



Title	Recursive Parametric Frequency/Spectrum Estimation for Nonstationary Signals With Impulsive Components Using Variable Forgetting Factor
Author(s)	Zhang, ZG; Chan, SC
Citation	IEEE Transactions on Instrumentation and Measurement, 2013, v. 62 n. 12, p. 3251-3264
Issued Date	2013
URL	http://hdl.handle.net/10722/202894
Rights	IEEE Transactions on Instrumentation and Measurement. Copyright © Institute of Electrical and Electronics Engineers.

Recursive Parametric Frequency/Spectrum Estimation for Nonstationary Signals With Impulsive Components Using Variable Forgetting Factor

Zhi Guo Zhang, *Member, IEEE*, and Shing-Chow Chan, *Member, IEEE*

Abstract—This paper proposes a general and computationally efficient parametric model-based framework for recursive frequency/spectrum estimation and feature detection of nonstationary signals, which may contain different extents of nonstationarities and impulsive components. The estimation of time-varying frequency or spectrum is formulated as a time-varying linear model identification problem, where the spectral information is estimated from the model coefficients. We then employ a QR-decomposition-based recursive least M-estimate (QRRLM) algorithm for recursive estimation of the time-varying model coefficients in impulsive environment using M-estimation. New variable forgetting factor (VFF) schemes are developed to improve the tracking performance of the QRRLM method in nonstationary environment and we use theoretical derivation and simulations to prove that the proposed VFF schemes can approach the optimal VFF selection. The resultant VFF-QRRLM algorithm is able to restrain and isolate impulsive components whereas it is able to handle different extents of spectral variations. Simulation results show that the proposed VFF-QRRLM algorithm is more robust and accurate than conventional recursive least squares-based methods in estimating both time-varying narrowband frequency components and broadband spectral components with impulsive components. Potential applications of the proposed method can be found in power quality monitoring, online fault detection and speech analysis.

Index Terms—M-estimation, recursive frequency estimation, spectrum estimation, time-varying linear model, variable forgetting factor.

I. INTRODUCTION

FREQUENCY and spectrum estimation of nonstationary signals have a wide range of practical applications in power quality monitoring [1], speech processing [2], biomedical signal analysis [3], vibration analysis and measurement [4] and so forth. Commonly used frequency/spectrum estimation methods can be broadly classified into two categories: nonparametric and parametric methods [5]–[7]. Nonparametric methods assume that the signal is composed of a specific

type of basis functions, such as sinusoids in discrete Fourier transform or wavelets in wavelet transform (WT) and the magnitude and phase of the signal at certain frequency are estimated by means of filtering or fitting. On the other hand, parametric methods, such as autoregressive (AR) or autoregressive moving averages (ARMAs) modeling, assume that a certain model generates the signal and the spectrum is computed from the model coefficients that are estimated by fitting the data with the model. Generally, if the assumed model is appropriate and the signal-to-noise ratio (SNR) is high,¹ parametric methods, which are the subject of this paper, have better frequency resolution than nonparametric methods in the analysis of narrowband signals. In parametric frequency and spectrum estimation methods, linear models in which the output depends linearly on the input and model coefficients are most widely used. For instance, in AR spectral estimation, the output is a linear combination of a set of immediate past outputs plus a noise-like excitation. The least squares (LS) method usually estimates the coefficients of the linear model, which minimizes the mean squared error of the noise-like sequence. In the context of frequency estimation, a commonly used approach is based on linear prediction (LP) [5] where the frequencies of a set of sinusoids are determined from the roots of a polynomial formed from the LP coefficients (LPCs).

In analyzing real-world signals and systems, classical linear-model-based parametric frequency and spectrum estimation methods often encounter two fundamental problems²: 1) nonstationarity and 2) sensitivity to impulsive components.

- 1) Firstly, most parametric frequency and spectrum estimators are intended for batch processing of a block of stationary data, and hence they are unsuitable for estimating signals with time-varying spectrum. An easy way to deal with time-varying spectrum is the sliding window approach [8], [9], but the automatic window selection for the best tradeoff between time resolution and frequency resolution is still an unaddressed problem.

Manuscript received September 12, 2012; revised March 18, 2013; accepted March 19, 2013. Date of publication August 1, 2013; date of current version November 6, 2013. The Associate Editor coordinating the review process was Dr. Rik Pintelon.

The authors are with the University of Hong Kong, Hong Kong (e-mail: zgzhong@eee.hku.hk).

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.2013.2272398

¹At very low SNR, it is generally known that the performance of parametric methods may degrade considerably and nonparametric methods may be more preferable [7].

²Another fundamental issue is model order selection, which can be addressed using the Akaike information criterion, Bayesian information criterion and related criteria [7].

In the context of parametric estimation using the linear model, the coefficients of the model are time varying, resulting in a time-varying linear model (TVLM) [10], [11]. The recursive estimation of TVLMs under noise is nontrivial as it usually involves the selection of an appropriate window size for fitting to seek a proper tradeoff between bias and variance. If the window size is too small, the variance of the estimated model coefficients will be large. On the other hand, the model may be unable to explain the observations in a large window, resulting in large bias. In [10] and [11], a new local polynomial modeling method can approach the optimal bias-variance tradeoff and can offer good performance for both slowly and fast varying signals, but its arithmetic complexity is high. For low-complexity implementation, recursive techniques such as adaptive filtering [12]–[15] are desirable. However, the selection of the forgetting factor, which controls the size of the exponential window in adaptive filtering, remains a great challenge.

- 2) Secondly, conventional LS-based estimators are sensitive to impulsive components, which violate the usual assumption of additive Gaussian noise. Furthermore, the adverse influence of impulsive components, which act as innovative variables for succeeding measurements in the AR or ARMA model, will last for a long period. Therefore, the performance of the LS-based methods will be severely degraded when such impulsive components are encountered. Much research has been devoted to mitigate their adverse effect on parametric spectral analysis [16]–[19], but online tracking of time-varying frequency/spectrum in presence of impulses has not been studied.

These two problems are particularly important in practical applications such as power quality monitoring, vibration and speech analysis and so forth. For example, power signals may contain harmful frequency variations caused by faults of power transmission systems and may be contaminated with impulsive transients caused by lightning [20]. Thus, realtime tracking of fundamental frequency and harmonics of power signals in presence of impulses is crucial for monitoring power quality [21]. Similar problems also exist in speech communication and recognition systems [22], [23], where the impulsive components may originate from transmission errors or adverse environments. Although several recursive frequency/spectrum estimation methods are available [8], [9], [14], [15], they still do not have the desired properties for addressing these two problems.

In this paper, we propose a general parametric framework for recursive frequency and spectrum estimation of nonstationary signals, which may contain different extents of spectral variations and impulsive components. The proposed framework is based on a recursive least M-estimation (RLM) for identifying the TVLM using: 1) the M-estimation function to safeguard the coefficients from adverse influence of the impulsive components and 2) variable and adaptive exponential windows to deal with different extents of spectral variations. The RLM algorithm [24] and [25] is

implemented using the QR decomposition (QRD) structure that leads to lower roundoff error and more efficient hardware realization. To address the limited tracking performance of the conventional QR decomposition-based recursive least M-estimate (QRRLM) in nonstationary environments, a new variable forgetting factor (VFF) control scheme is developed. The forgetting factor (FF), which is related to the size of the exponential window, is chosen from a possible candidate set to minimize the M-estimate errors (MMEs). It was shown by theoretical analysis that the performance of this MME-VFF scheme is close to that of the optimal parameter. Because the MME-VFF scheme has a high computational complexity, a low-complexity VFF scheme based on approximated derivative variable forgetting factor (AD-VFF) of coefficients is also introduced for real-time processing. Compared with the MME-VFF scheme, the AD-VFF scheme has a much lower complexity and slightly degraded performance. Simulation results show that the VFF-QRRLM algorithm can significantly improve the tracking performance over the conventional recursive least squares (RLS)/RLM algorithm in recursive frequency and spectrum estimation. In addition, other advanced adaptive filtering techniques, such as state regularization (SR) [26] and local polynomial regression (LPR) [27] can be incorporated into VFF-QRRLM to enhance its performance.

To illustrate the effectiveness of the proposed VFF-QRRLM method, we focus on two classical problems: 1) robust frequency estimation of real-valued multiple sinusoids with impulsive components and 2) robust ARMA spectral estimation with impulsive components. The first problem is important as LP-based models are well-suited for describing signals containing multiple sinusoids [28], [29], which occur frequently in detecting electromechanical modes in power systems to ensure stability [1]. The second problem is more general and it applies to signals with both narrowband and broadband spectra. The simulation results show that the VFF-QRRLM algorithm can satisfactorily estimate the time-varying frequency or spectrum and is particularly suitable for practical applications where online spectral feature detection is required. Compared with conventional methods such as RLS/RLM with fixed forgetting factors, the VFF counterpart in the proposed method can offer a faster response and hence a shorter lag. Compared with conventional nonparametric spectral methods (such as short-time Fourier transform and WT) that have large time lags in online spectral estimation because of the two-sided time window used, the VFF-QRRLM method only makes use of past data and thus is capable of extracting time-varying spectral features with much shorter lags.³

This paper is organized as follows. Section II is devoted to the problems of parametric frequency and spectrum estimation to be addressed. The proposed VFF-QRRLM algorithm will be introduced in Section III. Simulation results are presented in Section IV. Finally, conclusion is drawn in Section V.

³Though the filtering delay can be reduced by low-delay filterbanks and wavelets [30], [31], the delay is closely tied up with its length and hence stopband attenuation or performance. It should be noted that for certain situations such as phasor estimation, the non-parametric methods can be tailored to substantially reduce the estimation delay [32], [33].

II. TVLM BASED FREQUENCY AND SPECTRUM ESTIMATION

A. Parametric Model (Linear Prediction)-Based Frequency Estimation

We start with the recursive frequency estimation of a multiple sinusoidal signal from its real-valued discrete-time measurements

$$z(n) = s(n) + \varepsilon(n), \quad n = 1, 2, \dots, N \quad (1)$$

where $s(n) = \sum_{m=1}^M \alpha_m(n) \cos[\omega_m(n)n + \phi_m(n)]$ are the corresponding discrete-time samples of the signal with $\alpha_m(n) > 0$, $\omega_m(n) = 2\pi f_m(n)/F_s \in (0, \pi)$, and $\phi_m(n) \in [0, 2\pi]$ being, respectively, the time-varying amplitude, frequency (normalized by sampling frequency F_s), and phase of the m th frequency component. For simplicity, we assume that the number of sinusoids M is known and the frequencies are distinct, i.e., $\omega_{m1}(n) \neq \omega_{m2}(n)$ when $m1 \neq m2$. Conventionally, $\varepsilon(n)$ is assumed to be an additive white Gaussian noise with zero mean and variance σ_ε^2 . It was observed by Prony that $s(n)$ in (1) can be expressed as a linear combination of its $2M$ past samples based on the idea of LP [28] as follows:

$$s(n) = \sum_{i=1}^{2M} b_i(n)s(n-i) \quad (2)$$

where $b_i(n)$, $i = 1, \dots, 2M$, are the time-varying LPC. Through the symmetric property of the LPCs of real-valued sinusoids, $b_i(n) = b_{2M-i}(n)$, (2) can be rewritten as

$$s(n) + s(n-2M) = \sum_{i=1}^{M-1} b_i(n)[s(n-i) + s(n-2M+i)] + b_M(n)s(n-M). \quad (3)$$

Combining (1)–(3), one gets the following TVLM:

$$y(n) = \mathbf{x}^T(n)\boldsymbol{\beta}(n) + e(n) \quad (4)$$

where $y(n) = z(n) + z(n-2M)$, $\mathbf{x}(n) = [z(n-1) + z(n-2M+1), \dots, z(n-M+1) + z(n-M-1), z(n-M)]^T$, $\boldsymbol{\beta}(n) = [b_1(n), \dots, b_M(n)]^T$, and $e(n)$ is the prediction residual given by

$$e(n) = \varepsilon(n) + \varepsilon(n-2M) - \sum_{i=1}^{M-1} b_i(n)[\varepsilon(n-i) + \varepsilon(n-2M+i)] + b_M(n)\varepsilon(n-M). \quad (5)$$

In the conventional LP approach, the signals are assumed to be stationary and (4) can be solved using the LS estimator.⁴ On the other hand, if $\varepsilon(n)$ contains impulsive components, then robust regression techniques [34] should be employed.

⁴If $\varepsilon(n)$ is a zero-mean, independent and identically distributed (i.i.d.) Gaussian random process, then $e(n)$ is also a zero-mean Gaussian process. It was shown in [28] that $e(n)$ is correlated and a weighted LS estimator, which approximates the maximum likelihood estimator through estimating the covariance of the error and whitening the error, can be adopted for a better estimation accuracy. However, the estimation of the covariance of $e(n)$ involves an iterative process and hence a higher complexity. Moreover, from our simulation, it was found that the performance of the weighted LS estimator is only slightly better than that of the LS estimator. Therefore, we shall focus on the LS estimator in this paper.

After $\beta(n)$ is estimated, the frequency of the sinusoids can be computed by solving the following polynomial equation:

$$\sum_{i=0}^{2M} b_i(n) \exp(-j\omega_m(n)i) = 0 \quad (6)$$

where $b_i(n) = b_{2M-i}(n)$, $i = 1, \dots, 2M$, $b_0(n) = -1$, and $j = \sqrt{-1}$. After estimating the frequencies of the sinusoidal signals, the amplitudes and phases can be estimated using LS fitting [28], if $\varepsilon(n)$ is Gaussian.

B. Parametric Spectrum Estimation

In parametric spectrum estimation, a discrete-time signal $z(n)$, which may consist of both broadband and narrowband components, is modeled as the following time-varying ARMA process:

$$z(n) + \sum_{p=1}^P a(n, p)z(n-p) = \sum_{q=1}^Q c(n, q)\varepsilon(n-q) + \varepsilon(n) \quad (7)$$

where $a(n, p)$ and $c(n, q)$ are, respectively, the time-varying AR and MA coefficients, P and Q are, respectively, the orders of the AR and MA processes, and $\varepsilon(n)$ is the additive white noise with zero mean and variance σ_ε^2 . Equation (7) can be written as a TVLM: $y(n) = \mathbf{x}^T(n)\boldsymbol{\beta}(n) + e(n)$ as (4) where $y(n) = z(n)$, $\mathbf{x}(n) = [-z(n-1), \dots, -z(n-P), e(n-1), \dots, e(n-Q)]^T$, $\boldsymbol{\beta}(n) = [-a(n, 1), \dots, -a(n, P), c(n, 1), \dots, c(n, Q)]^T$, and $e(n) = \varepsilon(n)$. In the regression vector $\mathbf{x}(n)$, $e(n-q)$ is also unknown and needs to be estimated recursively from the estimated coefficients $\hat{\boldsymbol{\beta}}(n)$ as $\hat{e}(n) = y(n) - \mathbf{x}^T(n)\hat{\boldsymbol{\beta}}(n)$. The spectrum of $z(n)$ can then be calculated from the estimated $\hat{a}(n, p)$ and $c(n, q)$ as

$$P(n, \omega) = \sigma^2(n) \frac{\left| 1 + \sum_{q=1}^Q \hat{c}(n, q)e^{-jq\omega} \right|^2}{\left| 1 + \sum_{p=1}^P \hat{a}(n, p)e^{-jp\omega} \right|^2}. \quad (8)$$

III. VFF-QRRLM ALGORITHM

We have shown in the previous section that both the recursive frequency estimation and recursive spectrum estimation can be formulated as a problem of solving a TVLM. In this section, a low-complexity recursive algorithm, called the VFF-QRRLM algorithm is introduced to address the two problems (nonstationarity and sensitivity to impulses) in identifying a TVLM.

A. Recursive Least M -Estimation

In the LP-based frequency estimation problem (1) and the ARMA-based spectrum estimation problem (7), the additive noise $e(n)$ is usually assumed to be Gaussian distributed. For recursive estimation of a TVLM under Gaussian noise, the LS

cost function in the form of the sum of exponentially weighted LS error is commonly used as follows:

$$J_{LS}[\beta(n)] = \sum_{i=1}^n \lambda^{n-i} e^2(i) \quad (9)$$

where λ is a FF between 0 and 1. It can be seen that the FF is used to control the effective size of the exponential window by weighting distant measurements, say the i th one, by λ^{n-i} . Because of the recursive nature of (9), it can be solved using the celebrated RLS algorithm [6]. It is well known, however, that the LS cost function is sensitive to impulses and a more robust approach is to employ M-estimation [34]. The cost function in M-estimation consists of the following sum of exponentially weighted M-estimate errors:

$$J_{ME}[\beta(n)] = \sum_{i=1}^n \lambda^{n-i} \rho(e(i)) \quad (10)$$

where $\rho(e)$ is an M-estimate function usually chosen as an appropriate function to reduce the sensitivity of the estimator to non-Gaussian distributed noise. An effective M-estimation function is the Hampel's three parts redescending function (Fig. 1) as follows:

$$\rho(e) = \begin{cases} \frac{e^2}{2} & 0 \leq |e| < \xi_a \\ \xi_a |e| - \frac{\xi_a^2}{2} & \xi_a \leq |e| < \xi_b \\ \xi_a [(\xi_b + \xi_c) - |e|] + \frac{(|e| - \xi_c)^2}{2} & \xi_b \leq |e| < \xi_c \\ \frac{\xi_a(\xi_b + \xi_c)}{2} - \frac{\xi_c^2}{2} & \xi_c \leq |e| \end{cases} \quad (11)$$

where ξ_a , ξ_b , and ξ_c are the thresholds to control the degree of impulse suppression. The thresholds can be estimated based on the variance of the impulse-free prediction error, which can be recursively estimated as [24]

$$\hat{\sigma}^2(n) = \lambda_\sigma \hat{\sigma}^2(n-1) + c_1(1 - \lambda_\sigma) \text{median}[A_e(n)] \quad (12)$$

where λ_σ is the FF of the updating process, $A_e(n) = \{\hat{e}^2(n), \dots, \hat{e}^2(n - N_w + 1)\}$, $\hat{e}(n) = y(n) - \mathbf{x}^T(n)\hat{\beta}(n-1)$, $c_1 = 1.483(1 + 5/(N_w - 1))$ is a finite sample correction factor, and N_w is the length of the estimation window. Usually N_w is chosen between 5 and 11, so that the system can restrain two to five consecutive impulses and it can be increased if necessary, at the expense of slower adaptation to system changes [24]. For the Hampel's M-estimate function, these thresholds can be set as $\xi_a = 1.96 \cdot \hat{\sigma}$, $\xi_b = 2.24 \cdot \hat{\sigma}$, and $\xi_c = 2.58 \cdot \hat{\sigma}$, which correspond to 95%, 97.5%, and 99% confidence intervals of a Gaussian distribution $N(0, \hat{\sigma})$. More details about M-estimation functions and parameter selection can be found in [24].

The solution of (10) can be obtained by setting its derivative with respect to $\beta(n)$ to zero, resulting in the following M-estimate normal equation:

$$\tilde{\mathbf{R}}_{ME}(n)\beta(n) = \tilde{\mathbf{P}}_{ME}(n) \quad (13)$$

where $\tilde{\mathbf{R}}_{ME}(n) = \lambda \tilde{\mathbf{R}}_{ME}(n-1) + \tilde{q}(n)\mathbf{x}(n)\mathbf{x}^T(n)$, $\tilde{\mathbf{P}}_{ME}(n) = \lambda \tilde{\mathbf{P}}_{ME}(n-1) + \tilde{q}(n)y(n)\mathbf{x}(n)$, and $\tilde{q}(n) = q(\hat{e}(n)) = \rho'(\hat{e}(n))/\hat{e}(n)$. Therefore, a similar derivation of the conventional RLS algorithm can be followed to obtain the RLM

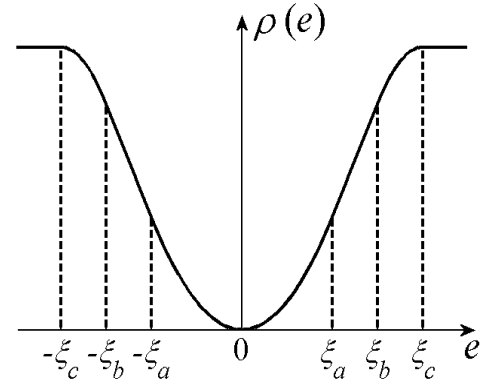


Fig. 1. Hampel's three parts redescending function.

algorithm [24]. Here, we propose to implement the RLM algorithm using QRD because the QRD holds a higher numerical stability over the direct implementation of RLS [6]. In addition, QRD simplifies an approximate but computationally efficient implementation of the SR techniques, which will be detailed in Section III-D.

The arithmetic complexity of the VFF-QRRLM algorithm is of the order $O(\Gamma^2)$, where Γ is the dimension of coefficients $\beta(n)$. The convergence performance of the RLM algorithm under the contaminated Gaussian (CG) impulsive noise model has been analyzed in [24] and [36].⁵ The CG noise model has the following probability density function:

$$P(e) \sim (1 - p_r)N(0, \sigma_g) + p_r N(0, \sigma_{im}) \quad (14)$$

where σ_g^2 and σ_{im}^2 are, respectively, the variances of two independent zero mean Gaussian processes with $\sigma_{im}^2 > \sigma_g^2$. p_r is a small positive value between 0 and 1 and it denotes the occurrence probability of the impulsive component with variance σ_{im}^2 . In this paper, the CG noise model is adopted to simulate the additive Gaussian noise (with variance of σ_g^2) plus the impulsive component (with variance σ_{im}^2). It should be noted that the impulsive components are only additive to the measurements but not served as excitation variables to the ARMA model.

B. VFF Schemes

1) *Minimum M-Estimate VFF*: To cope with nonstationary TVLM using the RLS/RLM algorithm, one can select appropriately the FF λ to achieve a proper tradeoff between estimation bias and variance. Ideally, for Gaussian noise, the optimal FF is the one that minimizes the sum of squared filtering error as follows:

$$\lambda_{\text{Ideal}} = \arg \min_{\lambda \in \Lambda} \sum_{n=1}^N [y_0(n) - \hat{y}_\lambda(n)]^2 \quad (15)$$

⁵In [24] and [36] we have theoretically and experimentally shown that the RLM algorithm is more robust than the RLS algorithm under such additive CG noise. For the LP and AR models, the convergence analysis in [36] will apply directly. This will allow us to obtain the evolution equations describing the mean and covariance of the weight vector over time and their steady state MSE. For the ARMA model, the coupling between the excitation and regression vector makes the analysis very involved, and due to page limitation, we do not analyze it further here.

where $\hat{y}_\lambda(n) = \mathbf{x}^T(n)\hat{\beta}_\lambda(n-1)$ is the a prior estimate of $y_0(n) = \mathbf{x}^T(n)\beta(n)$ based on a FF λ and Λ is a FF set containing all possible FFs. In practice, Λ is limited to a series of candidate FFs and one needs to use the noisy measurements $y(n)$ since the noise-free measurements $y_0(n)$ are unavailable. In [37], it has been shown theoretically and experimentally that the data-driven FF λ obtained from (15) using the noisy measurements performs nearly as well as the ideal FF λ_{Ideal} in presence of Gaussian noise. In addition, it is also shown in [37] that, instead of using a global FF, a locally adaptive λ (or local FF for short), which uses the most recent L measurements instead of all the measurements, can better adapt to the time-varying coefficients and achieve better performance.

If the measurements are contaminated with CG or impulsive noise, it is preferred to use the M-estimate instead of the LS estimate, which suggests the following MME data-driven global FF:

$$\lambda = \arg \min_{\lambda \in \Lambda} \sum_{n=1}^N \rho(y(n) - \hat{y}_\lambda(n)) \quad (16)$$

and local FF as follows:

$$\lambda(n) = \arg \min_{\lambda \in \Lambda} \sum_{i=n-L+1}^n \rho(y(i) - \hat{y}_\lambda(i)) \quad (17)$$

where L is the length of a moderately short period of neighboring time points preceding n .

In the Appendix, we show that such MME data-driven FF can achieve nearly as good performance as the ideal FF in presence of impulsive noise, providing a theoretical justification of the proposed MME-VFF schemes. The MME-VFF method and the coupled analysis are actually very general and can be applied to the selection of other possible parameters, such as the step-size in the least mean squares algorithm and the number of measurements in the Kalman filter.

To find the MME-VFF at time point n based on Section III-B.2, one needs to start from time point $n-L$ and estimate the coefficients $\hat{\beta}_\lambda(i)$, $i = n-L, \dots, n-1$, with every λ in the set Λ . Then, the mean of M-estimate error of the most recent L samples is calculated to determine $\lambda(n)$ according to Section III-B.2. Although the MME-VFF scheme works well in practice, it has a high computational complexity. More precisely, if the number of candidate FFs in Λ is N_Λ , the MME-VFF scheme requires extra $N_\Lambda L$ times of QRD-based recursions for calculating the VFF at each time instant. Therefore, an alternative low-complexity empirical VFF scheme is introduced next.

2) *Approximated Derivative Variable Forgetting Factor:* A reasonable and effective approach to determine the FF is based on the AD of the estimated model coefficients as it is a measure of the time variation of the underlying time-varying model parameters. Intuitively, if the variation of $\hat{\beta}(n)$ is large, a small FF should be used and vice versa. Such an AD scheme has been employed in [37] for selecting the number of measurements in updating the Kalman filter and, in this paper, the AD scheme is adopted to select the FFs of QRRLM. More precisely, the variation of coefficient $\hat{\beta}(n)$ is approximated as $\delta(n) = \hat{\beta}(n) - \hat{\beta}(n-1)$, where $\hat{\beta}(n) = \lambda_\beta \hat{\beta}(n-1) +$

$(1-\lambda_\beta)\hat{\beta}(n-1)$ is the smoothed past coefficient estimates and $0 < \lambda_\beta \leq 1$ is the corresponding FF of the smoothing process. We can see that if $\hat{\beta}(n)$ changes rapidly, the magnitude of the AD of $\|\delta(n)\|$, $G_\delta(n) = \|\delta(n)\| - \|\delta(n-1)\|$, will be large, and vice versa. A more robust measure of the coefficient variation can be obtained by smoothing and normalizing $G_\delta(n)$ as follows:

$$\bar{G}_{\delta,N}(n) = \bar{G}_\delta(n) / \bar{G}_{\delta 0} \quad (18)$$

where $\bar{G}_\delta(n)$ is the average of $G_\delta(n)$ over a smoothing window of length N_S and $\bar{G}_{\delta 0}$ is the median of the first N_T estimates (training samples) of $\bar{G}_\delta(n)$ for normalization. $\bar{G}_{\delta 0}$ is obtained as the median value to avoid the influence of large estimation error when the algorithm has not converged at the beginning. After the training period, the FF $\lambda(n)$ can be updated at each time instant as

$$\lambda(n) = \lambda_{\min} + [1 - g(\bar{G}_\delta(n))](\lambda_{\max} - \lambda_{\min}) \quad (19)$$

where λ_{\min} and λ_{\max} are, respectively, the lower and upper bounds of $\lambda(n)$ and $g(u) = \min\{\max\{u, 0\}, 1\}$ is a clipping function used to keep $\lambda(n)$ between λ_{\min} and λ_{\max} . During the training period ($n < N_T$), a fixed FF, such as $(\lambda_{\min} + \lambda_{\max})/2$, can be used.

C. RLM Based Frequency and Spectrum Estimation

We now discuss the applications of the proposed VFF-QRRLM algorithm to the problem of recursive frequency and spectrum estimation. Since both the LP-based frequency estimation and ARMA-based spectrum estimation use past data measurements to estimate the current coefficients, impulsive measurements, once detected, should be replaced with their approximated impulse-free counterpart to avoid the adverse effects from propagating to subsequent estimates. The impulse-free measurements can be approximated as

$$\tilde{y}(n) = \tilde{q}(n)y(n) - (1 - \tilde{q}(n))\hat{y}(n) \quad (20)$$

where $\hat{y}(n) = \mathbf{x}^T(n)\hat{\beta}(n)$, $\tilde{q}(n) = q(\hat{e}(n)) = \rho'(\hat{e}(n))/\hat{e}(n)$, and $\hat{e}(n) = y(n) - \mathbf{x}^T(n)\hat{\beta}(n-1)$. It can be seen that $\tilde{q}(n)$ will be equal to one when the prediction (*a priori*) error $\hat{e}(n)$ using $\hat{\beta}(n-1)$ is small, which means that the measurement is less likely to be corrupted by impulses and hence $\tilde{y}(n)$ will be equal to $y(n)$. In contrast, if $\hat{e}(n)$ is very large, it is likely that $y(n)$ is corrupted by impulse and hence $\tilde{q}(n)$ will assume a small value, so that impulse-contaminated measurement will be replaced by its impulse-free estimate $\tilde{y}(n)$.

In addition, in the recursive ARMA-based spectrum estimation, the a posteriori estimation error $\tilde{e}(n)$, $\tilde{e}(n) = y(n) - \hat{y}(n)$, are also a part of the regression vector and thus it should be safeguarded from impulse-contaminated measurements in a similar manner. More precisely, we estimate the impulse-free estimation error as $\tilde{e}(n) = \tilde{y}(n) - \hat{y}(n)$. The robust estimation of amplitudes and phases in impulsive environment can be achieved using M-estimation cost function instead of the LS cost function as well [34]. Finally, because the impulsive components may also contain meaningful information, we propose to approximate the impulse components as the difference between the impulse-free estimation noise $\tilde{e}(n)$ and

TABLE I
VFF-QRRLM FOR FREQUENCY/SPECTRUM ESTIMATION

Initialization

$\mathbf{R}(0) = \sqrt{\delta} \mathbf{I}$, δ is a small positive constant and \mathbf{I} is an identity matrix; $\mathbf{d}(0) = \mathbf{0}$; $\hat{\boldsymbol{\beta}}(0) = \mathbf{0}$; $\lambda(0) = 1$.

Recursion (At each time n)

- 1). Update the FF $\lambda(n)$ with the MME-VFF scheme (17) or the AD-VFF scheme (19).
- 2). Calculate the prediction error $\hat{e}(n)$ and approximate the impulse-free variance $\hat{\sigma}^2(n)$ as (12).
- 3). Update the M-estimation function.
- 4). Construct the recession vector $\mathbf{x}(n)$ from impulse-free past measurement and noise, and estimate $\hat{\boldsymbol{\beta}}(n)$ using QRD*.
- 5). Estimate frequency $\omega_m(n)$, amplitude, and phase from $\hat{\boldsymbol{\beta}}(n)$ as (6).
- 6). Approximate impulse-free measurement $\tilde{y}(n)$ and impulse-free noise $\tilde{e}(n)$ as (20).

* QRD (for any n):

- i). Given $\mathbf{R}(n-1)$, $\mathbf{d}(n-1)$, $\hat{\boldsymbol{\beta}}(n-1)$, $y(n)$ and $\mathbf{x}(n)$, $\lambda_e(n)$, and $q(e(n))$, compute

$$\begin{bmatrix} \mathbf{R}(n) & \mathbf{d}(n) \\ \mathbf{0} & v(n) \end{bmatrix} = \mathbf{Q}(n) \begin{bmatrix} \sqrt{\lambda(n)} \mathbf{R}(n-1) & \sqrt{\lambda(n)} \mathbf{d}(n-1) \\ \sqrt{\tilde{q}(n)} \mathbf{x}^T(n) & \sqrt{\tilde{q}(n)} y(n) \end{bmatrix}$$

where $\mathbf{Q}(n)$ can be calculated by Givens rotation.

- ii). (for VFF-SR-QRRLM only)

$$\begin{bmatrix} \mathbf{R}'(n) & \mathbf{d}'(n) \\ \mathbf{0} & v'(n) \end{bmatrix} = \mathbf{Q}'(n) \begin{bmatrix} \mathbf{R}(n) & \mathbf{d}(n) \\ \sqrt{\tilde{q}(n)\kappa(n)\Gamma E_\gamma} & \sqrt{\tilde{q}(n)} \hat{\boldsymbol{\beta}}_\gamma(n-1) \end{bmatrix}$$

where $\hat{\boldsymbol{\beta}}_\gamma(n-1)$ is the γ -th entry of $\hat{\boldsymbol{\beta}}(n-1)$, $\gamma = (n \bmod \Gamma) + 1$ for sequential update, and \mathbf{E}_γ is a $\Gamma \times 1$ vector with its γ -th ($\gamma = 1, 2, \dots, \Gamma$) entry equal to 1 and other entries equal to 0.

- iii). Solve $\hat{\boldsymbol{\beta}}(n)$ from $\mathbf{d}(n) = \mathbf{R}(n)\hat{\boldsymbol{\beta}}(n)$ or $\mathbf{d}'(n) = \mathbf{R}'(n)\hat{\boldsymbol{\beta}}(n)$ (VFF-SR-QRRLM) by back-substitution.

the estimation noise $\tilde{e}(n)$. The final proposed VFF-QRRLM algorithm for recursive frequency/spectrum estimation is listed in Table I.

D. State Regularization

The proposed two VFF schemes can also be applied to variants of the QRRLM algorithms, for example, the SR-QRRLM. The SR technique was developed in [26] to address the inherent problem of conventional RLS when the input is not persistently excited that results in an input covariance matrix being ill-conditioned. In such situations, the

variance of the RLS estimator will increase considerably or even become unstable. The SR technique aims at reducing the estimation variance of RLS estimator by an adaptive regularization term without introducing any asymptotic bias. To incorporate the SR technique into QRRLM, we can rewrite the M-estimation normal function of (13) as

$$[\tilde{\mathbf{R}}_{\text{ME}}(n) + \kappa(n)\mathbf{I}]\boldsymbol{\beta}(n) = \tilde{\mathbf{P}}_{\text{ME}}(n) + \boldsymbol{\beta}(n-1) \quad (21)$$

where $\kappa(n)$ is the regularization parameter that can be selected adaptively as in [26]. In the new state-regularized M-estimation normal equation, $\kappa(n)\mathbf{I}$ is the regularization term for addressing the possible ill-conditioned problem of $\tilde{\mathbf{R}}_{\text{ME}}(n)$ whereas $\boldsymbol{\beta}(n-1)$ acts as prior information to compensate for the extra bias introduced by the regularization term. Equation (21) can be recursively updated by an approximate but computationally efficient QRD implementation (the second QRD in Table I). The idea underlying is that, in a period of Γ time points, updating each row of $\tilde{\mathbf{q}}(n)\Gamma\kappa(n)\mathbf{I}$ in the second QRD at each time point approximates the updating with $\tilde{\mathbf{q}}(n)\kappa(n)\mathbf{I}$ at every time point. Such implementation can reduce the complexity of the second QRD in VFF-SR-QRRLM from $O(\Gamma^3)$ to $O(\Gamma^2)$. Therefore, compared with the VFF-QRRLM algorithm, the VFF-SR-QRRLM only needs one more QRD operation in each recursion, as listed in Table I.

E. Practical Issues

1) *Complexity Analysis*: The conventional QRRLS algorithm has a complexity of $O(\Gamma^2)$ at each time instant. The QRRLM algorithm also includes the calculations of $\hat{e}(n)$, the noise variance in (12) and the M-estimation weight, which only require several flops or has a complexity of $O(\Gamma)$. Therefore, the QRRLM has a complexity of $O(\Gamma^2)$ as well. The computationally expensive part is the MME-VFF scheme, because it requires calculating extra $N_\Lambda L$ times of QRD-based recursions for calculating the VFF at each time ($L = 32$ and $N_\Lambda = 7$ in our simulations). In the AD-VFF scheme, (18), (19), and other calculation only need a few flops. Therefore, the AD-VFF scheme will not increase the computational complexity substantially and its complexity is still $O(\Gamma^2)$. In summary, the AD-VFF-QRRLM algorithm is computationally efficient and it can achieve comparable results to the theoretically nearly optimal but computationally expensive MME-VFF-QRRLM algorithm.

2) *Selection of Model Order*: The order of AR or MA models is an important parameter for the proposed VFF-QRRLM method as it is susceptible to overfitting. To avoid the latter problem, the model order can generally be selected by certain model complexity measure such as the Akaike information criterion or Bayesian information criterion [7], which imposes a penalty on the model complexity during data fitting. In the context of frequency estimation, the model order (i.e., the number of frequency components) can also be determined by other spectral analysis techniques. For example, an exact model order ESPRIT has been developed in [39] to give an initial estimate of the number of components in the signal. The method in [39] can also be used to provide initial values for the VFF-QRRLM algorithm for a faster convergence.

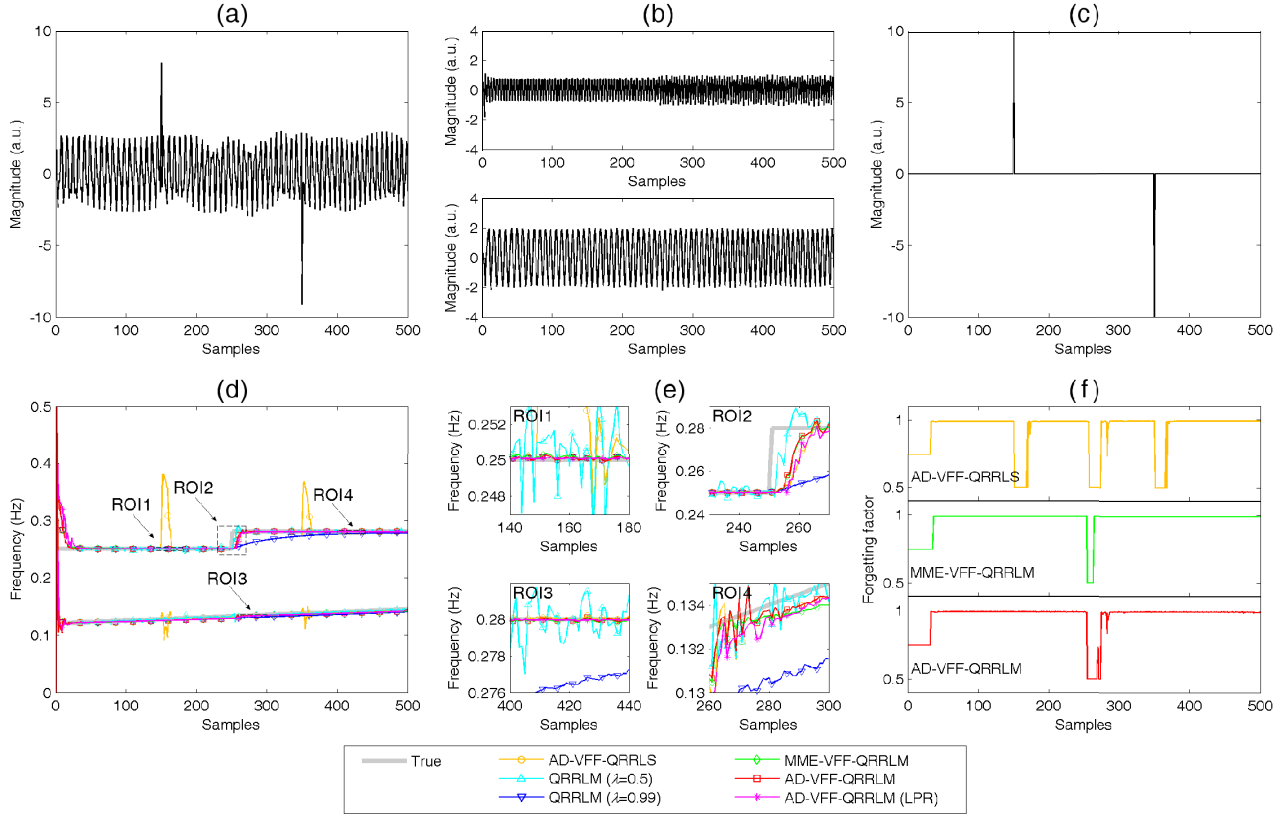


Fig. 2. Comparison between various recursive frequency tracking methods on a two-component sinusoidal signal in presence of impulses. (a) Data measurements. (b) Two sinusoidal components reconstructed from AD-VFF-QRRLM. (c) Impulsive components reconstructed from AD-VFF-QRRLM. (d) Estimated frequency. (e) Enlarged regions of interests of (d). (f) VFF used.

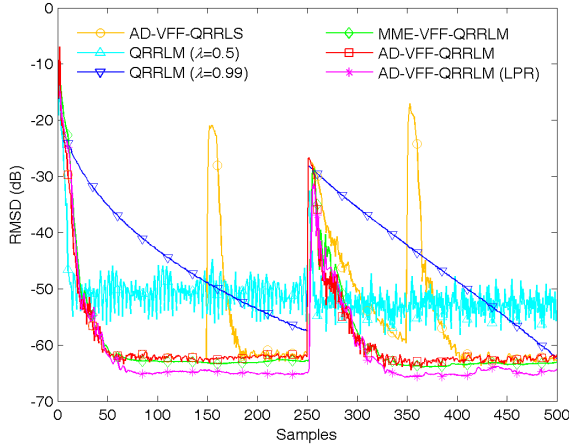


Fig. 3. RMSD curves of various recursive frequency tracking methods (SNR = 30 dB).

3) Selection of Hyperparameters Used in VFF: In the VFF schemes, the minimum and maximum FFs, λ_{\min} and λ_{\max} , determine the ability of the algorithm in tracking fast-varying frequency components and estimating stable frequency components. In our simulations, λ_{\min} and λ_{\max} are selected as 0.5 and 0.99, which correspond to effective window lengths of 2 and 100, respectively. That is to say, the VFF-QRRLM coupled with $\lambda_{\min} = 0.5$ and $\lambda_{\max} = 0.99$ is capable of dealing with signals that are stationary (i.e., frequency components

or spectra are unchanged) within 2 to 100 samples. In other applications, λ_{\min} and λ_{\max} can be determined according to prior knowledge (such as the sampling rate, the longest, and shortest period of local stationarity) about the signal. In the AD-VFF scheme, the length of training samples N_T can be as long as possible to obtain an accurate value for normalization, but an unnecessarily large N_T may include sharp changes or impulsive components in the training. In our simulations, N_T was set to 32, which implies the difference between the true mean and the sample mean is less than half of the standard deviation with a probability of $>99.5\%$ [40]. Similarly, the average window size N_S is set as 32 to obtain a tradeoff between statistical power and the capability of detecting sharp changes. The FF λ_β for updating the variation of coefficient can be selected as 0.9 to avoid over smoothing or under smoothing.

In the MME-VFF scheme, the optimal FF is selected from a number of candidate FFs in a set N_Λ . More FFs in the candidate set will lead to more refined estimation results, but will also increase the computational complexity. To achieve a good tradeoff between performance and complexity, we usually select 3 to 5 FFs between λ_{\min} and λ_{\max} . The window parameter L for obtaining the local MME in (17) is selected as 32 for a tradeoff between tracking performance and estimation accuracy.

Our intensive simulations show that the proposed selection and estimation of hyperparameters in VFF schemes gave

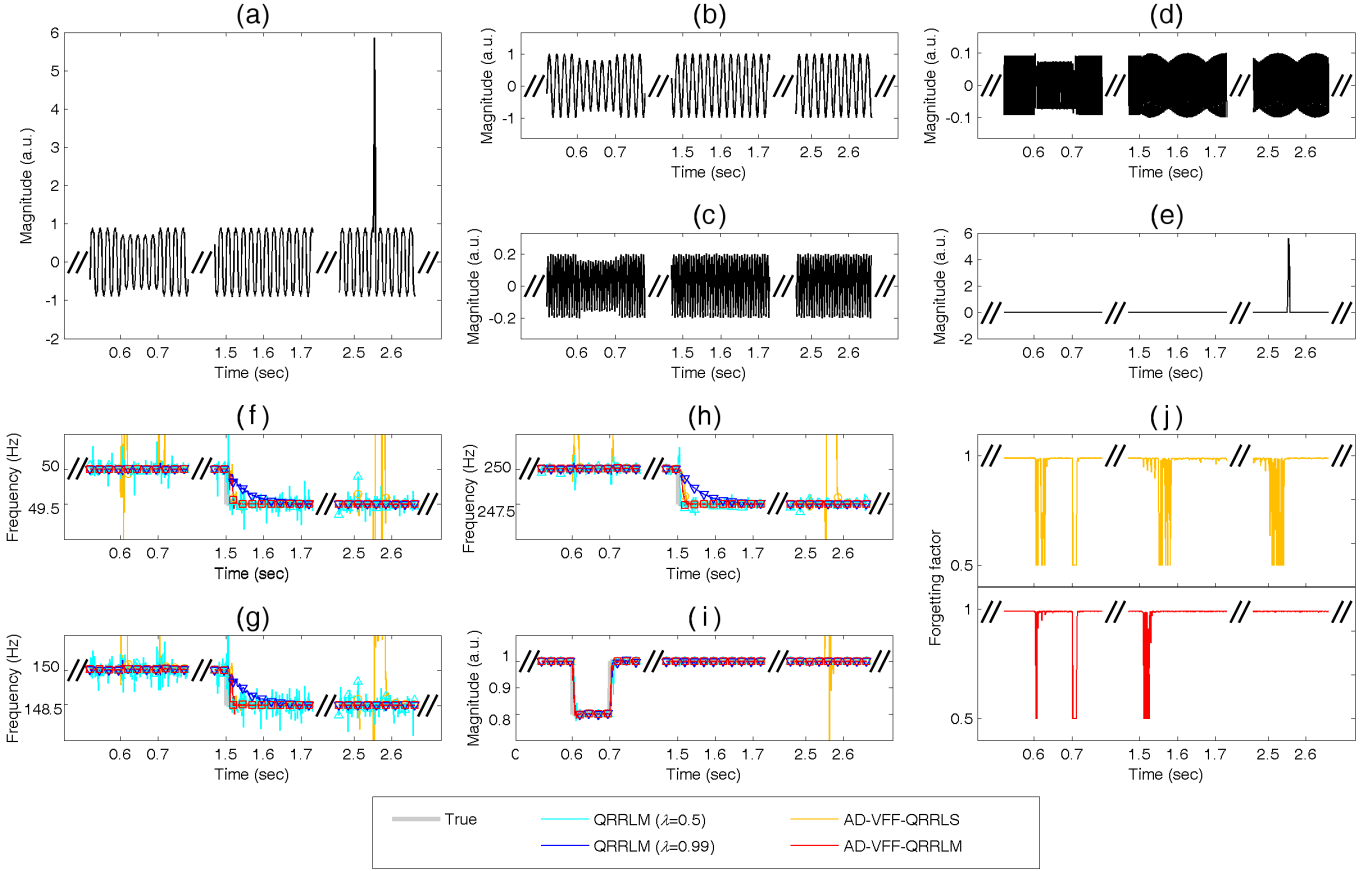


Fig. 4. Comparison between various recursive frequency tracking methods on a power signal in presence of harmonics and impulses. (a) Power measurements. (b)–(d) Fundamental, third harmonic, and fifth harmonic components reconstructed from AD-VFF-QRRLM. (e) Impulsive component reconstructed from AD-VFF-QRRLM. (f)–(h) Estimated frequency of the fundamental, the third harmonic, and the fifth harmonic components. (i) Estimated amplitude of the fundamental component. (j) VFF used.

satisfactory results and the proposed method was not sensitive to small variations of most of hyperparameters. Except for λ_{\min} and λ_{\max} , other hyperparameters do not have a substantial effect on the performance of the VFF-QRRLM algorithm.

4) *Postprocessing and Trajectory Smoothing*: Frequency and amplitude of time-varying sinusoidal components or spectral peaks convey important information regarding the trajectories of sinusoidal or spectral components constituting the signal under study. Although the VFF-QRRLM can achieve a reasonably good bias-variance tradeoff, variations still exist in the estimated trajectories because of noise and other imperfections. Therefore, it is advantageous to perform further smoothing of the trajectories to suppress these variations. The trajectory smoothing can be achieved by LPR [27], [41], and [42], which can achieve a better bias-variance tradeoff in smoothing. The theory and implementation of the LPR method are omitted and interested readers can refer to [27], [41], and [42] for details.

IV. EXPERIMENTAL RESULTS

A. LP-Based Frequency Estimation (Simulations)

The sinusoidal signal under test has 500 samples and the sampling rate is 1. It consists of two sinusoidal

components ($M = 2$) and their amplitudes and phases are $\alpha_m = [1, 2]$ and $\phi_m = [\pi/4, \pi/3]$, respectively. One sinusoidal component has a sudden frequency change from 0.28 to 0.25 Hz at the 250th sample, while the other sinusoidal component has a linearly increasing frequency, as shown in Fig. 2(d). An additive Gaussian white noise with zero mean and a SNR of 30 dB is added. Two impulsive components with signal-to-impulse ratio (SIR, defined as the ratio between power of impulse and power of signal) of -20 dB are added at the 150th and the 350th samples for illustration.

In the VFF-QRRLM algorithm, the hyperparameters used for MME-VFF is $\Lambda = [0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99]$, $L = 32$; the hyperparameters used for AD-VFF selection are $N_T = N_S = 32$, $\lambda_{\min} = 0.5$, $\lambda_{\max} = 0.99$, $\lambda_\beta = 0.9$; the hyperparameters used for estimation of impulse-free error in (12) are $\lambda_\sigma = 0.9$ and $N_w = 32$. The performances of the proposed VFF-QRRLM algorithms are compared with the LS version, VFF-QRRLS, and QRRLM with a fixed FF (λ_{\min} or λ_{\max}). The performance of LPR-based trajectory smoothing for estimated frequency is evaluated as well. After trajectory smoothing, the frequency estimates from MME-VFF-QRRLM and AD-VFF-QRRLM are very similar. Thus, we only show the results of LPR trajectory smoothing for AD-VFF-QRRLM.

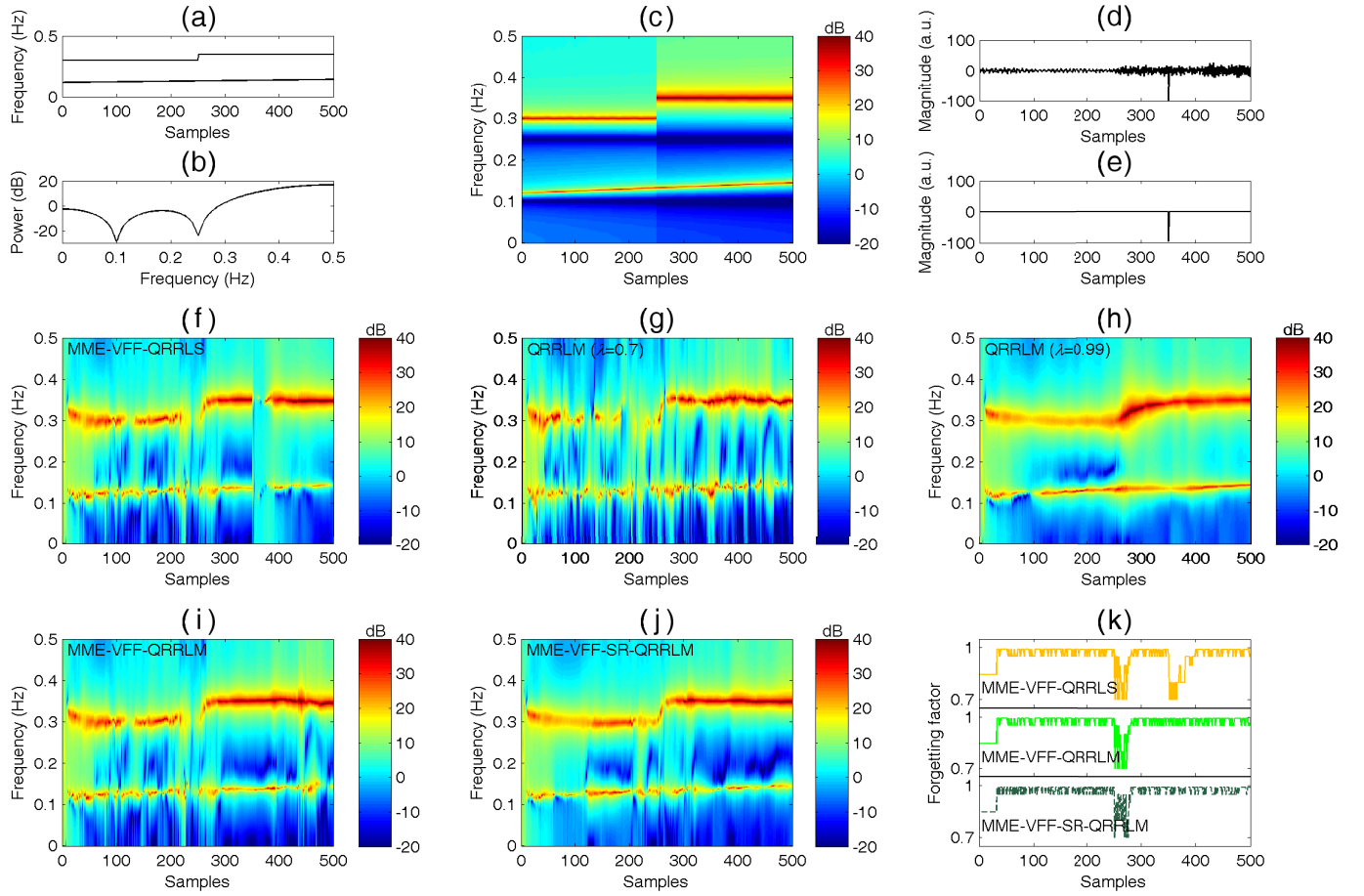


Fig. 5. Comparison between various recursive spectrum tracking methods on a time-varying ARMA-generated signal in presence of impulses. (a) Two sinusoidal components used to generate the time-varying AR coefficients. (b) Spectrum of MA coefficients. (c) True time-varying power spectrum. (d) Data measurements. (e) Reconstructed impulsive component from MME-VFF-SR-QRRLM. (f)–(j) Time-varying power spectra obtained by MME-VFF-QRRLS, QRRLM ($\lambda = 0.7$), QRRLM ($\lambda = 0.99$), MME-VFF-QRRLM, and MME-VFF-SR-QRRLM, respectively. (k) VFF used.

A root mean squared deviation (RMSD) criterion is used to quantitatively compare different frequency tracking methods, and it is given by

$$\text{RMSD}(n) = \sqrt{\frac{1}{M} \sum_{m=1}^M |f_m(n) - \hat{f}_m(n)|^2} \quad (22)$$

where $\hat{f}_m(n)$ is the m th estimated frequency. The RMSD curves from averages of 100 independent Monte Carlo runs are shown in Fig. 3. Fig. 2 shows the tracking performance of different methods for one example of testing signal, where the two impulsive components have amplitudes 10 and -10 . It can be seen from Figs. 2 and 3 that the QRRLM method can well suppress the adverse effects of impulses and effectively separate sinusoidal components and impulses. On the other hand, the tracking performance of QRRLS is seriously degraded in presence of impulses. We can also see that, although the QRRLM with a small FF can track the sudden frequency change quickly, its variance is very large. On the other hand, the QRRLM with a large FF has relatively smaller variance when tracking static frequency but it has a long lag in tracking the sudden frequency change. Both the MME- and AD-VFF schemes can obtain good results

in tracking both static and fast varying frequencies using a small FF for fast frequency changes and a large FF for static frequency components. In addition, LPR-based trajectory smoothing can further reduce the variance of static frequency estimation without introducing a long lag for sharp frequency change. The performance of the VFF-QRRLM method is also tested under different amount of additive Gaussian noise, and the results can be found in the supplementary materials (http://www.eee.hku.hk/~zgzhang/publication/tim2013_supp.pdf).

B. LP-Based Frequency Estimation (Application to Power Quality Monitoring)

Next, the effectiveness of the VFF-QRRLM method is illustrated in estimating frequency variations of power signals in presence of both harmonics and impulsive transients. Our simulation is carried out in PSCAD/EMTDC, which is an electromagnetic power transient software widely used for simulating power system transients. The power signal has a fundamental frequency of 50 Hz, a third harmonic component (150 Hz), and a fifth harmonic component (250 Hz). The sampling frequency of the power signal is 1000 Hz. The amplitude of the fundamental component decreases from 1 to 0.8 at

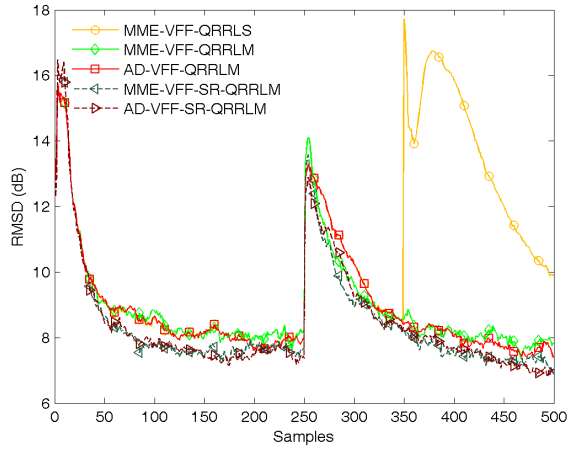


Fig. 6. RMSD curves of various recursive spectrum tracking methods.

0.6 s and recovers to 1 at 0.7 s, which simulates a typical voltage dip event. The amplitudes of the third and the fifth harmonic components are 20% and 10% of the amplitude of the fundamental component. At 1.5 s, the fundamental frequency jumps sharply from 50 to 49.5 Hz. The signal is also contaminated with an impulse transient, which has an amplitude of 10, at 0.7 s. A Gaussian white noise with zero mean and a SNR of 60 dB, which is in the normal SNR range (50–70 dB) of power signals, is added. The signal is also contaminated with an impulse transient, which has an amplitude of 5.5, at 2.575 s. We can see from Fig. 4 that the proposed VFF-QRRLM method can track the frequency change and amplitude change very quickly (~ 15 ms) and can achieve stable frequency estimates in the steady state. It is not only robust against impulsive transients, but also able to isolate the impulsive transient and harmonics effectively, thereby providing rich information about the status of the power quality.

C. ARMA-Based Spectrum Estimation (Simulations)

We now test the performance of the VFF-QRRLM and VFF-SR-QRRLM methods in recursive ARMA-based spectrum estimation. First, we use two time-varying sinusoidal components, as shown in Fig. 5(a), to generate a fourth-order time-varying AR coefficient vector. The MA coefficient vector is $[-1.5857, 1.9208, -1.5229, 0.9224]$, which implies a spectrum, as shown in Fig. 5(b). The true time-varying power spectrum of the time-varying ARMA process is shown in Fig. 5(c). Fig. 5(d) shows an example of testing signal simulated using the ARMA process, where the variance of the excitation noise is 1. An impulsive component with an amplitude of -100 is added at the 350th sample. The hyperparameters for VFF are $\lambda_{\min} = 0.7$, $\lambda_{\max} = 0.99$, and $\Lambda = [0.7, 0.8, 0.9, 0.95, 0.99]$. Other hyperparameters used in the recursive spectrum estimation are the same as those in the recursive frequency estimation.

The RMSD criterion for the recursive spectrum estimation is given by

$$\text{RMSD}(n) = \sqrt{\frac{1}{N_{\omega}} \sum_{n_{\omega}=1}^{N_{\omega}} |P(n, \omega) - \hat{P}(n, \omega)|^2} \quad (23)$$

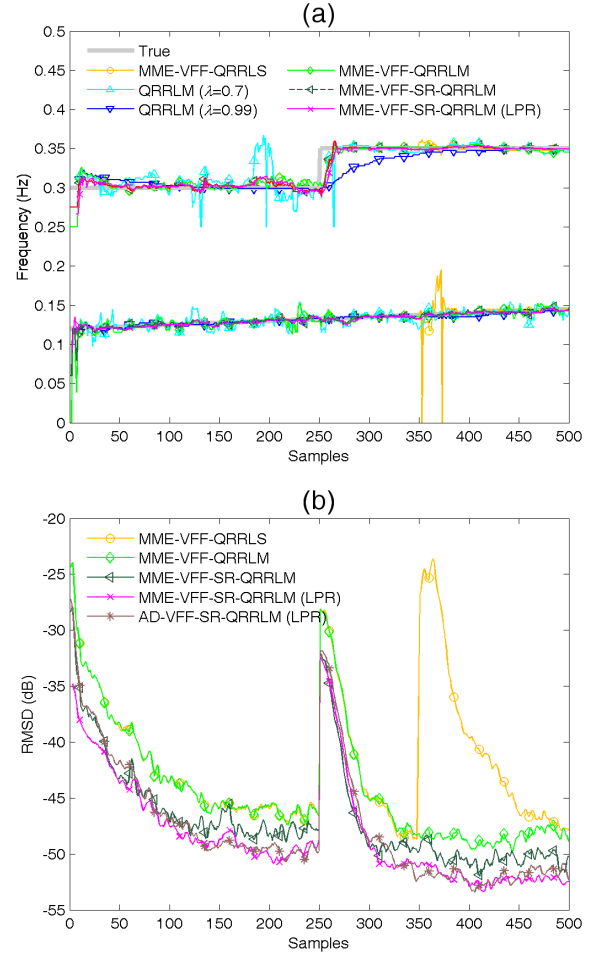


Fig. 7. (a) Extracted spectral peaks obtained from various recursive spectrum tracking methods. (b) RMSD curves of various recursive spectrum tracking methods.

where $\hat{P}(n, \omega)$ is the estimated spectrum and N_{ω} is the number of frequency bins evaluated. The RMSD curves from averages of 100 independent Monte Carlo runs are shown in Fig. 6. In each independent run, one impulse with SIR of -20 dB is added at the 350th samples for illustration.

We can see from Figs. 5 and 6 that the spectrum of the VFF-QRRLS method is seriously degraded by the impulsive component after 350th sample. The QRRLM method can effectively restrain the adverse influence of impulses on the spectrum estimated, but its performance still heavily depends on the FF. A small FF leads to a large variance in the spectrum estimates, but it can track the sharp changes of spectrum rapidly. On the contrary, a large FF has a small estimation variance, which is especially noticeable at the steady state, but it blurs the time resolution of the fast-varying spectrum (~ 250 th sample). The proposed VFF-QRRLM method has a satisfactory performance in the example, thanks to its advantages of robustness and adaptability. We can also see that the VFF-SR-QRRLM method can further stabilize the estimation especially when the signal level is small (~ 100 – 150 ms in this example). From Fig. 6, we can further see that the VFF-SR-QRRLM method can achieve better performance than other methods under test while the two VFF schemes (MME and AD) have comparable results. We can notice from Fig. 6 that the performance of

some methods in ARMA estimation become worse (diverges to some extent) after a largely deviated estimate (such as VFF-QRRLS after the impulse and QRRLM with a FF of 0.99 after the sharp spectrum change). The reason is that such ARMA methods use the estimation residual as regression vector for subsequent estimation. Thus, the adverse effects of deviated estimates will propagate over a long period.

Next, we extract the frequencies of two spectral peaks by detecting the maximum values in two frequency ranges (0–0.25 Hz) and (0.25–5 Hz) and compare the estimation accuracy of extracted frequencies obtained using different recursive spectrum tracking algorithms. It can be seen from Fig. 7 that: 1) the VFF-QRRLM algorithms can achieve better performance than the VFF-QRRLS algorithm and the QRRLM algorithm with a fixed FF; 2) two VFF schemes (MME and AD) have similar results; and 3) the LPR smoothing can lead to a better accuracy. The proposed VFF-QRRLM method can also yield improved performance in its applications to recursive fault detection and speech analysis, and the relevant results can be found in the supplementary materials (http://www.eee.hku.hk/~zgzhang/publication/tim2013_supp.pdf).

V. CONCLUSION

A general parametric model-based recursive frequency/spectrum estimation approach for nonstationary signals in presence of impulsive components is proposed in this paper. A TVLM is used to describe the time-varying frequency or spectrum, and a VFF-QRRLM method is proposed to identify the TVLM. The VFF-QRRLM method is capable of achieving time-varying frequency/spectrum estimates with: 1) adaptability to different extents of frequency/spectrum variations and 2) robustness against impulsive components. The former is achieved by the VFF control scheme while the latter is due to the M-estimation used. It is expected that this robust recursive frequency/spectrum estimation method will find applications in online frequency/spectrum estimation and its potential applications in power quality monitoring have been illustrated by simulation.

APPENDIX

PERFORMANCE ANALYSIS OF MME-VFF

We show that the performance of the proposed MME-VFF scheme is close to that of the optimal FF parameter in the sense that it minimizes the sum of squared prediction error [37]. In other words, the FFs selected by the MME-VFF scheme can achieve nearly optimal results.

A. $\Theta_0(\lambda)$: Ideal Measure for Selecting the Optimal FF

We first derive the ideal measure (the sum of squared prediction error) for selecting the optimal FF. Consider the TVLM $y(n) = x^T(n)\beta(n) + e(n)$ with $n = 1, \dots, N$ samples in presence of the CG noise of (14). In the MMEVFF scheme, the following sum of M-estimate weighted prediction errors is

minimized

$$\begin{aligned}\Theta(\lambda) &= \sum_{n=1}^N \tilde{q}(n)[y(n) - \hat{y}_\lambda(n)]^2 \\ &= \sum_{n=1}^N \tilde{q}(n)[y_0(n) - \hat{y}_\lambda(n) + e(n)]^2\end{aligned}\quad (A1)$$

where $y_0(n) = x^T(n)\beta(n)$ is the noise-free measurement, $\hat{y}_\lambda(n) = x^T(n)\hat{\beta}_\lambda(n-1)$ is a *a priori* estimate of $y_0(n)$ based on λ , $e(n)$ is the additive noise, and $\tilde{q}(n) = q(\hat{e}(n)) = \rho'(\hat{e}(n))/\hat{e}(n)$ is the M-estimate weight ($0 < \tilde{q}(n) \leq 1$). In the ideal case, the noise $e(n)$ is absent and $q(n) = 1$, then $\Theta_0(\lambda) = \sum_{n=1}^N [y_0(n) - \hat{y}_\lambda(n)]^2$ is a reasonable measure for selecting the optimal FF since it measures the squared error between the true signal and the one predicted by the given FF.

B. $\Theta'_0(\lambda)\Theta_0(\lambda)$: Robust Counterpart for Selecting the Optimal FF in the Presence of CG Noise

We show that the difference between $\Theta(\lambda)$ and the ideal measure $\Theta_0(\lambda)$ is small for a sufficiently large N even in presence of noise $e(n)$. To this end, we first rewrite (A1) as

$$\Theta(\lambda) = \Theta'_0(\lambda) + \Theta_C(\lambda) + \Theta_E \quad (A2)$$

where $\Theta'_0(\lambda) = \sum_{n=1}^N \tilde{q}(n)[y_0(n) - \hat{y}_\lambda(n)]^2$, $\Theta_C(\lambda) = 2 \sum_{n=1}^N \tilde{q}(n)[y_0(n) - \hat{y}_\lambda(n)]e(n)$, and $\Theta_E = \sum_{n=1}^N \tilde{q}(n)e^2(n)$. We now check the three terms respectively to see how they influence the selection of the optimal FF.

Note that, $q(n)$ will normally (with a probability of $1 - p_r$) assume a value of one, except when an exceptional large value of $e(n)$ (impulsive component) is encountered so that its value will be much smaller than one or even equal to zero. Therefore, we can assume that $q(n)$ are independent of λ . Consequently, Θ_E is independent of λ and can be removed from the objective function of (A2) for finding the optimal λ . Thus, minimizing $\Theta(\lambda)$ is equivalent to minimizing $\Theta_\Sigma(\lambda) = \Theta'_0(\lambda) + \Theta_C(\lambda)$. Again, since $\tilde{q}(n)$ depends mainly on $e(n)$, $\Theta'_0(\lambda)$ approximately retains the reliable measurements and computes the corresponding sum of square errors between the true model and the predicted model. Hence, $\Theta'_0(\lambda)$ is a robust counterpart of $\Theta_0(\lambda)$ for selecting λ in presence of CG noise.

If we can prove that $\Theta_C(\lambda)$ is small compared with $\Theta'_0(\lambda)$ for a sufficiently large value of N , then minimizing $\Theta(\lambda)$ (or $\Theta_\Sigma(\lambda)$) can give nearly the same performance as minimizing $\Theta'_0(\lambda)$, which allows us to obtain a FF nearly as good as the optimal one obtained from $\Theta_0(\lambda)$. In [37], it is shown that this is in fact the case for the LS criterion. Here, we shall extend this analysis to the case of M-estimation. To achieve this, we will derive the upper bound of $\Theta_C(\lambda)$, and then prove that the difference between $\Theta'_0(\lambda)$ obtained from MME-VFF and that from optimal FF is small.

C. Upper Bound of $\Theta_C(\lambda)$ in Terms of $\Theta_0(\lambda)$

The main idea, following [37], is to bound $\Theta_C(\lambda)$ by employing recent results in deviation probability of martin-gales [43]. In particular, we note the cross term $\Theta_C(\lambda)$ is a

square integrable martingale

$$M_\lambda(n) = \sum_{i=1}^n \tilde{q}(i)[y_0(i) - \hat{y}_\lambda(i)]e(i) \quad (\text{A3})$$

with zero mean and finite power Gaussian increments and stopping time N . Note, $e(i)$ may have different power at different time instants. To model the impulsive noise, we assume that there are on average $p_r n$ such $e(i)$'s that have a large variance of σ_{im}^2 and others are nominal samples with noise variance σ_g^2 . The predictable quadratic variation $V_\lambda^2(n)$ of $M_\lambda(n)$, which is defined via its increments $\zeta(n) = M_\lambda(n) - M_\lambda(n-1)$, is given by

$$\begin{aligned} V_\lambda^2(n) &= \sum_{i=1}^n E[\zeta^2(i)|F_{i-1}] \\ &= \sum_{i=1}^n [y_0(i) - \hat{y}_\lambda(i)]^2 E[\tilde{q}^2(i)e^2(i)]. \end{aligned} \quad (\text{A4})$$

Next, we shall show that $V_\lambda^2(n)$ is bounded by $\Theta_0(\lambda)$.

If adaptive threshold selection is employed, then $\zeta_j(i) = k_{\zeta_j} \hat{\sigma}_e(i)$, where $\hat{\sigma}_e(i)$ is the standard deviation of the estimated impulse free (or nominal) noise. Then, we get $E[\tilde{q}^2(i)e^2(i)] = 1/\sqrt{2\pi} \sigma_e(i) \int_{-\infty}^{\infty} e^2 q^2(e + \delta(i)/k_{\zeta_i} \hat{\sigma}_e(i)) \exp(-e^2/2\sigma_e^2(i)) de$, where $\delta(i) = y_0(i) - \hat{y}_\lambda(i)$ and $\sigma_e^2(i)$ is the variance of $e(i)$. Using the change of variable $u = e/\sigma_e(i)$, we get $E[\tilde{q}^2(i)e^2(i)] = \sigma_e^2(i)/\sqrt{2\pi} \int_{-\infty}^{\infty} u^2 q^2(u\sigma_e(i) + \delta(i)/k_{\zeta_i} \hat{\sigma}_e(i)) \exp(-u^2/2) du$. Since $\hat{\sigma}_e^2(i)$ is a robust estimator of $(\bar{\delta}^2(i) + \sigma_{e_g}^2)$, where $\bar{\delta}^2(i)$ is the mean of $\delta^2(i)$ and $\sigma_{e_g}^2$ is the nominal Gaussian noise, we have for $\sigma_e(i) = \kappa \sigma_{e_g}$, that $E[\tilde{q}^2(i)e^2(i)] = S_\psi(\sigma_e^2(i))\sigma_e^2(i)$, where $S_\psi(\sigma_e^2(i)) = 1/\sqrt{2\pi} \int_{-\infty}^{\infty} u^2 q^2(\kappa u \sigma_{e_g} + \delta(i)/\kappa/k_{\zeta_i} \hat{\sigma}_e(i)) \exp(-u^2/2) du$. For nominal Gaussian noise, κ is nearly equal to 1 and k_ζ is chosen so that most of the area under the exponential function is covered and hence $S_\psi(\sigma_e^2(i)) \equiv S_\psi \approx 1$. For the impulsive component, $\kappa \gg 1$ and $S_\psi(\sigma_e^2(i))$ will be nearly zero. Provided that the variance of the impulses σ_{im}^2 is sufficiently large, $S_\psi(\sigma_e^2(i))$ is independent of $\bar{\delta}^2(i)$ and hence the FF. Consequently, we have the following:

$$V_\lambda^2(n) = \sum_{i=1}^n [y_0(i) - \hat{y}_\lambda(i)]^2 S_\psi(\sigma_e^2(i))\sigma_e^2(i). \quad (\text{A5})$$

For LS criterion, $S_\psi(\sigma_e^2(i)) = 1$.

Since $S_\psi(\sigma_e^2(i))\sigma_e^2(i)$ is finite, there exists a constant v^2 such that $S_\psi(\sigma_e^2(i))\sigma_e^2(i) \leq v^2$. Substituting into (A4) gives

$$V_\lambda^2(N) \leq v^2 \Theta_0(\lambda). \quad (\text{A6})$$

Equation (A6) will be used later to prove that $\Theta_C(\lambda)$ is bounded by $\sqrt{\Theta_0(\lambda)}$ in (A9).

To proceed further, we show that $\Theta_C(\lambda)$ is bounded by $\sqrt{\Theta_0(\lambda)}$ by making use of the following theorem from [43] (the proof is omitted for simplicity and the readers are referred to [43] for details).

Theorem 1 (Analogous to Theorem 2.1 in [43]): Let $M(n)$ be a discretetime martingale with conditionally Gaussian

increments and predictable quadratic variation $V(n)$, and N be fixed or the stopping time. For deterministic constants $b > 0$, $S \geq 0$, and $\gamma \geq 1$, we have the following:

$$P\{|M(N)| > \gamma V(N), b \leq V(N) \leq bS\} \leq \alpha_\lambda(\gamma) \quad (\text{A7})$$

where $\alpha_\lambda(\gamma) = 4\sqrt{e}(1 + \log S)\gamma \exp[-\gamma^2/(2)]$.

Applying Theorem 1 (A7) to $M_\lambda(n) = \Theta_C(\lambda)$ for all candidate λ 's in the set Λ , we get the following:

$$\begin{aligned} \sum_{\lambda \in \Lambda} P\{\Theta_C(N) > 2\gamma V_\lambda(N), b \leq V_\lambda(N) \leq bS\} \\ \leq \sum_{\lambda \in \Lambda} \alpha_\lambda(\gamma). \end{aligned} \quad (\text{A8})$$

Using (A6) and the definition of $\Theta_C(\lambda)$ in (A2), it follows from (A8) that the event:

$$\Theta_C(\lambda) \leq 2\gamma v \sqrt{\Theta_0(\lambda)} \quad \forall \lambda \in \Lambda \quad (\text{A9})$$

has a probability at least $1 - \sum_{\lambda \in \Lambda} \alpha_\lambda(\gamma)$. Equation (A9) will be used later in (A12) and (A13) to prove that the difference between $\Theta'_0(\lambda)$ obtained from MME-VFF and that from optimal FFs is small.

D. Upper Bound of $\Theta_C(\lambda)$ in Terms of $\Theta'_0(\lambda)$

Next, denote $\Delta\Theta'_0(\lambda)$ as the reduction of the ideal measure $\Theta_0(\lambda)$ due to robust estimation in suppressing the impulses [i.e., $\Delta\Theta'_0(\lambda) = \Theta_0(\lambda) - \Theta'_0(\lambda)$], we consider the event $B = \{\Delta\Theta'_0(\lambda) \leq c\Theta'_0(\lambda)\}$ for some finite non-negative constant c . The purpose of considering the event B is to avoid the pathological case that all the measurements are removed for small n due to small but finite probability variations. The condition of B is equivalent to $\sum_{n=1}^N (1 - \tilde{q}(n))[y_0(n) - \hat{y}_\lambda(n)]^2 \leq c \sum_{n=1}^N \tilde{q}(n)[y_0(n) - \hat{y}_\lambda(n)]^2$. As long as $\tilde{q}(n)$ is not identical to zero, there exists a nonnegative constant c such that the condition is satisfied. As $N \rightarrow \infty$, $P(B) \rightarrow 1$. The value of c mainly depends on the occurrence probability of impulses p_r and depends weakly on N for a sufficiently large N . To prove the latter, we note from the law of large number that

$$\begin{aligned} \sum_{n=1}^N (1 - \tilde{q}(n))[y_0(n) - \hat{y}_\lambda(n)]^2 &\rightarrow \Theta_0(\lambda) E[1 - \tilde{q}(n)] \\ &\text{as } N \rightarrow \infty \end{aligned} \quad (\text{A10})$$

and

$$\begin{aligned} \sum_{n=1}^N \tilde{q}(n)[y_0(n) - \hat{y}_\lambda(n)]^2 &\rightarrow \Theta_0(\lambda) E[\tilde{q}(n)] \\ &\text{as } N \rightarrow \infty. \end{aligned} \quad (\text{A11})$$

Therefore, the minimum value of c is $c' = (1 - E[\tilde{q}(n)])/E[\tilde{q}(n)]$, which is independent of N . Under the condition B , we have $\Theta'_0(\lambda) \leq \Theta_0(\lambda) \leq \Theta'_0(\lambda)(1 + c')$. Thus, we can get from (A9) that

$$\Theta_C(\lambda) \leq 2\gamma v \sqrt{1 + c'} \sqrt{\Theta'_0(\lambda)}. \quad (\text{A12})$$

E. Upper and Lower Bounds of $\Theta_\Sigma(\lambda)$ in Terms of $\Theta'_0(\lambda)$

Subsequently, using (A12), we get

$$\Theta_\Sigma(\lambda) = \Theta'_0(\lambda) + \Theta_C(\lambda) \leq \Theta'_0(\lambda) + 2\gamma v \sqrt{1+c'} \sqrt{\Theta'_0(\lambda)} + \gamma^2 v^2 (1+c'). \quad (\text{A13})$$

Taking square root on both sides of (A13) gives

$$\sqrt{\Theta_\Sigma(\lambda)} \leq \sqrt{\Theta'_0(\lambda)} + \sqrt{1+c'} \gamma v. \quad (\text{A14})$$

On the other hand, it can be seen that

$$\Theta_\Sigma(\lambda) \geq \Theta'_0(\lambda) - 2\gamma v \sqrt{1+c'} \sqrt{\Theta'_0(\lambda)}. \quad (\text{A15})$$

Using the identity, $\sqrt{a-b} \geq \sqrt{a} - b/\sqrt{a}$, for all positive a and b , it follows that:

$$\sqrt{\Theta_\Sigma(\lambda)} \geq \sqrt{\Theta'_0(\lambda) - 2\gamma v \sqrt{1+c'} \sqrt{\Theta'_0(\lambda)}} \geq \Theta'_0(\lambda) - 2\gamma v \sqrt{1+c'}. \quad (\text{A16})$$

Combining (A15) and (A16), we get the upper and lower bounds for $\sqrt{\Theta_\Sigma(\lambda)}$ as

$$\begin{aligned} \sqrt{\Theta'_0(\lambda)} - 2\gamma v \sqrt{1+c'} &\leq \sqrt{\Theta_\Sigma(\lambda)} \\ &\leq \sqrt{\Theta'_0(\lambda)} + \gamma v \sqrt{1+c'}, \text{ for each } \lambda \in \Lambda \Theta'_0(\lambda). \end{aligned} \quad (\text{A17})$$

F. Difference Between Obtained from MME-VFF and That from Optimal FFs is Small

Suppose the FF that minimizes $\Theta_\Sigma(\lambda) = \Theta'_0(\lambda) + \Theta_C(\lambda)$ is λ_Σ (i.e., the FF selected by MME-VFF) while the FF that minimizes $\Theta'_0(\lambda)$ is λ_0 (i.e., the optimal FF), we have $\Theta_\Sigma(\lambda_\Sigma) \leq \Theta_\Sigma(\lambda_0)$. Using (A17), we get the following upper bound of $\Theta'_0(\lambda_\Sigma)$:

$$\begin{aligned} \sqrt{\Theta'_0(\lambda_\Sigma)} &\leq \sqrt{\Theta_\Sigma(\lambda_\Sigma)} + 2\gamma v \sqrt{1+c'} \leq \sqrt{\Theta_\Sigma(\lambda_0)} \\ &\quad + 2\gamma v \sqrt{1+c'} \leq \sqrt{\Theta'_0(\lambda_0)} + 3\gamma v \sqrt{1+c'}. \end{aligned} \quad (\text{A18})$$

The deviation of $\Theta'_0(\lambda_\Sigma)$ from the optimal $\Theta'_0(\lambda_0)$ can be summarized as follows. It holds for every $\gamma > 1$ and every $v > 0$

$$\mathbb{P} \left\{ \sqrt{\Theta'_0(\lambda_\Sigma)} \geq \sqrt{\Theta'_0(\lambda_0)} + 3\gamma v \sqrt{1+c'}, A \right\} \leq \sum_{\lambda \in \Lambda} \alpha_\lambda(\gamma)$$

with $A = \bigcap_{\lambda \in \Lambda} \{b \leq V_\lambda(N) \leq bS\} \cap \{V_\lambda^2(N) \leq v^2 \Theta_0(\lambda)\} \cap B$, which is analogous to Theorem 1 in [37].

If $\gamma > \sqrt{2ar \log N}$ for a given positive r , then $\alpha_\lambda(\gamma) = o(N^{-r})$. When the number of elements in Λ is of order $O(N^r)$, $\sum_{\lambda \in \Lambda} \alpha_\lambda(\gamma) \rightarrow 0$. Thus, the extra term $3\gamma v \sqrt{1+c'}$ compared with $\Theta'_0(\lambda_0)$ is only of order $O(\log(N)^{1/2})$, which is small as compared with $\Theta'_0(\lambda_0)$ and can be neglected.

To conclude, among the three terms of the objective function $\Theta(\lambda)$ of (A1), $\Theta_C(\lambda)$ is sufficiently small compared with $\Theta'_0(\lambda)$ and Θ_E is independent of λ . Therefore, the data-driven optimal λ_Σ that minimizes $\Theta(\lambda)$ can achieve nearly as good performance as the theoretically optimal λ_0 that minimizes $\Theta'_0(\lambda)$.

G. Effect of Impulsive Noise on MME-VFF

Finally, we examine the effect of impulsive noise and robust estimation on the MME-VFF scheme. In the LS criterion, $\tilde{q}(n)$ and $S_\psi(\sigma_e^2(i))$ are equal to one and hence $v^2 = \max\{S_\psi(\sigma_e^2(i))\sigma_e^2(i)\} = \sigma_{im}^2$. Since the deviation of $\sqrt{\Theta_\Sigma(\lambda_\Sigma)}$ from the optimal measure $\sqrt{\Theta_\Sigma(\lambda_0)}$ is proportional to v as depicted in (A18) we expect that the impulsive noise has a substantial effect on the selection process. More samples are thus needed to obtain the right FF. On the other hand, for M-estimation $v^2 = \max\{S_\psi(\sigma_e^2(i))\sigma_e^2(i)\} \approx \sigma_g^2$, since $S_\psi(\sigma_e^2(i))$ will become zero for large $\sigma_e^2(i)$ and hence the effect is effectively mitigated. An unavoidable consequence is that more samples are also needed as those corrupted samples are discarded during the estimation. The estimator itself is however, safeguarded from the adverse influence of σ_{im}^2 , which could be arbitrarily large.

ACKNOWLEDGMENT

The authors would like to thank Dr. Y. H. Hou and Dr. X. Chen, who are with the Department of Electric and Electronic Engineering, the University of Hong Kong, Hong Kong, for providing them with the power system signals simulated by PSCAD/EMTDC. The authors would also like to thank the anonymous reviewers for their kind and constructive comments which helped to enhance this paper.

REFERENCES

- [1] M. D. Kusljevic, "A simple recursive algorithm for simultaneous magnitude and frequency estimation," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 6, pp. 1207–1214, Jun. 2008.
- [2] A. Watanabe, "Formant estimation method using inverse-filter control," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 4, pp. 317–326, Nov. 2001.
- [3] J. F. Thayer, J. J. Sollers, E. R. Padial, and J. Vila, "Estimating respiratory frequency from autoregressive spectral analysis of heart period," *IEEE Eng. Med. Biol. Mag.*, vol. 21, no. 4, pp. 41–45, Apr. 2002.
- [4] I. Santamaría-Caballero, C. J. Pantaleón-Prieto, J. Ibáñez-Díaz, and E. Gómez-Cosío, "Improved procedures for estimating amplitudes and phases of harmonics with application to vibration analysis," *IEEE Trans. Instrum. Meas.*, vol. 47, no. 1, pp. 209–214, Feb. 1998.
- [5] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Reading, MA, USA: Addison-Wesley, Jul. 1991.
- [6] J. G. Proakis, C. M. Rader, F. Ling, M. Moonen, I. K. Proudler, and C. L. Nikias, *Algorithms for Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
- [7] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.
- [8] I. Y.-H. Gu and M. H. J. Bollen, "Estimating interharmonics by using sliding-window ESPRIT," *IEEE Trans. Power Delivery*, vol. 23, no. 1, pp. 13–23, Jan. 2008.
- [9] A. Bracale, G. Carpinelli, I. Y.-H. Gu, and M. H. J. Bollen, "A new joint sliding-window ESPRIT and DFT scheme for waveform distortion assessment in power systems," *Electr. Power Syst. Res.*, vol. 88, pp. 112–120, Jul. 2012.
- [10] 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.
- [11] 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.
- [12] Z. G. Zhang, K. M. Tsui, S. C. Chan, W. Y. Lau, and M. Aboiy, "A novel method for non-stationary power spectral density estimation of cardiovascular signals based on the Kalman filter," *Med. Biol. Eng. Comput.*, vol. 46, no. 8, pp. 789–797, Aug. 2008.

- [13] Z. G. Zhang, H. T. Liu, S. C. Chan, K. D. K. Luk, and Y. Hu, "Time-dependent power spectral density estimation of surface electromyography during isometric muscle contraction: Methods and comparisons," *J. Electromyogr. Kinesiol.*, vol. 20, no. 1, pp. 89–101, Feb. 2010.
- [14] Z. G. Zhang, S. C. Chan, and K. M. Tsui, "A recursive frequency estimator using linear prediction and a Kalman-filter-based iterative algorithm," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 55, no. 6, pp. 576–580, Jun. 2008.
- [15] P. K. Dash, D. P. Swain, A. Routray, and A. C. Liew, "An adaptive neural network approach for the estimation of power system frequency," *Electr. Power Syst. Res.*, vol. 41, no. 3, pp. 203–210, Jun. 1997.
- [16] R. D. Martin and D. J. Thomson, "Robust-resistant spectrum estimation," *Proc. IEEE*, vol. 70, no. 9, pp. 1097–1115, Sep. 1982.
- [17] M. Niedzwiecki and K. Cisowski, "Adaptive scheme for elimination of broadband noise and impulsive disturbances from AR and ARMA signals," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 528–537, Mar. 1996.
- [18] S. J. Godsill and P. J. W. Rayner, "Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 352–372, Jul. 1998.
- [19] T. Cai, S. X. Duan, T. Ren, and F. R. Liu, "A robust parametric method for power harmonic estimation based on M-estimators," *Measurement*, vol. 43, no. 1, pp. 67–77, Jan. 2010.
- [20] *IEEE Recommended Practice for Monitoring Electric Power Quality*, IEEE Standard 1159, 1995.
- [21] 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.
- [22] S. V. Vaseghi and P. J. W. Rayner, "Detection and suppression of impulsive noise in speech communication systems," in *IEE Proc.*, vol. 137, no. 1, pp. 38–46, Feb. 1990.
- [23] S. V. Vaseghi and B. P. Milner, "Speech recognition in impulsive noise," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, May 1995, pp. 437–440.
- [24] 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.
- [25] W. Y. Lau, S. C. Chan, Z. G. Zhang, and C. H. Leung, "A new QR-decomposition based recursive frequency estimator for multiple sinusoids in impulsive noise environment," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2006, pp. 4248–4251.
- [26] S. C. Chan and Y. J. Chu, "A new state-regularized QRRLS algorithm with a variable forgetting factor," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 59, no. 3, pp. 183–187, Mar. 2012.
- [27] 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 multiresolution analysis of non-uniform data," *J. Signal Process. Syst.*, vol. 52, no. 3, pp. 263–280, Sep. 2008.
- [28] H. C. So, K. W. Chan, Y. T. Chan, and K. C. Ho, "Linear prediction approach for efficient frequency estimation of multiple real sinusoids: Algorithms and analyses," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2290–2305, Jul. 2005.
- [29] Y. T. Chan, J. M. M. Lavoie, and J. B. Plant, "A parameter estimation approach to estimation of frequencies of sinusoids," *IEEE Trans. Acoust. Speech, Signal Process.*, vol. 29, no. 2, pp. 214–219, Apr. 1981.
- [30] J. S. Mao, S. C. Chan, W. Liu, and K. L. Ho, "Design and multiplier-less implementation of a class of two-channel PR FIR filterbanks and wavelets with low system delay," *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3379–3394, Dec. 2000.
- [31] S. C. Chan, C. K. S. Pun, and K. L. Ho, "New design and realization techniques for a class of perfect reconstruction two-channel FIR filterbanks and wavelets bases," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 2135–2141, Jul. 2004.
- [32] J. Ren and M. Kezunovic, "Real time power system frequency and phasor estimation scheme using recursive wavelet transform," *IEEE Trans. Power Delivery*, vol. 26, no. 3, pp. 1392–1402, Jul. 2011.
- [33] J. Ren and M. Kezunovic, "An adaptive phasor estimator for power system waveforms containing transients," *IEEE Trans. Power Delivery*, vol. 27, no. 2, pp. 735–745, Apr. 2012.
- [34] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics: Theory and Methods*. New York, NY, USA: Wiley, 2006.
- [35] L. Ljung, *System Identification: Theory for the User*. Upper Saddle River, NJ, USA: Prentice-Hall, 1999.
- [36] S. C. Chan, Y. J. Chu, Z. G. Zhang, and K. M. Tsui, "A new variable regularized QR decomposition-based recursive least M-estimate algorithm—Performance analysis and acoustic applications," *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 5, pp. 907–922, Dec. 2000.
- [37] M. Y. Cheng, J. Q. Fan, and V. Spokoiny, "Dynamic nonparametric filtering with application to volatility estimation," in *Recent Advances and Trends in Nonparametric Statistics*, M. G. Akritas and D. N. Politis, Eds. Amsterdam, The Netherlands: Elsevier, 2003, pp. 315–333.
- [38] B. Liao, Z. G. Zhang, and S. C. Chan, "A new robust Kalman filter-based subspace tracking algorithm in impulsive noise environment," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 57, no. 9, pp. 740–744, Sep. 2010.
- [39] S. K. Jain and S. N. Singh, "Exact model order ESPRIT technique for harmonics and interharmonics estimation," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 7, pp. 1915–1923, Jul. 2012.
- [40] (2010). *NIST/SEMATECH E-Handbook of Statistical Methods* [Online]. Available: <http://www.itl.nist.gov/div898/handbook/>
- [41] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. London, U.K.: Chapman and Hall, 1996.
- [42] V. Katkovnik, "A new method for varying adaptive bandwidth selection," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2567–2571, Sep. 1999.
- [43] R. Lipster and V. Spokoiny, "Deviation probability bound for martingales with applications to statistical estimation," *Stat. Probab. Lett.*, vol. 46, no. 4, pp. 347–357, Feb. 2000.



Zhi Guo 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, Hong Kong, in 2008.

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



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

He joined the City Polytechnic of Hong Kong, 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, and is currently a Professor. He was a Visiting Researcher with Microsoft Corporation, Redmond, WA, USA, Microsoft, Beijing, China, University of Texas at Arlington, Arlington, TX, USA, and Nanyang Technological University, Singapore. His current 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.