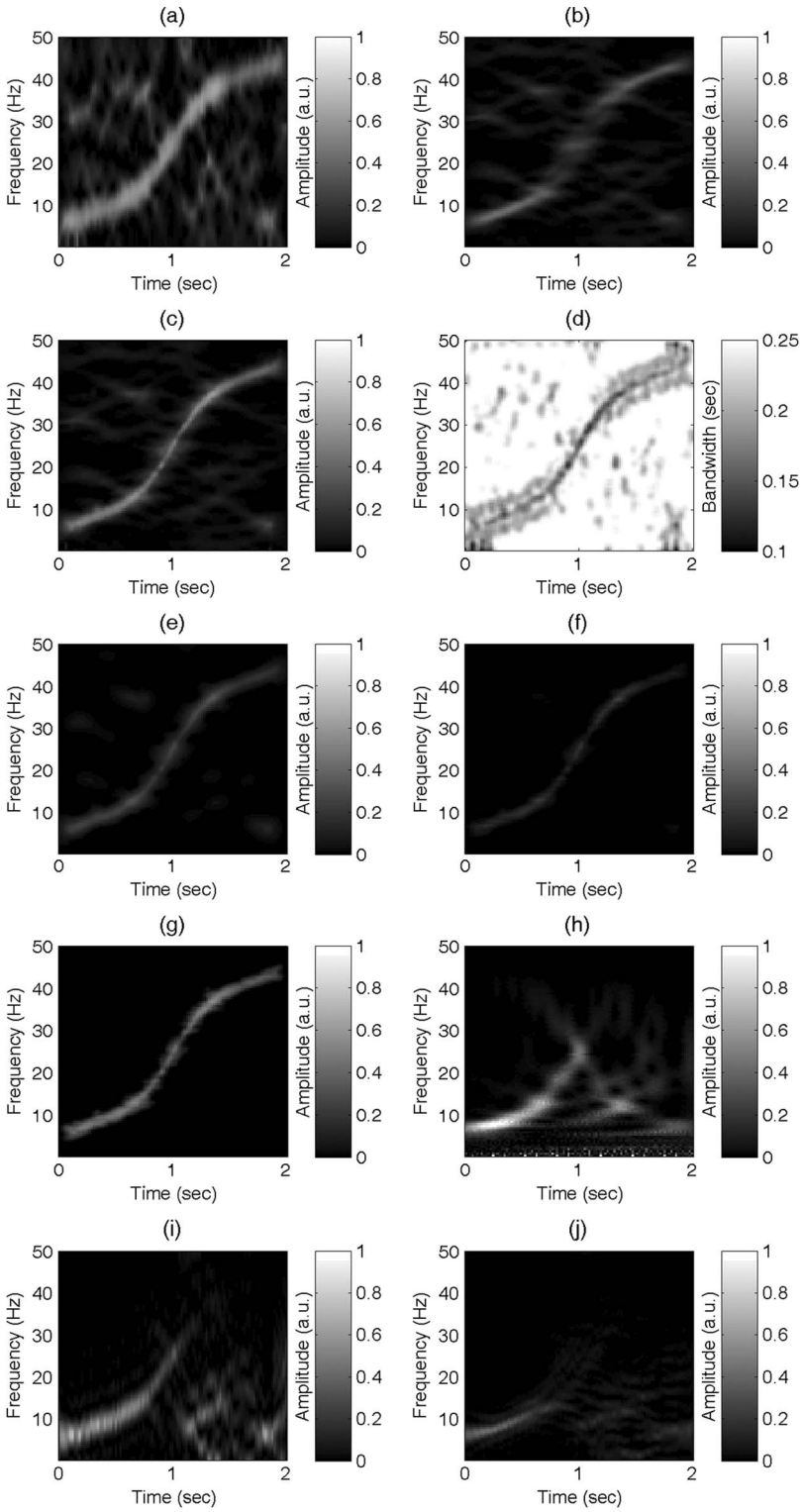


Fig. 8. Time–frequency representations of one example of nonuniformly sampled data in Fig. 7(a) using different TFA methods. (a) LS-WLP ($h = 0.1$ s). (b) LS-WLP ($h = 0.25$ s). (c) LS-AWLP (varying h). (d) ICI-selected h. (e) LS-Ridge-RAWLP (varying h). (f) LS-Lasso-RAWLP (varying h). (g) LS-SCAD-RAWLP (varying h). (h) WT. (i) STFT ($h = 0.1$ s). (j) STFT ($h = 0.25$ s).

regularization to provide a time-domain representation for the impulsive transient and a time–frequency representation for the oscillatory transient and the fundamental 50-Hz component. Meanwhile, the waveforms of the fundamental 50-Hz component and the oscillatory transient are reconstructed from the ME-RAWLP.

Fig. 11(e) shows that the LS-based AWLP has two defects: 1) It is considerably contaminated by the impulsive transient in a wide range of frequency, and 2) it cannot achieve a focused and sparse representation for the oscillatory transient in the time–frequency domain. These two problems can be well addressed by the ME-RAWLP method with SCAD regularization.

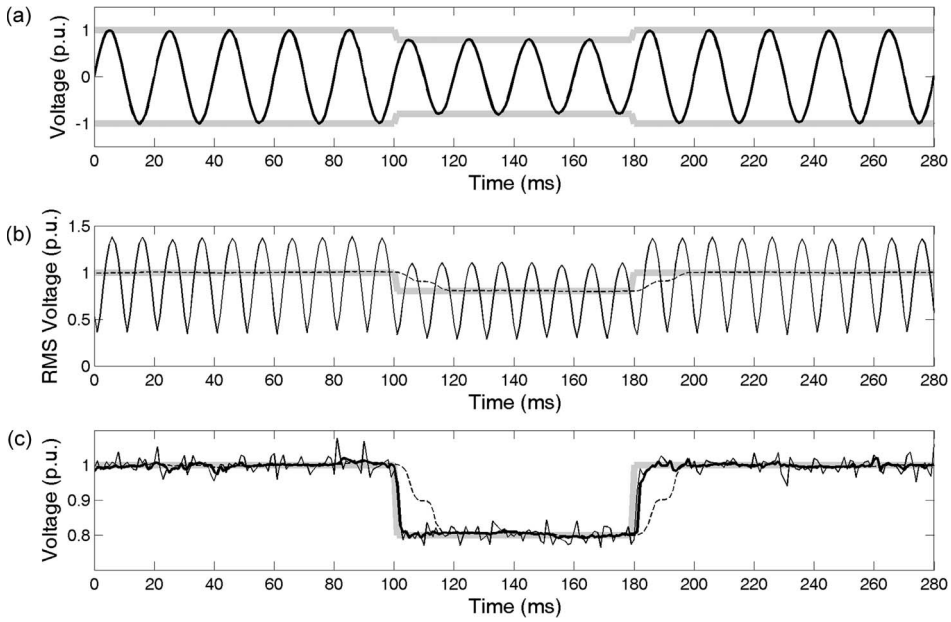


Fig. 9. Detection of a simulated voltage dip. (a) Simulated 50-Hz voltage waveforms with a dip between 100 and 180 ms. (b) RMS obtained using sliding windows of 1/8 cycle and 1 cycle (bold gray line: true voltage amplitude; thin dashed black line: RMS obtained using 1/8-cycle window; and thin solid black line: RMS obtained using 1-cycle window). (c) Voltage amplitudes estimated using WLP and AWLP (bold gray line: true voltage amplitude; thin dashed black line: amplitude estimated from WLP with 1/8-cycle window; thin solid black line: amplitude estimated from WLP with 1-cycle window; and bold solid black line: amplitude estimated from AWLP with adaptive windows).

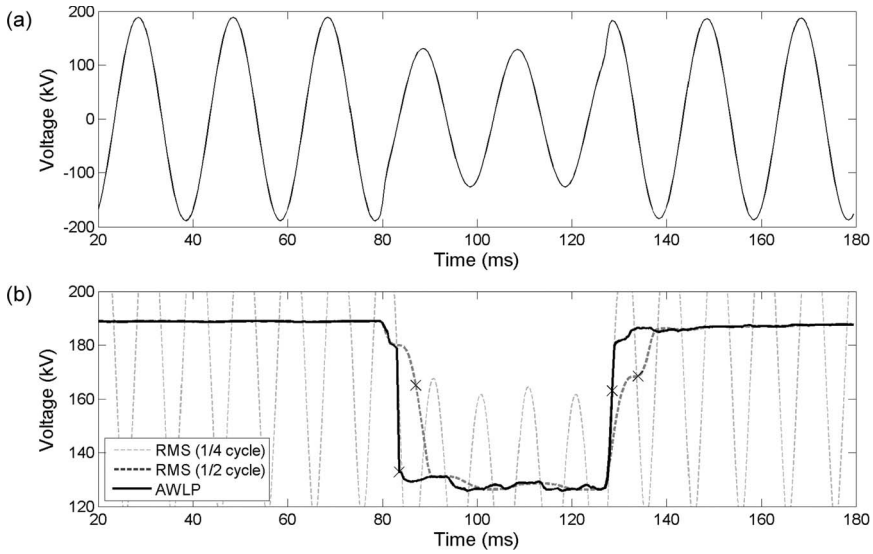


Fig. 10. Detection of a real voltage dip. (a) Voltage waveform with a dip between around 80 and 127 ms. (b) RMS obtained using sliding windows of 1/4 cycle (dashed thin line) and 1/2 cycle (dashed bold line) and amplitude estimated using AWLP (solid bold line); the cross signs denote the starting and end points of the dip (amplitude $< 90\%$ of the normal values; only for AWLP and RMS with 1/2 cycle). (b) Duration of the dip detected by AWLP is from 82.5 to 128.5 ms, the duration of the dip detected by RMS with 1/2 cycle is from 87 to 134 ms, and it is impossible to detect the duration of the dip from RMS with 1/4 cycle due to its huge variability.

As shown in Fig. 11(f), the impulsive transient is satisfactorily restrained, and the oscillatory transient has a distinct representation with very high time–frequency resolution. The reconstructed 50-Hz fundamental component and the 300-Hz oscillatory transient are shown in Fig. 11(b) and (c), respectively. The impulsive transient shown in Fig. 11(d) is estimated by subtracting the reconstructed 50-Hz fundamental component and the 300-Hz oscillatory transient from the original power waveform. We can see from Fig. 11(b)–(d) that the three components can be well separated via the ME-RAWLP method, providing accurate and complete data characteristics for power monitoring.

D. Applications to Speech Signal Analysis

In this experiment, a segment of speech signal is used for comparing the performances of different TFA methods when dealing with multicomponent sinusoidal signals. Speech signals are known to contain multiple sinusoidal components (fundamental and harmonics), so they are good candidates for

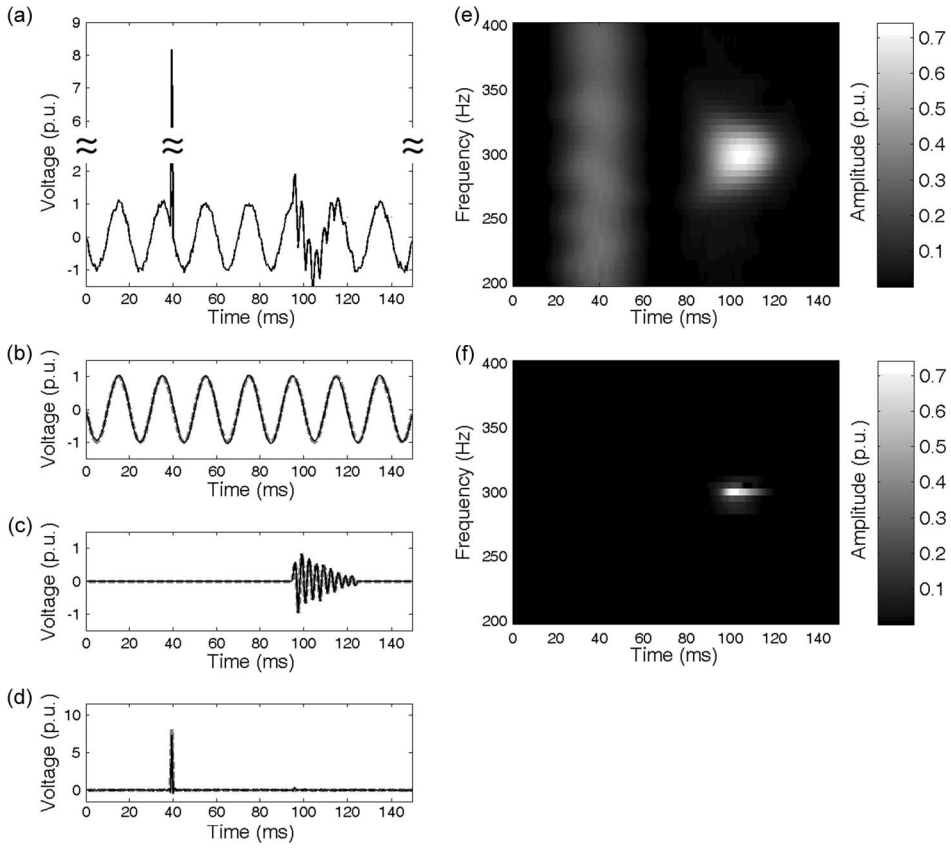


Fig. 11. Time–frequency representations and separation of impulsive and transient power transients. (a) Simulated 50-Hz power waveforms with an impulsive transient at 40 ms and an oscillatory transient from 95 to 125 ms. (b) True 50-Hz voltage component (bold gray dashed line) and its estimate reconstructed from ME-RAWLP (black solid line). (c) True oscillatory transient (bold gray dashed line) and its estimate reconstructed from ME-RAWLP (black solid line). (d) True impulsive transient (bold gray dashed line) and its estimate (black solid line). (e) LS-AWLP. (f) ME-RAWLP with SCAD regularization.

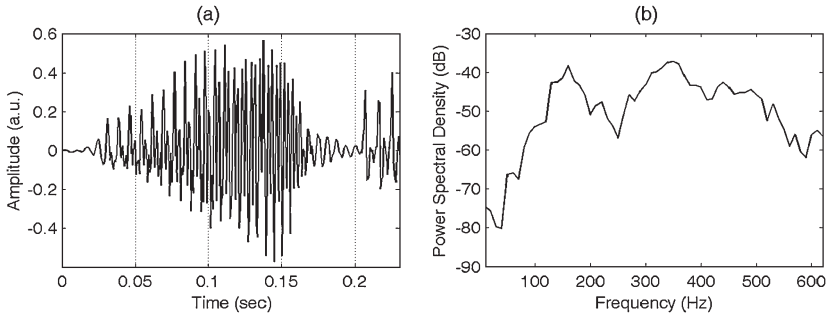


Fig. 12. Speech signal consisting of two phonemes "tea" and "o" ("o" starts from 0.2 s) and its Lomb periodogram. (a) Data samples. (b) Lomb periodogram.

evaluating the performances of TFA methods. Conventionally, STFT and WT are used for disclosing time-varying frequency components of speech signals, but they still suffer from the problem of selection of a fixed window [37]. Here, we will show that the proposed AWLP/RAWLP method is a good alternative to address the window selection problem and can achieve better time–frequency representations for multiple sinusoidal components in the real data to be analyzed hereinafter. The speech signal under test contains two phonemes, namely, "tea" and "o" (Fig. 12), and it is downloaded from http://authors.phptr.com/quatieri/audio.html. The duration of the speech is around 0.23 s, and the second phoneme "o" starts from 0.2 s. The sampling rate of the speech is 1250 Hz. In the AWLP method, Epanechnikov windows are employed, and the bandwidth set is chosen as $\boldsymbol{H} = \{0.01, 0.015, 0.02, 0.025\}$ s. The evaluated time instants are set as the sampling time instants, and the evaluated frequency instants are from 5 to 625 Hz with a step of 5 Hz.

We can see from Fig. 13 that the proposed AWLP and RAWLP methods can obtain good results when analyzing the multicomponent speech signal. First, the adaptive windows selected by the ICI technique are small around 0.2 s so that the two phonemes can be well separated. In addition, the adaptive windows selected make clear the onset and offset time of these frequency components. Second, when SCAD regularization is used in RAWLP, the frequency components are more pronounced and distinct because small coefficients are set to be zero by the SCAD regularization.

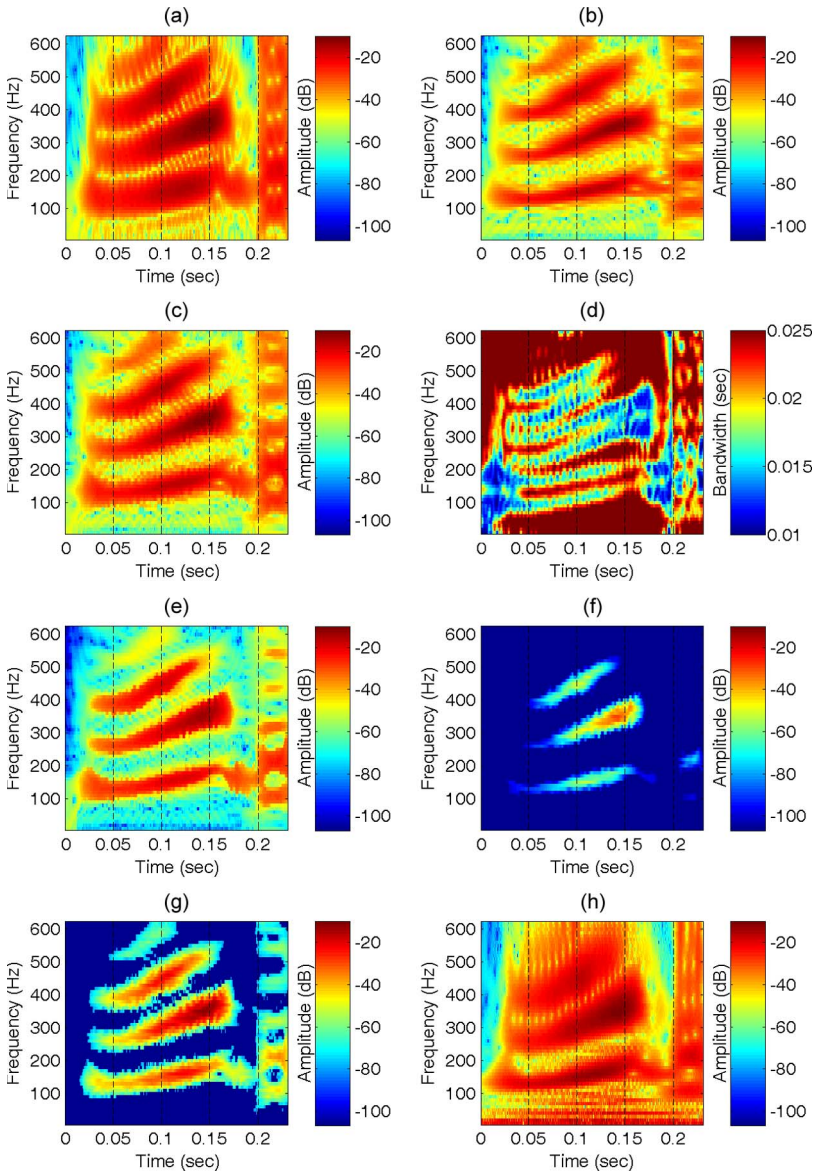


Fig. 13. Time–frequency representations of the speech signal in Fig. 12(a) using different TFA methods. The amplitudes of time–frequency representations are shown in logarithmical scale. (a) LS-WLP ($h = 0.01$ s). (b) LS-WLP ($h = 0.025$ s). (c) LS-AWLP (varying h). (d) ICI-selected h. (e) LS-Ridge-RAWLP (varying h). (f) LS-Lasso-RAWLP (varying h). (g) LS-SCAD-RAWLP (varying h). (h) WT.

E. Parallel Implementation Using GPU

As mentioned before, the proposed method is performed at each time and frequency point in parallel. Therefore, the RAWLP method is divided into a lot of subtasks according to the time and frequency, which will be executed on GPUs in parallel. These subtasks are called work items in OpenCL, the number of which is equal to the total number of time–frequency points. To achieve a better performance, the exchange of intermediate data is limited on the GPU. In the experiments, two PCs with different configurations are used. The first PC has an Intel i7-920 CPU and an AMD Radeon HD 6950 GPU. The other one has an Intel i7-990X CPU and a Nvidia GTX 580 GPU. The parameter selection follows the suggestions in Table II. The time cost of the GPU implementation is summarized in Fig. 14. It can be seen that if the number of time–frequency points computed per second is 1.0×10^5, it only requires about 95 ms. In other words, if the number of frequency points is 1024, we can compute about 98 spectra in about 95 ms. This means that we can compute a spectrum of 1024 points with around 1028 spectra per second. This is very satisfactory for many real-time signal processing applications such as vibration analysis and speech analysis. If the number of frequency points is reduced to 512, the number of spectra that can be computed per second will be increased to around 2056. This shows the flexibility of the proposed method in achieving different tradeoffs between time resolution of the spectrum and the number of frequency points computed. Moreover, the evaluated frequency points can be nonuniformly spaced and arbitrarily selected. It can also be seen from Fig. 14 that the computational time increases quite linearly with the number of work items, which suggests that the achieved speed-up is quite constant for different numbers of work items.

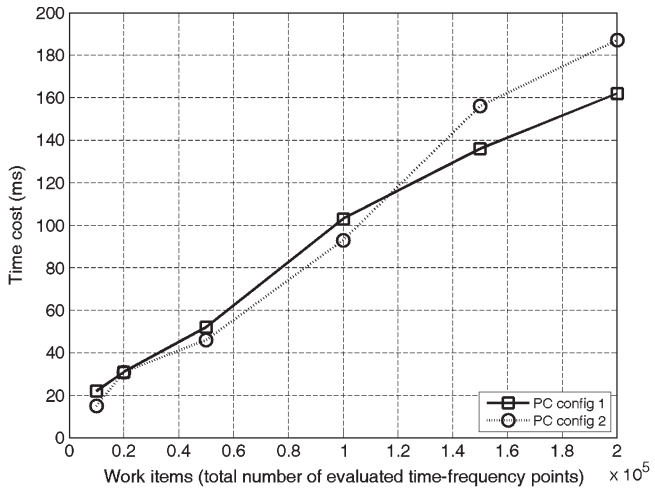


Fig. 14. Time cost of GPU implementation of the RAWLP method.

Overall, we can summarize the properties and advantages of the proposed WLP methods as follows.

1) The WLS-based AWLP can achieve a good tradeoff between time resolution and frequency resolution according to local time–frequency characteristics. Thus, it can be used for TFA of general data sequences.
2) The ridge-based RAWLP can effectively reduce the variability of AWLP at the expense of extra bias. If it is not required to correctly estimate the amplitude of sinusoids but only required to indicate the location of energy concentration in the time–frequency domain, ridge-based RAWLP can be a good choice.
3) The lasso-based RAWLP will lead to a sparse time–frequency representation, and so, it is more suitable for sinusoid data. In addition, it will induce a large bias term to the amplitude estimate. Thus, compared with the SCAD-based RAWLP, it does not have evident advantages.
4) The SCAD-based RAWLP can lead to a sparse time–frequency representation as well, but it has a much smaller estimation bias than lasso-based RAWLP. It is therefore more preferred to analyze sinusoid data, particularly when SNR is high.
5) M-estimation-based RAWLP can effectively restrain the adverse effects of impulsive components in the time–frequency representation. Although ME-based WLP methods are, in particular, useful in dealing with data contaminated with outliers, they can also achieve similar performances to the LS-based WLP methods when analyzing data containing no impulsive noise.
6) The Lomb periodogram and the ICI technique are most suitable for data with Gaussian-distributed residuals, and thus, the AWLP-based methods are very suitable for single-component sinusoidal data. For multiple-component sinusoidal data and nonsinusoidal data, the performance of AWLP-based method may be affected to some extent, but in general, they can still provide valuable information of the signal spectrum.
7) Parallel implementation of proposed methods using GPU can be easily achieved to considerably reduce the executional time of the WLP-based methods which facilitates real-time applications.

VII. Conclusion and Discussion

A new class of WLP methods for TFA of nonstationary signals possibly with impulsive components has been presented. The conventional Lomb periodogram is extended to a TFA method through WLS fitting and adaptive bandwidth selection using the ICI technique, which adapts to the local time–frequency characteristics of nonstationary signals to be analyzed. Regularized M-estimation method is further adopted instead of LS estimation to improve the performance of AWLP by reducing the variance, enhancing sparsity, and restraining the adverse effect of impulsive components. Simulation results show the advantages of the proposed method over the conventional Lomb periodogram or other conventional TFA methods in adaptive time–frequency resolution, sparse representation for sinusoids, robustness to impulsive components, applicability to nonuniformly sampled data, and fast computation time by using parallel computing. Its potential applications to power quality monitoring and analysis are also demonstrated.

In the future, it is desirable and possible to set up a more general linear model involving a dense set of sinusoids (instead of one sine and one cosine in the proposed method) and rely on regularization to estimate the model coefficients and the spectrum. In principle, each frequency will have its own adaptive window, which significantly complicates the selection of all multiple window parameters in the general linear model. The proposed method adopted a simplified linear model of one single sinusoid, and thus, it is relatively easy to estimate the adaptive window at each frequency independently. The adaptive bandwidth selection in the general linear model requires further investigation and will be left for future work. In addition, we may also consider employing other LS techniques such as total LS to address the error-in-variable problem and recursive LS for reducing the arithmetic complexity.

Acknowledgment

The authors would like to thank Dr. Z. Du and Dr. Y. Liang, who are with the School of Electric Power, South China University of Technology, for providing the real voltage dip data.

References

[1] T. Radil, P. M. Ramos, F. M. Janeiro, and A. C. Serra, "PQ monitoring system for real-time detection and classification of disturbances in a single-phase power system," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 8, pp. 1725–1733, Aug. 2008.
[2] Y. H. Gu and M. H. J. Bollen, "Time-frequency and time-scale domain analysis of voltage disturbances," *IEEE Trans. Power Deliv.*, vol. 15, no. 4, pp. 1279–1284, Oct. 2000.
[3] M. H. J. Bollen, I. Y. H. Gu, S. Santoso, M. F. Mcgranaghan, P. A. Crossley, M. V. Ribeiro, and P. F. Ribeiro, "Bridging the gap between signal and power," *IEEE Signal Process. Mag.*, vol. 26, no. 4, pp. 12–31, Jul. 2009.
[4] D. Granados-Lieberman, R. J. Romero-Troncoso, R. A. Osornio-Rios, A. Garcia-Perez, and E. Cabal-Yepez, "Techniques and methodologies for power quality analysis and disturbances classification in power systems: A review," *IET Gener. Transm. Distrib.*, vol. 5, no. 4, pp. 519–529, Apr. 2011.

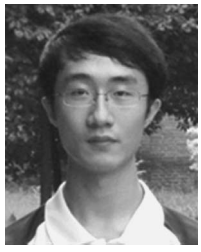
[5] M. Pineda-Sanchez, M. Riera-Guasp, J. A. Antonino-Daviu, J. Roger-Folch, J. Perez-Cruz, and R. Puche-Panadero, "Diagnosis of induction motor faults in the fractional Fourier domain," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 8, pp. 2065–2075, Aug. 2010.
[6] A. Chatterjee, R. Fournier, A. Naït-Ali, and P. Siarry, "A postural information-based biometric authentication system employing S-transform, radial basis function network, and extended Kalman filtering," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 12, pp. 3131–3138, Dec. 2010.
[7] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York: Springer-Verlag, 1987.
[8] M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1999.
[9] E. Sejdic, I. Djurovic, and J. Jiang, "Time-frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Process.*, vol. 19, no. 1, pp. 153–183, Jan. 2009.
[10] B. Boashash, *Time Frequency Signal Analysis: Methods and Applications*. Melbourne, Australia: Longman Cheshire, 1992.
[11] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation: Optimal kernel design," *IEEE Trans. Signal Process.*, vol. 41, no. 4, pp. 1589–1602, Apr. 1993.
[12] S. Mallat, *A Wavelet Tour of Signal Processing*. Burlington, MA: Academic, 1998.
[13] F. Marvasti, *Nonuniform Sampling, Theory and Practice*. New York: Kluwer, 2001.
[14] P. Babu and P. Stoica, "Spectral analysis of non-uniformly sampled data—A review," *Digital Signal Process.*, vol. 20, no. 2, pp. 359–378, Mar. 2010.
[15] P. Stoica, J. Li, and H. He, "Spectral analysis of non-uniformly sampled data: A new approach versus the periodogram," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 843–858, Mar. 2009.
[16] P. Laguna, G. B. Moody, and R. G. Mark, "Power spectral density of unevenly sampled data by least-square analysis: Performance and application to heart rate signals," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 6, pp. 698–715, Jun. 1998.
[17] P. M. T. Broersen, "Practical aspects of the spectral analysis of irregularly sampled data with time-series models," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 5, pp. 1380–1388, May 2009.
[18] E. F. Glynn, J. Chen, and A. R. Mushegian, "Detecting periodic patterns in unevenly spaced gene expression time series using Lomb-Scargle periodograms," *Bioinformatics*, vol. 22, no. 3, pp. 310–316, Feb. 2006.
[19] A. Holland and M. Aboy, "A novel recursive Fourier transform for non-uniform sampled signals: Application to heart rate variability spectrum estimation," *Med. Biol. Eng. Comput.*, vol. 47, no. 7, pp. 697–707, Jul. 2009.
[20] N. R. Lomb, "Least-squares frequency analysis of unequally spaced data," *Astrophys. Space Sci.*, vol. 39, pp. 447–462, Feb. 1976.
[21] Z. G. Zhang and S. C. Chan, "Robust adaptive Lomb periodogram for time-frequency analysis of signals with sinusoidal and transient components," in *Proc. IEEE ICASSP*, Philadelphia, PA, Mar. 18–23, 2005, vol. 4, pp. iv/493–iv/496.
[22] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Am. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001.
[23] S. C. Chan, Z. G. Zhang, and Y. J. Chu, "A new transform-domain regularized recursive least M-estimate algorithm for robust linear estimation," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 58, no. 2, pp. 120–124, Feb. 2011.
[24] V. Katkovnik, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2567–2571, Sep. 1999.
[25] A. Goldenshluger and A. Nemirovski, "On spatial adaptive estimation of nonparametric regression," *Math. Methods Stat.*, vol. 6, no. 2, pp. 135–170, 1997.
[26] Z. G. Zhang, S. C. Chan, K. L. Ho, and K. C. Ho, "On bandwidth selection in local polynomial regression analysis and its application to multi-resolution analysis of non-uniform data," *J. Signal Process. Syst.*, vol. 52, no. 3, pp. 263–280, Sep. 2008.
[27] Z. G. Zhang, S. C. Chan, and Z. Y. Zhu, "A new two-stage method for restoration of images corrupted by Gaussian and impulse noises using local polynomial regression and edge preserving regularization," in *Proc. IEEE ISCAS*, May 24–27, 2009, pp. 948–951.
[28] S. C. Chan and Z. G. Zhang, "Local polynomial modeling and variable bandwidth selection for time-varying linear systems," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 3, pp. 1102–1117, Mar. 2011.
[29] Z. G. Zhang, Y. S. Hung, and S. C. Chan, "Local polynomial modelling of time-varying autoregressive models with application to time-frequency analysis of event-related EEG," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 557–566, Mar. 2011.
[30] L. Stanković, "Performance analysis of the adaptive algorithm for bias-to-variance tradeoff," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1228–1234, May 2004.
[31] V. Katkovnik and V. Spokoiny, "Spatially adaptive estimation via fitted local likelihood techniques," *IEEE Trans. Signal Process.*, vol. 56, no. 3, pp. 873–886, Mar. 2008.
[32] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Stat. Soc B.*, vol. 58, no. 1, pp. 267–288, 1996.
[33] V. Katkovnik, I. Djurović, and L. Stanković, "Robust time-frequency distributions," in *Time–Frequency Signal Analysis and Processing: A Comprehensive Reference*, B. Boashash, Ed. London, U.K.: Elsevier, 2003, ch. ch.0.3, pp. 392–399.
[34] *IEEE Recommended Practice for Monitoring Electrical Power Quality*, IEEE Std. 1159-1995, Jun. 1995.
[35] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. London, U.K.: Chapman & Hall, 1996.
[36] S. C. Chan and Y. X. Zou, "A recursive least M-estimate algorithm for robust adaptive filtering in impulse noise: Fast algorithm and convergence performance analysis," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 975–991, Apr. 2004.
[37] T. F. Quatieri, *Discrete-time Speech Signal Processing: Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall, 2002, p. 542.

Zhiguo Zhang (S'05–M'07) received the B.Sc. degree in electrical and electronic engineering from Tianjin University, Tianjin, China, in 2000, the M.Eng. degree in electrical and electronic engineering from the University of Science and Technology of China, Hefei, China, in 2003, and the Ph.D. degree from the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, in 2008.

He is currently a Research Assistant Professor with The University of Hong Kong. His research interests include biomedical signal processing, neural engineering, and computational neuroscience.

Shing-Chow Chan (S'87–M'92) received the B.Sc. (Eng.) and Ph.D. degrees from the University of Hong Kong, Pokfulam, Hong Kong, in 1986 and 1992, respectively.

He joined the City Polytechnic of Hong Kong, Kowloon, Hong Kong, in 1990, as an Assistant Lecturer and later as a University Lecturer. Since 1994, he has been with the Department of Electrical and Electronic Engineering, The University of Hong Kong, where he is currently a Professor. He was a Visiting Researcher with Microsoft Corporation, Redmond, WA, Microsoft, Beijing, China, University of Texas at Arlington, Arlington, and Nanyang Technological University, Singapore. His research interests include fast transform algorithms, filter design and realization, multirate and biomedical signal processing, communications and array signal processing, high-speed AD converter architecture, bioinformatics, and image-based rendering.



Chong Wang received the B.Eng. degree from Zhejiang University of Technology, Hangzhou, China, in 2007 and the M.Eng. degree from the University of Science and Technology of China, Hefei, China, in 2010. He is currently working toward the Ph.D. degree in the Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong.

His main research interests include digital signal processing, image and video restoration, and parallel computing.