Asymptotic Analysis of a Fast Algorithm for Efficient Multiple Frequency Estimation[1]

Ta-Hsin Li[2], *Member, IEEE,* and Kai-Sheng Song[3]

Published in

IEEE Transactions on Information Theory

vol. 48, no. 10, pp. 2709–2720, 2002.

[1]This work was supported in part by the National Science Foundation under Grant DMS 9817552. The authors thank the associate editor and the anonymous reviewers for helpful comments and suggestions.

[2]T. H. Li was with the Department of Statistics and Applied Probability, University of California at Santa Barbara. He is now with the Department of Mathematical Sciences, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: thl@watson.ibm.com).

[3]K. S. Song is with the Department of Statistics, Florida State University, Tallahassee, FL 32306 USA (e-mail: kssong@stat.fsu.edu).

Abstract

Based on an asymptotic analysis of the contraction mapping (CM) method of Li and Kedem (*IEEE Trans. Inform. Theory*, vol. 39, pp. 989–998, 1993), a bandwidth shrinkage rule is proposed for fast and accurate estimation of the frequencies of multiple sinusoids from noisy measurements. The CM frequency estimates are defined as the fixed-points of a contractive mapping formed by the lag-one autocorrelation coefficient calculated from the output of a parametric filter applied to the observed time series. With judiciously chosen bandwidth parameters according to the asymptotic analysis, the algorithm is shown to be able to accommodate possibly poor initial values of precision $\mathcal{O}(n^{-1/3})$ and converge to a final estimate whose accuracy is arbitrarily close to $\mathcal{O}(n^{-3/2})$, the optimal error rate for frequency estimation under the Gaussian assumption. The total computational complexity of the algorithm is shown to be $\mathcal{O}(n \log n)$, which is comparable to that of n-point FFT. A novelty in the asymptotic analysis is that it accommodates closely-spaced frequencies by allowing not only the filter bandwidth but also the frequency separation to be functions of the sample size n. This enables an assessment of the accuracy of the frequency estimates for given bandwidths and initial values in situations where some or all of the frequencies are close to each other.

Key words and phrases. Autoregressive, filter, fixed-points, harmonic analysis, martingale difference, mixed spectrum, nonlinear regression, signal processing, spectral analysis.

Running Head. Multiple Frequency Estimation

I. INTRODUCTION

Consider a time series $\{y_1, \cdots, y_n\}$ obtained from the following random process:

$$y_t := \sum_{k=1}^{p} \beta_k \cos(\omega_k t + \phi_k) + \varepsilon_t, \tag{1}$$

where β_k, ω_k, and ϕ_k are unknown constants satisfying $\beta_k > 0$, $0 < \omega_1 < \cdots < \omega_p < \pi$, and $\phi_k \in (-\pi, \pi]$, and $\{\varepsilon_t\}$ is a zero-mean stationary process. This 'multiple-sinusoid-plus-noise' model has important scientific and engineering applications in, for example, radar and sonar signal processing and rotating machinery.

A fundamental problem in these applications is to accurately estimate the unknown frequencies ω_k. In particular, an accuracy of $o_P(n^{-1})$ is required for reliable assessment of the amplitudes of the sinusoids, as demonstrated in [1] and [2]. Traditional methods of obtaining such accurate frequency estimates include the maximization of periodogram (MP) as a continuous function of the frequency variable and the minimization of the error sum of squares by nonlinear least-squares (NLS) regression (which coincides with the maximum likelihood method if $\{\varepsilon_t\}$ is Gaussian white noise). Both MP and NLS are statistically efficient for frequency estimation in the sense that the estimation error achieves asymptotically the Cramér-Rao lower bound (derived under the Gaussian white noise assumption) that can be expressed as $\mathcal{O}_P(n^{-3/2})$ (e.g., [3]–[7]). Unfortunately, the computational requirements of these methods are quite prohibitive, not only because iterative optimization algorithms are needed to compute the estimates, but more importantly because extremely precise initial values, typically of accuracy $o(n^{-1})$, which cannot be obtained by n-point FFT, are required to ensure convergence (e.g., [1], [4], [8], and [9]). Furthermore, the MP and NLS estimates cannot be easily updated upon the arrival of new observations without re-processing the entire data record. These considerations have motivated the proposal of many alternative methods in both statistical and signal processing literature.

Iterative filtering (IF) is a favorite approach in signal processing to developing alternative methods of frequency estimation that are computationally efficient (e.g., [10]–[15]). A typical IF algorithm repeats the steps of enhancing the sinusoids with a bandpass filter and estimating the frequencies on the basis of the filtered data. Since recursive filters are often employed by IF algorithms, the frequency estimates can be easily updated upon the arrival of new observations in order to track possible frequency changes (e.g., [16]–[19]). The general premise of the IF approach is that as the frequency estimates become more accurate, the filter, which depends on the frequency estimates, would enhance the sinusoids more effectively and thus further improve the precision of frequency estimation in the subsequent cycle of iteration. This, indeed, has been vindicated by many numerical studies in the literature. What remains largely an open question is how to design the filter on the basis of the available frequency estimates so that the entire iterative scheme would converge to a solution of improved accuracy.

Because the sinusoids are localized in the frequency domain, bandpass filters are often employed to enhance them. One can use a filter with multiple passbands to simultaneously estimate all frequencies (e.g., [14] and [20]), or a filter with single passband to sequentially estimate each frequency (e.g., [21]). The first approach may have higher frequency resolution, as indicated by many numerical studies, but at the expense of greater computational complexity. In this paper, we focus on the second approach.

In essence, the second approach is an application of single-frequency estimation methods to the multiple frequency case by regarding all but one sinusoids as interference and lumping them into the noise term in (1). Since this approach relies on the bandpass filter to suppress both the noise and the interfering sinusoids, the bandwidth selection becomes an important issue. If the bandwidth is too large, the noise and the interfering sinusoids would not be effectively suppressed and the resulting frequency estimates would be inaccurate. On the other hand, if the bandwidth is too small, the desired sinusoid could be filtered out by a filter designed on the basis of poor frequency estimates and the iteration would not converge to the desired solution.

The main contribution of this paper is to analytically quantify the role of bandwidth in determining the required initial precision that ensures the convergence of an IF algorithm and the accuracy of the final frequency estimates after convergence. The IF algorithm that we focus on in this paper is the contraction mapping (CM) method of Li and Kedem [22]. This method employs a second-order autoregressive (AR) filter endowed with a bandwidth parameter (for other filters, see [23]–[25]). Statistical and numerical properties of the CM method in the case of single sinusoid have been studied by Li and Kedem [22], Li, Kedem, and Yakowitz [26], and most recently, by Song and Li [2], [27]. These studies show that if the bandwidth is judiciously adjusted with the iteration, the CM method can accommodate poor initial guesses of accuracy $\mathcal{O}_P(1)$ and converge to a final frequency estimate whose accuracy is arbitrarily close to $\mathcal{O}_P(n^{-3/2})$.

To investigate the CM method in the case of multiple sinusoids, one has to overcome two major obstacles. First, the interfering sinusoids have very different statistical properties from the noise (e.g., discrete versus continuous spectrum). Second, the interfering frequencies may reside in a close vicinity of the frequency to be estimated. To deal with the first problem, the interaction of the sinusoids among themselves and with the noise has to be carefully evaluated. To accommodate the second problem, we assume that the minimum distance among the frequencies may depend on the sample size n and may decrease to zero as n tends to infinity. Under this assumption, the bandwidth must also depend on the separation of the frequencies in order to suppress the interference. Consequently, the required initial precision for the CM iteration to converge depends not only on the bandwidth parameter but also on the frequency separation. It is shown that when the frequencies are not too close to each other (as compared to the filter bandwidth), the CM method retains its capability of producing accurate frequency estimates whose accuracy is arbitrarily close to $\mathcal{O}_P(n^{-3/2})$. The convergence is guaranteed as long as the initial precision is $o_P(n^{-1/3})$. This requirement is easily satisfied by any root-n consistent estimates, including those produced by the multivariate IF method in [14] and by the singular-value-decomposition-based methods such as MUSIC and ESPRIT (e.g., [28]).

The rest of the paper is organized as follows. In Sec. II, we introduce the CM frequency estimator. In Sec. III, we present our main contributions in the form of five theorems and a resulting bandwidth shrinkage rule that leads to a three-step algorithm capable of improving poor initial values of accuracy $o_P(n^{-1/3})$ to produce a final frequency estimator whose accuracy is arbitrarily close to $\mathcal{O}_P(n^{-3/2})$. A simulation example is also given in this section to demonstrate the algorithm. Finally, in Sec. IV, we provide the proof of the theorems on the basis of some preliminary propositions. The proof of these propositions are outlined in Appendix I. Interested readers are referred to [29] for a complete proof of the propositions. Some useful expressions, which are important in the proof of the propositions and interesting on their own right, are given in Appendix II without proof.

II. THE CM FREQUENCY ESTIMATOR

For any given $\eta \in (0,1)$ and $\alpha := \cos\omega \in \mathcal{A} := (-2\eta(1+\eta^2)^{-1}, 2\eta(1+\eta^2)^{-1})$, let $\{y_t(\alpha)\}$ be obtained recursively from the observations $\{y_1,\cdots,y_n\}$ according to

$$y_t(\alpha) + 2\,\theta(\alpha)\,\eta\,y_{t-1}(\alpha) + \eta^2 y_{t-2}(\alpha) = y_t \quad (t=1,\cdots,n), \tag{2}$$

where $y_{-1}(\alpha) = y_0(\alpha) := 0$ and

$$\theta(\alpha) := -\frac{1+\eta^2}{2\eta}\,\alpha := -\cos\lambda. \tag{3}$$

Note that (2) defines a causal stable AR(2) filter with transfer function $(1+2\,\theta(\alpha)\,\eta\,\mathfrak{B}+\eta^2\mathfrak{B}^2)^{-1}$, where $\mathfrak{B}$ is the backward-shift operator such that $\mathfrak{B}y_t = y_{t-1}$. Note also that $\lambda \in (0,\pi)$ in (3) is uniquely determined by $\eta \in (0,1)$ and $\alpha \in \mathcal{A}$.

Let the lag-one autocorrelation coefficient of $\{y_t(\alpha)\}$ be estimated by

$$\rho_n(\alpha) := \frac{\sum_{t=1}^n y_{t-1}(\alpha)\,\{y_t(\alpha)+\eta^2 y_{t-2}(\alpha)\}}{(1+\eta^2)\sum_{t=1}^n y_{t-1}^2(\alpha)}. \tag{4}$$

This estimator minimizes the weighted sum of the forward and backward prediction error sums of squares defined by $e_n^2(\rho) := \sum_{t=1}^n \{y_t(\alpha)-\rho\,y_{t-1}(\alpha)\}^2 + \eta^2\sum_{t=1}^n\{y_{t-2}(\alpha)-\rho\,y_{t-1}(\alpha)\}^2$, where η^2 plays the role of a weight that discounts the contribution of the backward prediction errors. The CM method in [22] produces the frequency estimates from the fixed-point iteration

$$\hat{\alpha}_n^{(m)} := \rho_n(\hat{\alpha}_n^{(m-1)}) \quad (m=1,2,\cdots). \tag{5}$$

Suppose that with an initial guess $\hat{\alpha}_n^{(0)}$ in some neighborhood of $\alpha_k := \cos \omega_k$ the sequence $\{\hat{\alpha}_n^{(m)}\}$ converges to a fixed-point $\hat{\alpha}_n$ as $m \to \infty$. Then, since $\hat{\alpha}_n$ can be regarded as an estimator of α_k, the frequency $\omega_k = \arccos(\alpha_k)$ can be estimated by

$$\hat{\omega}_n := \arccos(\hat{\alpha}_n). \tag{6}$$

The convergence of (5) depends crucially on how close the initial value $\hat{\alpha}_n^{(0)}$ is to α_k. In other words, it depends on the accuracy of $\hat{\alpha}_n^{(0)}$ as an estimator of α_k. This initial accuracy required for convergence is in turn determined by the bandwidth parameter η. Numerical experiments in [14] indicate that the closer is η to unity, the more stringent is the requirement on $\hat{\alpha}_n^{(0)}$ and the more accurate is the resulting $\hat{\omega}_n$. Quantification of this relationship in the presence of interfering sinusoids and noise is a main objective of this paper.

III. MAIN RESULTS

We assume that η is a function of n such that $\eta \to 1^-$ as $n \to \infty$. An equivalent assumption is that $\delta := 1 - \eta \to 0^+$. This assumption is necessary in order to achieve the optimal error rate for frequency estimation. Furthermore, for any $k, \ell \in \{1, \cdots, p\}$, let $\omega_{k\ell}^{\pm} := \omega_k \pm \omega_\ell$, $\Delta_{k\ell} := |\omega_{k\ell}^-| = \Delta_{\ell k}$, $\Delta_k := \min\{\Delta_{k\ell} : \ell \neq k\}$, and $\Delta := \min\{\Delta_k : k = 1, \cdots, p\}$. We assume, for analysis purposes, that $\Delta_{k\ell}$, Δ_k, and Δ may depend on n and may tend to zero as $n \to \infty$. This assumption is made in order to model possible frequency clustering (i.e., frequencies that are closely spaced relative to the sample size n). Finally, for technical reasons, we assume that $\{\varepsilon_t\}$ is a martingale difference sequence with respect to some filtration $\{\mathfrak{F}_t\}$ such that $E\{\varepsilon_t^2 | \mathfrak{F}_{t-1}\} = \sigma_\varepsilon^2$ almost surely and $E\{\varepsilon_t^4\} < \infty$ for all t. This assumption is less restrictive than the usual $i.i.d.$ assumption because it can be satisfied as long as the ε_t are independent, but not necessarily identically distributed, with mean zero, variance σ_ε^2, and finite fourth-order moments (in this case $\mathfrak{F}_t$ is the sigma field generated by $\{\varepsilon_\tau, \tau \leq t\}$). Note that a martingale difference sequence is a white (uncorrelated) noise process, due to the fact that $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-\tau}) = E\{E(\varepsilon_t \varepsilon_{t-\tau} | \mathfrak{F}_{t-1})\} = E\{E(\varepsilon_t | \mathfrak{F}_{t-1}) \varepsilon_{t-\tau}\} = 0$ for all $\tau > 0$. Note also that the complete knowledge about the distribution of $\{\varepsilon_t\}$, which may or may not be Gaussian, is not required in our analysis.

A. Asymptotic Properties

This section contains five theorems regarding some asymptotic properties of the CM estimator. The first theorem concerns the existence of the CM estimator $\hat{\alpha}_n$ as a fixed-point of $\rho_n(\alpha)$ and the convergence of the CM iteration (5) to $\hat{\alpha}_n$ for a given initial value $\hat{\alpha}_n^{(0)}$.

Theorem 1 Let $\mathcal{A}_{nk} := \{\alpha : |\alpha - \alpha_k| \le a\delta^{\varepsilon}\} \subset \mathcal{A}$ be a neighborhood of α_k, where $a > 0$ and $\varepsilon \in (1, \frac{3}{2})$ are constants. Assume that as $n \to \infty$, $n\eta^n = \mathcal{O}(1)$, $\delta^{3-2\varepsilon} \log n \to 0$, and $\Delta^{-2}\delta^{2-\varepsilon} = \mathcal{O}(1)$. Then, for sufficiently large n, the mapping $\alpha \mapsto \rho_n(\alpha)$ has almost surely a unique fixed-point $\hat{\alpha}_n$ in $\mathcal{A}_{nk}$ such that $\rho_n(\hat{\alpha}_n) = \hat{\alpha}_n$; and for any $\hat{\alpha}_n^{(0)} \in \mathcal{A}_{nk}$, the probability that the sequence $\{\hat{\alpha}_n^{(m)}\}$ defined by (5) converges to $\hat{\alpha}_n$ as $m \to \infty$ is equal to unity.

Note that $n\eta^n = \mathcal{O}(1)$ implies $\delta n \to \infty$. Therefore, Theorem 1 requires that δ approach zero slower than n^{-1}. On the other hand, it also requires that δ approach zero faster than $(\log n)^{-1/(3-2\varepsilon)}$ so that $\delta^{3-2\varepsilon} \log n \to 0$. Both conditions are satisfied with the choice of $\delta = \mathcal{O}(n^{-\nu})$ for any fixed $\nu \in (0,1)$. For a given δ, Theorem 1 requires that the minimum separation of the frequencies be at least $\mathcal{O}(\delta^{1-\varepsilon/2})$, or $\mathcal{O}(n^{-\nu+\varepsilon\nu/2})$ if $\delta = \mathcal{O}(n^{-\nu})$.

The next theorem shows that the CM estimator is strongly consistent.

Theorem 2 Assume that the conditions in Theorem 1 are satisfied. Let $\hat{\omega}_n$ be defined by (6) where $\hat{\alpha}_n$ is the fixed-point of $\rho_n(\alpha)$ in $\mathcal{A}_{nk}$ obtained from (5). Then, for any $d \le \varepsilon$, $\delta^{-d}(\hat{\omega}_n - \omega_k) \to 0$ almost surely as $n \to \infty$. In particular, $\hat{\omega}_n \to \omega_k$ almost surely as $n \to \infty$.

In practice, the initial values may be provided by another estimation procedure. It is more appropriate in such cases to regard $\hat{\alpha}_n^{(0)}$ as a random variable rather than a constant. For random initial values, Theorems 1 and 2 can be modified as follows.

Theorem 3 Let the conditions in Theorem 1 be satisfied. For any $\hat{\alpha}_n^{(0)}$, if $P\{\hat{\alpha}_n^{(0)} \in \mathcal{A}_{nk}\} \to 1$ as $n \to \infty$, then, the probability that the sequence $\{\hat{\alpha}_n^{(m)}\}$ converges to $\hat{\alpha}_n$ as $m \to \infty$ approaches unity as $n \to \infty$. Moreover, for any $d \le \varepsilon$, $\delta^{-d}(\hat{\omega}_n - \omega_k) \to 0$ in probability as $n \to \infty$.

Depending on how quickly δ tends to zero, different rates of weak convergence to normality can be established for the CM estimator. Two useful cases are considered in the following.

Theorem 4 Assume that the conditions in Theorem 1 are satisfied. If, in addition, $\delta^2 n \to \infty$, $\delta^{5-2r} n = \mathcal{O}(1)$, and $\Delta^{-4}\delta^r \to 0$ for some constant $r \in (0,1]$, then $\delta^{-3/2} n^{1/2}(\hat{\omega}_n - \omega_k - \delta^2 \eta^{-1} b_k^{-1}) \xrightarrow{D} \mathcal{N}(0, \gamma_k^{-2})$ as $n \to \infty$, where $\gamma_k := \frac{1}{2}\beta_k^2/\sigma_\varepsilon^2$ and $b_k := 2\beta_k^2/\sum_{\ell \ne k}\beta_\ell^2\{\cot(\frac{1}{2}\omega_{k\ell}^-)) + \cot(\frac{1}{2}\omega_{k\ell}^+)\}$ are the signal-to-noise ratio and the signal-to-interference ratio of the kth sinusoid, respectively.

The requirements in Theorem 4 can be satisfied by $\delta = \mathcal{O}(n^{-\nu})$ for any $\nu \in [\frac{1}{5-2r}, \frac{1}{2})$. According to Theorem 3, the initial precision for the CM iteration to converge can be expressed as $\mathcal{O}_P(n^{-\varepsilon\nu})$. This requirement is satisfied by any estimator whose accuracy is $o_P(n^{-1/(5-2r)})$, which obviously includes all root-n consistent estimators. With such initial values, the iteration (5) is guaranteed by Theorem 3 to converge to the desired CM estimator, at least with probability tending to unity. By Theorem 4, the error of $\hat{\omega}_n$ takes the form

$$\hat{\omega}_n - \omega_k = \max\{\mathcal{O}_P(n^{-2\nu}), \mathcal{O}_P(n^{-(1+3\nu)/2})\} = \mathcal{O}_P(n^{-2\nu}),$$

which is always as small as $o_P(n^{-2/5})$ because $\nu \in [\frac{1}{5-2r}, \frac{1}{2}) \subseteq (\frac{1}{5}, \frac{1}{2})$ for any $r \in (0,1]$. To achieve this error rate, it is required by Theorem 4 that $\Delta^{-1} = o(\delta^{-r/4})$, i.e., the separation of the frequencies be greater than $\mathcal{O}(\delta^{r/4})$. Note that $\Delta^{-4}\delta^r \to 0$ for any $r \in (0,1]$ implies $\Delta^{-2}\delta^{2-\varepsilon} \to 0$. Therefore, the frequency separation condition is stronger in Theorem 4 than in Theorem 1.

Under three different scenerios, Table 1 summarizes the role of bandwidth selection in determining the required initial accuracy, the error rate of the resulting CM estimator, and the required frequency separation. It shows in particular that if a root-n consistent estimator is employed as the initial guess, then the CM iteration is guaranteed to converge and the resulting error rate can be made arbitrarily close to $\mathcal{O}_P(n^{-1})$ by choosing ν near $\frac{1}{2}^-$ (the third row in Table 1), provided the frequency separation is greater than $\mathcal{O}(n^{-1/8})$.

Note that due to the interference from other sinusoids, the CM estimator in Theorem 4 is not as precise as it would be in the case of single sinusoid for the same bandwidth [2]. The interference appears in Theorem 4 as the deterministic bias term $\delta^2\eta^{-1}b_k^{-1}$. This term dominates the random error that takes the form $\mathcal{O}(\delta^{3/2}n^{-1/2})$ and thus determines the precision of $\hat{\omega}_n$. Since the bias tends to zero as $n\to\infty$, $\hat{\omega}_n$ remains to be consistent for estimating ω_k, as ensured by Theorem 3. Except the bias, the asymptotic distribution of $\hat{\omega}_n$ in Theorem 4 is the same as in the single-frequency case discussed in [2].

Theorem 4 requires that δ approach zero at least as fast as $n^{-1/(5-2r)}$ but slower than $n^{-1/2}$. The next theorem concerns two situations in which δ approaches zero faster than $n^{-1/2}$.

Theorem 5 Let the conditions in Theorem 1 be satisfied. (a) If $\delta^2 n\to 0$ and $\Delta^{-4}\delta\to 0$, then $\delta^{-1/2}n\,(\hat{\omega}_n - \omega_k - \delta^2\eta^{-1}b_k^{-1}) \xrightarrow{D} \mathcal{N}(0,\gamma_k^{-1})$ as $n\to\infty$. (b) If $\delta^{3/2}n\to 0$ and $\Delta^{-4}\delta\to 0$, then $\delta^{-1/2}n\,(\hat{\omega}_n-\omega_k)\xrightarrow{D}\mathcal{N}(0,\gamma_k^{-1})$ as $n\to\infty$.

Again, the asymptotic distribution of $\hat{\omega}_n$ in Theorem 5 is the same as in the single-frequency case discussed in [2], except the interference-induced bias. The conditions in Theorem 5 can be satisfied by $\delta=\mathcal{O}(n^{-\nu})$ for any $\nu\in(\frac{1}{2},1)$. By Theorems 1 and 3, the required initial accuracy takes the form $\mathcal{O}(n^{-\varepsilon\nu})$ for almost sure convergence or $\mathcal{O}_P(n^{-\varepsilon\nu})$ for convergence with probability tending to unity. The error of the resulting $\hat{\omega}_n$ can be expressed as $\max\{\mathcal{O}_P(n^{-2\nu}),\mathcal{O}_P(n^{-1-\nu/2})\}$, which implies that

$$\hat{\omega}_n-\omega_k=\begin{cases}\mathcal{O}_P(n^{-2\nu}) & \text{if } \nu\in(\frac{1}{2},\frac{2}{3}),\\ \mathcal{O}_P(n^{-1-\nu/2}) & \text{if } \nu\in(\frac{2}{3},1).\end{cases}$$

Table 2 summarizes several scenarios of bandwidth selection under the conditions in Theorem 5. As can be seen, the error of $\hat{\omega}_n$ is always smaller than $\mathcal{O}_P(n^{-1})$ if $v \in (\frac{1}{2}, \frac{2}{3})$ and smaller than $\mathcal{O}_P(n^{-4/3})$ if $v \in (\frac{2}{3}, 1)$. Most importantly, by choosing v near 1^- (the fourth row in Table 2), the error rate can be made arbitrarily close to the optimal value $\mathcal{O}_P(n^{-3/2})$.

B. A Three-Step Algorithm

Based on the asymptotic results, we now propose a bandwidth selection rule that capitalizes on the ability of the CM estimator in accommodating poor initial values to produce improved frequency estimates. This leads to a three-step algorithm for achieving the optimal statistical efficiency with a computational complexity comparable to that of FFT.

As shown in Table 2, in order to approach the optimal error rate, the initial guess should be at least as accurate as $\mathcal{O}_P(n^{-1})$. Such an initial guess can be obtained from the CM iteration with any $v \in (\frac{1}{2}, \frac{2}{3})$, because Theorem 5 guarantees that the resulting estimator is always more accurate than $\mathcal{O}_P(n^{-1})$. To obtain the latter estimator from the CM iteration, the required initial accuracy is reduced to $o_P(n^{-1/2})$, which, according to Theorem 4, can be satisfied by the CM estimates with any $v \in [\frac{1}{3}, \frac{1}{2})$ when initialized by any estimates of precision $o_P(n^{-1/3})$ (see Table 1).

In summary, with three increasing values of v, namely

$$v_1 \in [\tfrac{1}{3}, \tfrac{1}{2}), \quad v_2 \in (\tfrac{1}{2}, \tfrac{2}{3}), \quad v_3 = 1^-,$$

the CM method is able to improve upon any initial estimates of precision $o_P(n^{-1/3})$ and converge to a final estimate whose accuracy is arbitrarily close to the optimal rate $\mathcal{O}_P(n^{-3/2})$. Note that if a root-n consistent estimator is employed as the first initial guess, then it suffices to take $v_1 \in (\frac{1}{3}, \frac{1}{2})$. Note also that the required frequency separation depends on the accuracy of the first initial guess: if that accuracy is merely $o_P(n^{-1/3})$, then the separation should be greater than $\mathcal{O}(n^{-1/12})$; if a root-n consistent estimator is employed as the first initial guess, then it suffices that the separation be greater than $\mathcal{O}(n^{-1/8})$.

Table 1: DIFFERENT SCENARIOS IN THEOREM 4

Bandwidth	Initial Accuracy	Error of $\hat{\omega}_n$	Δ^{-1}	$C_n(\alpha, \hat{\alpha}_n)$
$\nu = \frac{1}{5}^+$ $(r = 0^+)$	$o_P(n^{-1/5})$ $(\varepsilon = 1^+)$	$o_P(n^{-2/5})$	$\mathcal{O}(1)$	$o(1)$
$\nu = \frac{1}{3}$ $(r = 1)$	$o_P(n^{-1/3})$ $(\varepsilon = 1^+)$	$\mathcal{O}_P(n^{-2/3})$	$o(n^{1/12})$	$o(1)$
$\nu \in (\frac{1}{3}, \frac{1}{2})$ $(r = 1)$	$\mathcal{O}_P(n^{-1/2})$ $(\varepsilon = \frac{1}{2\nu})$	$\mathcal{O}_P(n^{-2\nu})$	$o(n^{\nu/4})$	$\mathcal{O}(n^{-1/2+\nu})$

Table 2: DIFFERENT SCENARIOS IN THEOREM 5

Bandwidth	Initial Accuracy	Error of $\hat{\omega}_n$	Δ^{-1}	$C_n(\alpha, \hat{\alpha}_n)$
$\nu = \frac{1}{2}^+$	$o_P(n^{-1/2})$ $(\varepsilon = 1^+)$	$o_P(n^{-1})$	$\mathcal{O}(n^{1/8})$	$o(1)$
$\nu \in (\frac{1}{2}, \frac{2}{3})$	$\mathcal{O}_P(n^{-2/3})$ $(\varepsilon = \frac{2}{3\nu})$	$\mathcal{O}_P(n^{-2\nu})$	$o(n^{\nu/4})$	$\mathcal{O}(n^{-3/2+\nu})$
$\nu = \frac{2}{3}^+$	$o_P(n^{-2/3})$ $(\varepsilon = 1^+)$	$o_P(n^{-4/3})$	$\mathcal{O}(n^{1/6})$	$o(1)$
$\nu \in (\frac{2}{3}, 1)$	$\mathcal{O}_P(n^{-1})$ $(\varepsilon = \frac{1}{\nu})$	$\mathcal{O}_P(n^{-1-\nu/2})$	$o(n^{\nu/4})$	$\mathcal{O}(n^{-1+\nu})$

The computational complexity of this three-step algorithm is comparable to that of n-point FFT, both taking the form $\mathcal{O}(n \log n)$. To prove this assertion, consider the expressions of the contraction coefficient $C_n(\alpha, \hat{\alpha}_n)$ given in Tables 1 and 2. According to (19) and (20),

$$\hat{\alpha}_n^{(m+1)} - \hat{\alpha}_n = C_n^{(m)} (\hat{\alpha}_n^{(m)} - \hat{\alpha}_n),$$

where $C_n^{(m)} := C_n(\hat{\alpha}_n^{(m)}, \hat{\alpha}_n)$. In Step 3 of the algorithm, we have $C_n^{(m)} = \mathcal{o}(1)$ because $\nu_3 = 1^-$ (Row 4 in Table 2). Therefore, the number of iterations required to achieve the desired accuracy $\mathcal{O}_P(n^{-3/2})$ from an initial value of accuracy $\mathcal{O}_P(n^{-1})$ can be expressed as $\mathcal{O}(\log n)$. Similarly, the number of iterations required in Step 2 takes the form $\mathcal{O}(1)$ because $C_n^{(m)} = \mathcal{O}(n^{-3/2+\nu_2})$ (Row 2 in Table 2). In Step 1, if the initial accuracy is $\mathcal{o}_P(n^{-1/3})$, then the required number of iterations is $\mathcal{O}(\log n)$ (Row 2 in Table 1); if the initial accuracy is $\mathcal{O}_P(n^{-1/2})$, then that number is reduced to $\mathcal{O}(1)$ (Row 3 in Table 1). Therefore, the total number of iterations required to achieve the optimal error rate from an initial accuracy $\mathcal{o}_P(n^{-1/3})$ or $\mathcal{O}_P(n^{-1/2})$ can be expressed as $\mathcal{O}(\log n)$. The overall complexity takes the form $\mathcal{O}(n \log n)$ because the complexity of updating the estimate in each iteration is $\mathcal{O}(n)$.

A simulation example is shown in Fig. 1 to demonstrate the algorithm. The time series in this example contains three equal-amplitude sinusoids of frequencies $\omega_1 = 2\pi \times 10.5/n$, $\omega_2 = 2\pi \times 11.5/n$, and $\omega_3 = 2\pi \times 20.5/n$, where $n = 64$. The noise is a zero-mean white Gaussian process, with the sample variance standardized so that the signal-to-noise ratio of each sinusoid is equal to -6 dB. Fig. 1 shows the trajectory of the normalized frequency estimates $\hat{\omega}_n^{(m)}/(2\pi)$, as functions of m, obtained with different initial values, where $\hat{\omega}_n^{(m)} := \arccos(\hat{\alpha}_n^{(m)})$. The CM iteration begins with $\eta_1 = 0.96$; after 6 iterations, the bandwidth parameter is increased to $\eta_2 = 0.98$, and after 6 additional iterations, it is increased to $\eta_3 = 0.99$. For each initial value, the iteration converges to one of the frequencies with the final (highest) accuracy determined by the last (smallest) bandwidth.

As suggested in [30], the convergence of the CM iteration can be accelerated by replacing $\rho_n(\alpha)$ in (5) with the modified mapping $\tilde{\rho}_n(\alpha) := \rho_n(\alpha)\,\mu + \alpha\,(1-\mu)$, where $\mu \neq 1$ is a constant. Note that $\tilde{\rho}_n(\alpha)$

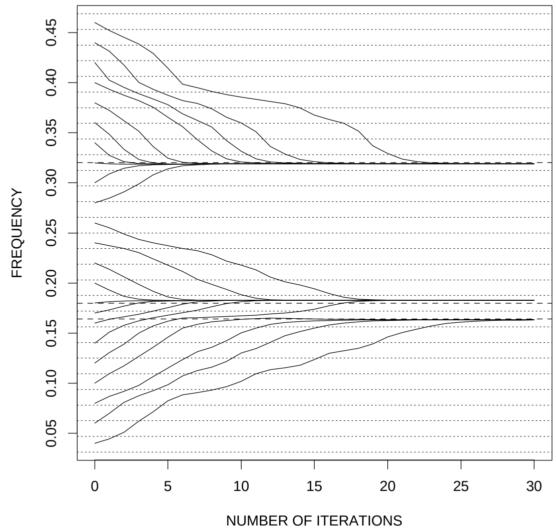

Figure 1: CM frequency estimates $\hat{\omega}_n^{(m)}/(2\pi)$ versus the number of iterations m for different initial values. Dashed lines represent the true frequencies and dotted lines represent the Fourier frequencies j/n $(j = 1, 2, \cdots)$. The sample size is $n = 64$, and the signal-to-noise ratio is -6 dB for each sinusoid. The bandwidth parameters are $\eta_1 = 0.96$ for $1 \le m \le 6$, $\eta_2 = 0.98$ for $7 \le m \le 12$, and $\eta_3 = 0.99$ for $m \ge 13$.

has the same fixed-points as $\rho_n(\alpha)$. Furthermore, since the contraction coefficient of $\tilde{\rho}_n(\alpha)$ is $\tilde{C}_n(\alpha, \hat{\alpha}_n) := 1 - \mu\{1 - C_n(\alpha, \hat{\alpha}_n)\}$, the CM iteration with the modified mapping is guaranteed to converge (under the conditions in Theorem 1) if μ satisfies $0 < \mu < 2/\{1 - C_n(\alpha, \hat{\alpha}_n)\}$. The choice of $\mu = 2$ is valid in particular when $C_n(\alpha, \hat{\alpha}_n) > 0$ (as is the case in Fig. 1). For $\mu = 2$, $|\tilde{C}_n(\alpha, \hat{\alpha}_n)| < C_n(\alpha, \hat{\alpha}_n)$ if $\frac{1}{3} < C_n(\alpha, \hat{\alpha}_n) < 1$. This means that accelerated convergence can be achieved with $\tilde{\rho}_n(\alpha)$ when the convergence with $\rho_n(\alpha)$ is slow (e.g., $C_n(\alpha, \hat{\alpha}_n) \approx 1$). Fig. 2 shows the trajectory of the CM estimates obtained with the modified mapping ($\mu = 2$) for the same data used in Fig. 1. Accelerated convergence is evident.

C. Remarks

So far, the ϕ_k in (1) are assumed to be constants. Alternatively, the ϕ_k can be 'randomized' by assuming that they are *i.i.d.* random variables with uniform distribution in $(-\pi, \pi]$ and are independent of $\{\varepsilon_t\}$. This leads

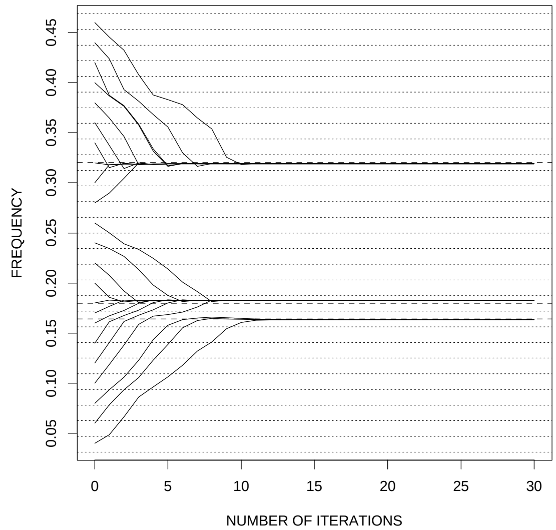

Figure 2: Same as in Fig. 1 except that the modified mapping $\tilde{\rho}_n(\alpha)$ with $\mu = 2$ is employed instead of $\rho_n(\alpha)$. The bandwidth parameters are $\eta_1 = 0.96$ for $1 \le m \le 4$, $\eta_2 = 0.98$ for $5 \le m \le 8$, and $\eta_3 = 0.99$ for $m \ge 9$.

to a stochastic signal model in which the sinusoids become (second-order) stationary random processes. The randomization does not alter our results presented in the previous sections because these results do not dependent on the values of ϕ_k. This can be easily justified, as demonstrated in [26], by first conditioning on the ϕ_k to obtain a probabilistic statement (e.g., an estimator exists almost surely or with probability tending to unity, or an estimator converges in distribution to a normal random variable whose mean and variance do not depend on the ϕ_k) and then taking the expected value of the conditional probabilities with respect to the ϕ_k. The same remark applies to the asymptotic error rate of MP and NLS frequency estimators.

Even though our discussion is focused on real-valued sinusoids, similar results can be obtained under the complex-valued sinusoid-plus-noise model $y_t = \sum_{k=1}^{p} \beta_k \exp\{i(\omega_k t + \phi_k)\} + \varepsilon_t$. In this case, it suffices to consider a complex AR(1) filter $(1 + \alpha \mathfrak{B})^{-1}$, where $\alpha := \eta \exp(i\omega)$ and $\eta \in (0,1)$. For this filter, the (ensemble) lag-one autocorrelation coefficient of the filtered (white) noise is equal to α. Therefore, the filter satisfies the "fundamental property" required by the CM method [22]. This property implies that when $p = 1$ the lag-one autocorrelation coefficient of the filtered $\{y_t\}$ forms a contractive mapping whose unique fixed-point is equal to $\eta\exp(i\omega_1)$. Therefore, as in the real-valued case, frequency estimators can be obtained from the sample lag-one autocorrelation of the filtered observations. Note that the parameter η controls the bandwidth of the AR(1) filter in the same way as it does the AR(2) filter.

IV. Proof of the Theorems

The theorems in Sec. III are proved in this section on the basis of some preliminary propositions. The proof of these propositions are outlined in Appendix I (see [29] for a complete proof).

First, we introduce some useful notation. Let Λ_{nk} be the set of $\lambda \in (0,\pi)$ determined by (3) with $\alpha \in \mathcal{A}_{nk} \subset \mathcal{A}$. Since $\eta \to 1$ and hence $\mathcal{A} \to (-1,1)$ as $n \to \infty$, it follows that α_k becomes an interior point of $\mathcal{A}$ for large n. Furthermore, since the length of $\mathcal{A}_{nk}$ decreases with the increase of n, the interval $\mathcal{A}_{nk}$, there exists a closed subinterval $\mathcal{A}_k$, which is independent of n, such that $\mathcal{A}_{nk} \subset \mathcal{A}_k$ for large n. As a result, there is a closed subinterval $\Lambda_k \subset (0,\pi)$, which is independent of n, such that $\Lambda_{nk} \subset \Lambda_k$ for large n. Therefore, any $\lambda \in \Lambda_{nk}$ can be uniformly bounded away from 0 and π for large n.

Moreover, according to (2) and (3), we can write

$$y_t(\alpha)+\eta^2 y_{t-2}(\alpha) = y_t + (1+\eta^2)\,\alpha\, y_{t-1}(\alpha).$$

Therefore, with $\lambda \in (0,\pi)$ defined by (3), the mapping $\rho_n(\alpha)$ in (4) can be expressed as

$$\rho_n(\alpha) = \alpha + (1+\eta^2)^{-1}\sin\lambda\,\frac{\Phi_n(\lambda)}{\Psi_n(\lambda)}, \tag{7}$$

where

$$\Psi_n(\lambda) := \sin^2\lambda \sum_{t=1}^{n} y_{t-1}^2(\alpha), \quad \Phi_n(\lambda) := \sin\lambda \sum_{t=1}^{n} y_t\, y_{t-1}(\alpha). \tag{8}$$

Equation (7) shows that the behavior of $\rho_n(\alpha)$ in a neighborhood of α_k is determined by the behavior of $\Psi_n(\lambda)$ and $\Phi_n(\lambda)$ in a neighborhood of $\lambda_k \in (0,\pi)$, where λ_k is defined by

$$\cos\lambda_k = \frac{1+\eta^2}{2\eta}\,\alpha_k. \tag{9}$$

The propositions in the following describe the behavior of $\Psi_n(\lambda)$ and $\Phi_n(\lambda)$ and are prerequisite to the proof of the theorems.

A. Preliminary Propositions

The first two propositions describe some asymptotic characteristics of $\Psi_n(\lambda)$ and $\Phi_n(\lambda)$.

Proposition 1 Let $\Psi_n(\lambda)$ be defined by (8) with $\lambda \in \Lambda_{nk}$, $\alpha \in \mathcal{A}_{nk}$, and $\varepsilon \in (1, \frac{3}{2})$. As $n \to \infty$, assume that $\delta \to 0$, $n(1-\delta)^n = \mathcal{O}(1)$, and $\Delta^{-1}\delta \to 0$. Then,

$$\begin{aligned}\Psi_n(\lambda) \;=\; & \tfrac{1}{8}\beta_k^2\eta^{-2}\delta^{-2}n + \mathcal{O}(\Delta_k^{-4}n) + \mathcal{O}(\Delta_k^{-2}\delta^{-1}n) + \mathcal{O}(\Delta_k^{-2}\delta^{-1/2}n\sqrt{\log n}) \\ & + \mathcal{O}(\delta^{-3}) + \mathcal{O}(\delta^{-3/2}n\sqrt{\log n}) + \mathcal{O}(\delta^{-1}n\log n) \\ & + (\lambda-\lambda_k)\{\mathcal{O}(\Delta_k^{-6}\delta^{-1}) + \mathcal{O}(\Delta_k^{-4}\delta^{-2}) + \mathcal{O}(\Delta_k^{-2}\delta^{-2}n) + \mathcal{O}(\delta^{\varepsilon-4}n) \\ & + \mathcal{O}(\delta^{-5/2}n\sqrt{\log n}) + \mathcal{O}(\delta^{-2}n\log n)\} \end{aligned} \tag{10}$$

almost surely and uniformly in $\lambda \in \Lambda_{nk}$ for sufficiently large n. Under the same assumptions,

$$\begin{aligned}\Psi_n(\lambda) - \Psi_n(\lambda') \;=\; & (\lambda-\lambda')\{\mathcal{O}(\Delta_k^{-6}\delta^{-1}) + \mathcal{O}(\Delta_k^{-4}\delta^{-2}) + \mathcal{O}(\Delta_k^{-2}\delta^{-2}n) + \mathcal{O}(\delta^{\varepsilon-4}n) \\ & + \mathcal{O}(\delta^{-5/2}n\sqrt{\log n}) + \mathcal{O}(\delta^{-2}n\log n)\} \end{aligned} \tag{11}$$

almost surely and uniformly in $\lambda, \lambda' \in \Lambda_{nk}$.

Proposition 2 Let $\Phi_n(\lambda)$ be defined by (8). If the conditions in Proposition 1 are satisfied, then

$$\begin{aligned}\Phi_n(\lambda) \;=\; & n\eta^{-1}\xi_k + \mathcal{O}(\Delta_k^{-4}) + \mathcal{O}(\Delta_k^{-2}\delta^{-1}) + \mathcal{O}(\Delta_k^{-4}\delta^{2}n) \\ & + \mathcal{O}(\delta^{-1/2}n\sqrt{\log n}) + \mathcal{O}(\delta^{-1}\sqrt{n\log n}) + \mathcal{O}(\delta^{-3/2}\sqrt{\log n}) \\ & + (\lambda-\lambda_k)\{\tfrac{1}{4}\beta_k^2\delta^{-2}n + \mathcal{O}(\Delta_k^{-4}n) + \mathcal{O}(\Delta_k^{-2}\delta^{-2}) + \mathcal{O}(\delta^{-3}) + \mathcal{O}(\delta^{\varepsilon-3}n) \\ & + \mathcal{O}(\delta^{-3/2}n\sqrt{\log n}) + \mathcal{O}(\delta^{-2}\sqrt{n\log n}) + \mathcal{O}(\delta^{-5/2}\sqrt{\log n})\} \end{aligned} \tag{12}$$

almost surely and uniformly in $\lambda \in \Lambda_{nk}$ for sufficiently large n, where

$$\xi_k := \tfrac{1}{8} \sum_{\ell \neq k} \beta_\ell^2 \{\cot(\tfrac{1}{2}\omega_{k\ell}^-) + \cot(\tfrac{1}{2}\omega_{k\ell}^+)\}.$$

Under the same assumptions,

$$\begin{aligned}\Phi_n(\lambda) - \Phi_n(\lambda') \;=\; & (\lambda - \lambda')\{\tfrac{1}{4}\beta_k^2 \delta^{-2} n + \mathcal{O}(\Delta_k^{-4} n) + \mathcal{O}(\Delta_k^{-2}\delta^{-2}) + \mathcal{O}(\delta^{-3}) + \mathcal{O}(\delta^{\varepsilon-3} n) \\ & + \mathcal{O}(\delta^{-3/2} n \sqrt{\log n}) + \mathcal{O}(\delta^{-2}\sqrt{n \log n}) + \mathcal{O}(\delta^{-5/2}\sqrt{\log n})\} \end{aligned} \tag{13}$$

almost surely and uniformly in $\lambda, \lambda' \in \Lambda_{nk}$.

The next result is presented without proof because it can be easily obtained from Propositions 1 and 2 together with the fact that $\lambda - \lambda_k = \mathcal{O}(\delta^\varepsilon)$ for any $\lambda \in \Lambda_{nk}$.

Corollary 1 Let the conditions in Proposition 1 be satisfied. If, in addition, $\delta^{3-2\varepsilon} \log n \to 0$ and $\Delta_k^{-2}\delta \to 0$ as $n \to \infty$, then

$$\begin{aligned}
\Psi_n(\lambda) &= \tfrac{1}{8}\beta_k^2 \eta^{-2}\delta^{-2} n \{1 + \mathcal{O}(\Delta_k^{-2}\delta) + \mathcal{O}(\delta^{-1} n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{2\varepsilon-2})\}, \\
\Psi_n(\lambda_k) &= \tfrac{1}{8}\beta_k^2 \eta^{-2}\delta^{-2} n \{1 + \mathcal{O}(\Delta_k^{-2}\delta) + \mathcal{O}(\delta^{-1} n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n})\}, \\
\Psi_n(\lambda)\Psi_n(\lambda') &= \tfrac{1}{64}\beta_k^4 \eta^{-4}\delta^{-4} n^2 \{1 + \mathcal{O}(\Delta_k^{-2}\delta) + \mathcal{O}(\delta^{-1} n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{2\varepsilon-2})\}, \\
\Psi_n(\lambda) - \Psi_n(\lambda') &= (\lambda - \lambda')\{\mathcal{O}(\Delta_k^{-2}\delta^{-2} n) + \mathcal{O}(\delta^{\varepsilon-4} n)\}, \\
\Phi_n(\lambda) &= \mathcal{O}(\Delta_k^{-2}\delta^{-1}) + \mathcal{O}(\delta^{\varepsilon-2} n), \\
\Phi_n(\lambda_k) &= \mathcal{O}(\Delta_k^{-2}\delta^{-1}) + \mathcal{O}(\delta^{-1/2} n \sqrt{\log n}), \\
\Phi_n(\lambda) - \Phi_n(\lambda') &= (\lambda - \lambda')\tfrac{1}{4}\beta_k^2 \delta^{-2} n \{1 + \mathcal{O}(\Delta_k^{-4}\delta^2) + \mathcal{O}(\delta^{-1} n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{\varepsilon-1})\},
\end{aligned}$$

almost surely and uniformly in $\lambda, \lambda' \in \Lambda_{nk}$ for sufficiently large n.

The next two propositions play an important role in establishing the asymptotic normality of the CM frequency estimator. One of them is cited from the literature without proof.

Proposition 3 Under the conditions in Proposition 1,

$$\begin{aligned}\Phi_n(\lambda_k) &= W_{n1}+W_{n2}+n\eta^{-1}\xi_k+\mathcal{O}_P(\Delta_k^{-4})+\mathcal{O}_P(\Delta_k^{-2}\delta^{-1}) \\ &\quad +\mathcal{O}_P(\Delta_k^{-4}\delta^2 n)+\mathcal{O}_P(\Delta_k^{-2}n^{1/2}),\end{aligned}$$

where

$$W_{n1} := \tfrac{1}{2}\beta_k\delta\, g(\lambda_k-\omega_k)\sum_{t=1}^{n}\varepsilon_t\,(\eta^{n-t}-\eta^{t-1})\,\sin(t\omega_k+\phi_k), \tag{14}$$

$$W_{n2} := \sum_{t=1}^{n}\sum_{j=0}^{t-1}\eta^{j-1}\sin(j\lambda_k)\,\varepsilon_t\,\varepsilon_{t-j}, \tag{15}$$

and $g(\lambda):=(1-2\eta\cos\lambda+\eta^2)^{-1}$.

Proposition 4 [2] Let W_{n1} and W_{n2} be defined by (14) and (15), respectively. Then, under the conditions in Proposition 1, $\delta^{3/2}W_{n1}\xrightarrow{D}\mathcal{N}(0,\frac{1}{8}\beta_k^2\sigma_\varepsilon^2)$ and $\delta^{1/2}n^{-1/2}W_{n2}\xrightarrow{D}\mathcal{N}(0,\frac{1}{4}\sigma_\varepsilon^4)$ as $n\to\infty$.

The final proposition, cited without proof, describes some useful relations between λ and α.

Proposition 5 [27] Let $\mathcal{A}_{nk}$ be defined in Theorem 1 with $\alpha_k=\cos\omega_k$ and $\varepsilon>0$. If $\eta\to 1$ ($\delta=1-\eta\to 0$) as $n\to\infty$, then the following assertions are valid.

(a) Let λ_k be defined by (9), then

$$\lambda_k-\omega_k=-\tfrac{1}{2}\eta^{-1}\delta^2\alpha_k(1-\alpha_k^2)^{-1/2}+\mathcal{O}(\delta^4).$$

Moreover, there exist constants $c_0>0$ and $n_0>0$ such that $|\lambda_k-\omega_k|\le c_0\delta^2$ for $n>n_0$.

(b) Let $\lambda,\lambda'\in\Lambda_{nk}$ be determined by $\alpha,\alpha'\in\mathcal{A}_{nk}$ according to (3), then

$$\lambda-\lambda'=-\frac{1}{\sin\lambda'}\,\frac{1+\eta^2}{2\eta}\,(\alpha-\alpha')\left\{1+\frac{\xi\sin\lambda'}{2(1-\xi^2)^{3/2}}\,\frac{1+\eta^2}{2\eta}\,(\alpha-\alpha')\right\}$$

where $\xi\in(-1,1)$ depends on λ and λ' and there exist constants $0<c<1$ and $n_0>0$ such that $\xi^2\le c$ for all $\lambda,\lambda'\in\Lambda_{nk}$ and for $n>n_0$.

Equipped with these propositions, let us now prove the theorems.

B. Proof of Theorem 1

It suffices to show that $\rho_n(\alpha)$ is a contractive mapping in $\mathcal{A}_{nk}$. This can be done by proving that the following inequalities hold almost surely for sufficiently large n (e.g., [31], p. 251, Theorem 5.2.3):

$$|\rho_n(\alpha) - \rho_n(\alpha')| \le c\,|\alpha - \alpha'| \tag{16}$$

for all $\alpha, \alpha' \in \mathcal{A}_{nk}$, where $c \in (0,1)$ is a constant, and

$$|\rho_n(\alpha_k) - \alpha_k| \le (1-c)\,a\delta^{\varepsilon}, \tag{17}$$

where c is given in (16) and a is given in Theorem 1. Let us now prove these inequalities.

Proof of (16). It follows from (7) that

$$\rho_n(\alpha) - \rho_n(\alpha') = \alpha - \alpha' + R_n, \tag{18}$$

where

$$\begin{aligned} R_n &:= (J_1 + J_2 + J_3)\,\{(1+\eta^2)\,\Psi_n(\lambda)\,\Psi_n(\lambda')\}^{-1}, \\ J_1 &:= \{\Psi_n(\lambda') - \Psi_n(\lambda)\}\,\Phi_n(\lambda)\sin\lambda, \\ J_2 &:= \{\Phi_n(\lambda) - \Phi_n(\lambda')\}\,\Psi_n(\lambda)\sin\lambda, \\ J_3 &:= (\sin\lambda - \sin\lambda')\,\Phi_n(\lambda')\,\Psi_n(\lambda). \end{aligned}$$

Note that $\Delta^{-2}\delta^{2-\varepsilon} = \mathcal{O}(1)$ implies $\Delta^{-2}\delta \to 0$. Therefore, by Corollary 1,

$$\begin{aligned} J_1 &= (\lambda - \lambda')\,\{\mathcal{O}(\Delta_k^{-4}\delta^{-3}n) + \mathcal{O}(\Delta_k^{-2}\delta^{\varepsilon-5}n) + \mathcal{O}(\Delta_k^{-2}\delta^{\varepsilon-4}n^2) + \mathcal{O}(\delta^{2\varepsilon-6}n^2)\}, \\ J_2 &= (\lambda - \lambda')\sin\lambda\,\tfrac{1}{32}\beta_k^4\eta^{-2}\delta^{-4}n^2\{1 + \mathcal{O}(\Delta_k^{-2}\delta) + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{\varepsilon-1})\}, \\ J_3 &= (\lambda - \lambda')\,\{\mathcal{O}(\Delta_k^{-2}\delta^{-3}n) + \mathcal{O}(\delta^{\varepsilon-4}n^2)\}\}. \end{aligned}$$

Since $\sin\lambda$ can be bounded away from zero uniformly for all $\lambda \in \Lambda_{nk}$, it follows that

$$\begin{aligned} J_1 + J_2 + J_3 &= (\lambda - \lambda')\sin\lambda\,\tfrac{1}{32}\beta_k^4\eta^{-2}\delta^{-4}n^2\,\{1 + \mathcal{O}(\Delta_k^{-2}\delta) \\ &\quad + \mathcal{O}(\delta^{-1}n^{-1}) + \mathcal{O}(\delta^{1/2}\sqrt{\log n}) + \mathcal{O}(\delta^{\varepsilon-1})\}. \end{aligned}$$

This, combined with the expression for $\Psi_n(\lambda)\,\Psi_n(\lambda')$ in Corollary 1, leads to

$$\begin{aligned} R_n &= (\lambda-\lambda')\sin\lambda\, 2\eta^2(1+\eta^2)^{-1}\{1+\mathcal{O}(\Delta_k^{-2}\delta) \\ &\quad + \mathcal{O}(\delta^{-1}n^{-1})+\mathcal{O}(\delta^{1/2}\sqrt{\log n})+\mathcal{O}(\delta^{\varepsilon-1})\}. \end{aligned}$$

Furthermore, since $\alpha-\alpha'=\mathcal{O}(\delta^{\varepsilon})$, it follows from Proposition 5(b) that

$$\lambda-\lambda' = -(2\eta\sin\lambda)^{-1}(1+\eta^2)\,(\alpha-\alpha')\,\{1+\mathcal{O}(\delta^{\varepsilon})\}.$$

Substituting this expression in the foregoing equation yields

$$R_n = -\,\eta\,(\alpha-\alpha')\,\{1+\mathcal{O}(\Delta_k^{-2}\delta)+\ \mathcal{O}(\delta^{-1}n^{-1})+\mathcal{O}(\delta^{1/2}\sqrt{\log n})+\mathcal{O}(\delta^{\varepsilon-1})\}.$$

Therefore, (18) can be rewritten as

$$\rho_n(\alpha)-\rho_n(\alpha') = C_n(\alpha,\alpha')\,(\alpha-\alpha'), \tag{19}$$

where $C_n(\alpha,\alpha') := \{\rho_n(\alpha)-\rho_n(\alpha')\}/(\alpha-\alpha')$ can be expressed as

$$\begin{aligned} C_n(\alpha,\alpha') &= 1-\eta\,\{1+\mathcal{O}(\Delta_k^{-2}\delta)+\mathcal{O}(\delta^{-1}n^{-1})+\mathcal{O}(\delta^{1/2}\sqrt{\log n})+\mathcal{O}(\delta^{\varepsilon-1})\} \\ &= \delta+\mathcal{O}(\Delta_k^{-2}\delta)+\mathcal{O}(\delta^{-1}n^{-1})+\mathcal{O}(\delta^{1/2}\sqrt{\log n})+\mathcal{O}(\delta^{\varepsilon-1}). \end{aligned} \tag{20}$$

The proof is complete upon noting that $C_n(\alpha,\alpha') \stackrel{a.s.}{\to} 0$ uniformly in α, $\alpha' \in \mathcal{A}_{nk}$.

Proof of (17). It follows from (7) that $\rho_n(\alpha_k)-\alpha_k = \sin\lambda_k\,\Phi_n(\lambda_k)\,\{(1+\eta^2)\,\Psi_n(\lambda_k)\}^{-1}$, where λ_k is defined by (9). According to Corollary 1,

$$\frac{\Phi_n(\lambda_k)}{\Psi_n(\lambda_k)} = \mathcal{O}(\Delta_k^{-2}\delta n^{-1})+\mathcal{O}(\delta^{3/2}\sqrt{\log n}).$$

Since $\delta^{3-2\varepsilon}\log n \to 0$, one can write $\mathcal{O}(\delta^{3/2}\sqrt{\log n}) = \mathcal{o}(\delta^{\varepsilon})$. Since $\delta n\to\infty$ and $\Delta^{-2}\delta^{2-\varepsilon}=\mathcal{O}(1)$, one can write $\Delta_k^{-2}\delta^{1-\varepsilon}n^{-1}\to 0$ and hence $\Delta_k^{-2}\delta n^{-1} = \mathcal{o}(\delta^{\varepsilon})$. Combining these results yields

$$\rho_n(\alpha_k)-\alpha_k = \mathcal{o}(\delta^{\varepsilon})$$

almost surely for large n. The proof is thus complete.

C. Proof of Theorem 2

Let $\hat{\alpha}_n$ be the fixed-point of $\rho_n(\alpha)$ in $\mathcal{A}_{nk}$. Then, it follows from (19) that

$$\hat{\alpha}_n - \alpha_k = C_{nk}\,(\hat{\alpha}_n - \alpha_k) + \rho_n(\alpha_k) - \alpha_k,$$

where $C_{nk} := C_n(\hat{\alpha}_n, \alpha_k)$. This equation can be rewritten as

$$\hat{\alpha}_n - \alpha_k = (1 - C_{nk})^{-1}\{\rho_n(\alpha_k) - \alpha_k\},$$

which, combined with (7), leads to

$$\hat{\alpha}_n - \alpha_k = \{(1+\eta^2)\,(1 - C_{nk})\}^{-1} \sin\lambda_k \, \frac{\Phi_n(\lambda_k)}{\Psi_n(\lambda_k)}. \tag{21}$$

Moreover, by Corollary 1, $\delta^2 n^{-1}\Psi_n(\lambda_k) \to \frac{1}{8}\beta_k^2$ and $\Phi_n(\lambda_k) = \mathcal{O}(\Delta_k^{-2}\delta^{-1}) + \mathcal{O}(\delta^{-1/2} n \sqrt{\log n})$ almost surely as $n \to \infty$. Combining these results with the fact that $C_{nk} \to 0$ gives $\hat{\alpha}_n - \alpha_k = \mathcal{O}(\delta^2 n^{-1}\Phi_n(\lambda_k)) = \mathcal{O}(\Delta_k^{-2}\delta n^{-1}) + \mathcal{O}(\delta^{3/2}\sqrt{\log n})$. Finally. since $\delta^{3-2\varepsilon}\log n \to 0$, $\delta n \to \infty$, and $\Delta^{-2}\delta^{2-\varepsilon} = \mathcal{O}(1)$, it follows that

$$\delta^{-d}(\hat{\alpha}_n - \alpha_k) = \mathcal{O}(\Delta_k^{-2}\delta^{1-d} n^{-1}) + \mathcal{O}(\delta^{3/2-d}\sqrt{\log n}) \stackrel{a.s.}{\to} 0$$

for any $d \leq \varepsilon$. The proof is complete upon noting (6) and the fact that $\omega_k = \arccos(\alpha_k)$.

D. Proof of Theorem 3

If $\rho_n(\alpha)$ is contractive (in $\mathcal{A}_{nk}$), i.e., if it satisfies (16) and (17), then $\hat{\alpha}_n^{(m)} \to \hat{\alpha}_n$ as $m \to \infty$ for any $\hat{\alpha}_n^{(0)} \in \mathcal{A}_{nk}$. This implies that

$$\begin{aligned}
& P\Big\{\lim_{m\to\infty} \hat{\alpha}_n^{(m)} = \hat{\alpha}_n\Big\} \\
\geq \;& P\{\rho_n(\alpha) \text{ is contractive and } \hat{\alpha}_n^{(0)} \in \mathcal{A}_{nk}\} \\
= \;& P\{\rho_n(\alpha) \text{ is contractive}\} \\
& - P\{\rho_n(\alpha) \text{ is contractive and } \hat{\alpha}_n^{(0)} \notin \mathcal{A}_{nk}\} \\
\geq \;& P\{\rho_n(\alpha) \text{ is contractive}\} - P\{\hat{\alpha}_n^{(0)} \notin \mathcal{A}_{nk}\}.
\end{aligned}$$

By Theorem 1, $P\{\rho_n(\alpha)$ is contractive$\} = 1$ for large n. By assumption, $P\{\hat{\alpha}_n^{(0)} \in \mathcal{A}_{nk}\} \to 1$ as $n \to \infty$. Combining these results leads to $P\{\lim_{m\to\infty} \hat{\alpha}_n^{(m)} = \hat{\alpha}_n\} \to 1$ as $n \to \infty$. The second part of the assertion follows from a similar argument coupled with Theorem 2.

E. Proof of Theorem 4

Consider (21), and observe that $\sin\lambda_k \xrightarrow{a.s.} \sin\omega_k$ by Proposition 5(a), $\delta^2 n^{-1}\Psi_n(\lambda_k) \xrightarrow{a.s.} \frac{1}{8}\beta_k^2$ by Corollary 1, and $C_{nk} \xrightarrow{a.s.} 0$ by (20). Therefore, by Slutsky's theorem (e.g., [32], p. 337, Theorem 1.4), $\delta^{-3/2} n^{1/2}(\hat{\alpha}_n - \alpha_k)$ has the same asymptotic distribution as

$$Z_{n1} := 4\beta_k^{-2} \sin\omega_k\, \delta^{1/2} n^{-1/2} \Phi_n(\lambda_k). \tag{22}$$

Note that $\Phi_n(\lambda_k)$ has the expression in Proposition 3 where $W_{n1} = \mathcal{O}_P(\delta^{-3/2})$ by Proposition 4. Therefore, under the assumption that $\delta^2 n \to \infty$, $\delta^{5-2r} n = \mathcal{O}(1)$, and $\Delta^{-4}\delta^r \to 0$, it follows from Proposition 3 that $\delta^{1/2} n^{-1/2}\{\Phi_n(\lambda_k) - n\eta^{-1}\xi_k\}$ has the same asymptotic distribution as $\delta^{1/2} n^{-1/2} W_{n2}$, which, by Proposition 4, is $\mathcal{N}(0, \frac{1}{4}\sigma_\varepsilon^4)$. Therefore,

$$Z_{n1} - \delta^{1/2} n^{1/2} \eta^{-1} \xi_k' \xrightarrow{D} \mathcal{N}(0, \gamma_k^{-2} \sin^2\omega_k),$$

where $\xi_k' := 4\beta_k^{-2}\xi_k \sin\omega_k$. Combining these results yields

$$\delta^{-3/2} n^{1/2} (\hat{\alpha}_n - \alpha_k - \delta^2 \eta^{-1} \xi_k') \xrightarrow{D} \mathcal{N}(0, \gamma_k^{-2} \sin^2\omega_k).$$

Since $\xi_k' = b_k^{-1} \sin\omega_k$, the proof is complete upon noting that $\hat{\omega}_n - \omega_k$ has the same asymptotic distribution as $(\hat{\alpha}_n - \alpha_k)/\sin\omega_k$ by the delta method (e.g., [32], p. 337, Theorem 1.5).

F. Proof of Theorem 5

By an argument similar to the proof of Theorem 4, it can be shown from (21) that $\delta^{-1/2} n(\hat{\alpha}_n - \alpha_k)$ has the same asymptotic distribution as

$$Z_{n2} := 4\beta_k^{-2} \sin\omega_k\, \delta^{3/2} \Phi_n(\lambda_k),$$

where $\Phi_n(\lambda_k)$ has the expression in Proposition 3 and $W_{n2} = \mathcal{O}_P(\delta^{-1/2}n^{1/2})$ by Proposition 4. Therefore, under the assumption that $\delta^2 n \to 0$ and $\Delta^{-4}\delta \to 0$, it follows from Proposition 3 that $\delta^{3/2}\{\Phi_n(\lambda_k) - n\eta^{-1}\xi_k\}$ has the same asymptotic distribution as $\delta^{3/2}W_{n1}$, namely $\mathcal{N}(0, \frac{1}{8}\beta_k^2\sigma_\varepsilon^2)$ by Proposition 4. This implies that

$$Z_{n2} - \delta^{3/2}n\eta^{-1}\xi_k' \xrightarrow{D} \mathcal{N}(0, \gamma_k^{-1}\sin^2\omega_k),$$

and hence

$$\delta^{-1/2}n\,(\hat{\alpha}_n - \alpha_k - \delta^2\eta^{-1}\xi_k') \xrightarrow{D} \mathcal{N}(0, \gamma_k^{-1}\sin^2\omega_k).$$

An application of the delta method leads to $\delta^{-1/2}n\,(\hat{\omega}_n - \omega_k - \delta^2\eta^{-1}b_k^{-1}) \xrightarrow{D} \mathcal{N}(0, \gamma_k^{-1})$. Furthermore, if, in addition, $\delta^{3/2}n \to 0$, then $\delta^{-1/2}n\,(\hat{\omega}_n - \omega_k)$ has the same asymptotic distribution $\mathcal{N}(0, \gamma_k^{-1})$ because $\delta^{-1/2}n \times \delta^2\eta^{-1}b_k^{-1} \to 0$. The proof is complete.

APPENDIX I

PROOF OF PROPOSITIONS 1–3

It is important to note that $y_t(\alpha)$, which is defined by (2) and (3), can be written explicitly as

$$y_t(\alpha) = \frac{1}{\sin\lambda}\sum_{j=0}^{t}\eta^{j-1}\sin(j\lambda)\,y_{t-j+1}. \tag{23}$$

This expression can be verified simply by substituting (23) into the left-hand side of (2) and confirming that the substitution results in y_t, which is the right-hand side of (2). Finally, it is alway assumed in the sequel that $\lambda \in \Lambda_{nk}$ and $\varepsilon \in (1, \frac{3}{2})$.

A. Proof of Proposition 1

Let $c_k(t) := \beta_k\cos(\omega_k t + \phi_k)$ and $z_k(t) := \sum_{\ell\neq k} c_\ell(t)$. In estimating the kth frequency, $z_k(t)$ can be regarded as the interference from the other sinusoids. By replacing y_t in (23) with its definition given by (1), we can write

$$y_{t-1}(\alpha) = \frac{1}{\sin\lambda}\{u_t(\lambda) + v_t(\lambda) + w_t(\lambda)\}, \tag{24}$$

where

$$\begin{aligned}
u_t(\lambda) &:= \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, c_k(t-j), \\
v_t(\lambda) &:= \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, \varepsilon_{t-j}, \\
w_t(\lambda) &:= \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, z_k(t-j).
\end{aligned}$$

Note that $u_t(\lambda)$ and $v_t(\lambda)$ represent the contributions of the kth sinusoid and the noise, respectively; these terms remain the same as in the case of single sinusoid. The third term $w_t(\lambda)$ is the contribution of the other sinusoids; it is the extra term in the case of multiple sinusoids.

To prove (10) and (11) with $\Psi_n(\lambda)$ defined by (8), we first note that from (24),

$$\begin{aligned}
y_{t-1}^2(\alpha) &= \frac{1}{\sin^2\lambda}\{u_t^2(\lambda) + v_t^2(\lambda) + 2\,u_t(\lambda)\,v_t(\lambda) \\
&\quad + w_t^2(\lambda) + 2\,u_t(\lambda)\,w_t(\lambda) + 2\,v_t(\lambda)\,w_t(\lambda)\}.
\end{aligned} \tag{25}$$

As can be seen, the first three terms in (25) are the same as in the case of single sinusoid. The remaining terms involve the contribution of the other sinusoids. It suffices to show that these terms are asymptotically negligible. The main effort in the proof is to evaluate the contribution $w_t(\lambda)$ of the interfering sinusoids. The proof of Proposition 1 is outline as follows.

Proof of (10). By substituting (25) in the expression of $\Psi_n(\lambda)$ in (8), we can write

$$\begin{aligned}
\Psi_n(\lambda) &= V_n(\lambda) + \sum_{t=1}^{n}\{w_t^2(\lambda) + 2\,u_t(\lambda)\,w_t(\lambda) + 2\,v_t(\lambda)\,w_t(\lambda)\} \\
&:= V_n(\lambda) + T_1(\lambda) + T_2(\lambda) + T_3(\lambda),
\end{aligned}$$

where $V_n(\lambda) := \sum\{u_t^2(\lambda) + 2\,u_t(\lambda)\,v_t(\lambda) + v_t^2(\lambda)\}$. Since $V_n(\lambda)$ is the same as in the case of single sinusoid, by Proposition 1 in [2], $V_n(\lambda)$ has the asymptotic expression in the right-hand side of (10) with $\Delta_k := 1$.

Therefore, it suffices to show that the following expressions hold almost surely and uniformly in $\lambda \in \Lambda_{nk}$ for large n:

$$\begin{aligned} T_1(\lambda) &= \mathcal{O}(\Delta_k^{-4} n) + (\lambda - \lambda_k)\{\mathcal{O}(\Delta_k^{-2}\delta^{-2}n) + \mathcal{O}(\delta^{\varepsilon-4}n)\}, \\ T_2(\lambda) &= \mathcal{O}(\Delta_k^{-2}\delta^{-1}n) + (\lambda - \lambda_k)\{\mathcal{O}(\Delta_k^{-6}\delta^{-1}) + \mathcal{O}(\Delta_k^{-4}\delta^{-2}) \\ &\quad + \mathcal{O}(\Delta_k^{-2}\delta^{-2}n) + \mathcal{O}(\delta^{\varepsilon-4}n)\}, \\ T_3(\lambda) &= \mathcal{O}(\Delta_k^{-2}\delta^{-1/2}n\sqrt{\log n}) + (\lambda - \lambda_k)\,\mathcal{O}(\delta^{-5/2}n\sqrt{\log n}). \end{aligned}$$

A proof of these expressions is given in [29]. Note that the Taylor series expansion (TSE) technique plays an important role in the proof.

Proof of (11). Let $\Psi_n'(\lambda) := T_1(\lambda) + T_2(\lambda) + T_3(\lambda)$. Then, $\Psi_n(\lambda) = V_n(\lambda) + \Psi_n'(\lambda)$. According to Proposition 1 in [2], $V_n(\lambda) - V_n(\lambda')$ has the same asymptotic expression as in the right-hand side of (11) with $\Delta_k := 1$. Furthermore, the TSE of $\Psi_n'(\lambda)$ at λ' can be written as

$$\Psi_n'(\lambda) - \Psi_n'(\lambda') = (\lambda - \lambda')\{\dot{T}_1(\lambda^*) + \dot{T}_2(\lambda^*) + \dot{T}_3(\lambda^*)\},$$

where λ^* lies between λ and λ'. It can be shown [29] that

$$\begin{aligned} \dot{T}_1(\lambda^*) &= \mathcal{O}(\Delta_k^{-2}\delta^{-2}n) + \mathcal{O}(\delta^{\varepsilon-4}n), \\ \dot{T}_2(\lambda^*) &= \mathcal{O}(\Delta_k^{-6}\delta^{-1}) + \mathcal{O}(\Delta_k^{-4}\delta^{-2}) + \mathcal{O}(\Delta_k^{-2}\delta^{-2}n) + \mathcal{O}(\delta^{\varepsilon-4}n), \\ \dot{T}_3(\lambda^*) &= \mathcal{O}(\delta^{-5/2}n\sqrt{\log n}). \end{aligned}$$

Combining these results leads to

$$\begin{aligned} \Psi_n'(\lambda) - \Psi_n'(\lambda') &= (\lambda - \lambda')\{\mathcal{O}(\Delta_k^{-6}\delta^{-1}) + \mathcal{O}(\Delta_k^{-4}\delta^{-2}) \\ &\quad + \mathcal{O}(\Delta_k^{-2}\delta^{-2}n) + \mathcal{O}(\delta^{\varepsilon-4}n) + \mathcal{O}(\delta^{-5/2}n\sqrt{\log n})\}. \end{aligned}$$

The proof is complete.

B. Proof of Proposition 2

It is easy to show from (1), (8), and (23) that

$$\Phi_n(\lambda) = \sum_{t=1}^{n} \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, y_t\, y_{t-j} = U_n(\lambda) + \sum_{i=1}^{5} S_i(\lambda), \tag{26}$$

where

$$\begin{aligned}
U_n(\lambda) &:= \sum_{t=1}^{n} \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, \{c_k(t) + \varepsilon_t\}\, \{c_k(t-j) + \varepsilon_{t-j}\}, \\
S_1(\lambda) &:= \sum_{t=1}^{n} \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, z_k(t)\, c_k(t-j) \\
S_2(\lambda) &:= \sum_{t=1}^{n} \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, z_k(t-j)\, \varepsilon_t, \\
S_3(\lambda) &:= \sum_{t=1}^{n} \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, z_k(t-j)\, c_k(t), \\
S_4(\lambda) &:= \sum_{t=1}^{n} \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, z_k(t)\, \varepsilon_{t-j}, \\
S_5(\lambda) &:= \sum_{t=1}^{n} \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\, z_k(t)\, z_k(t-j).
\end{aligned}$$

In these expressions, $n^{-1}S_1(\lambda)$ is the sample covariance between the kth filtered sinusoid and the interference; $n^{-1}S_2(\lambda)$ is the sample covariance between the noise and the filtered interference; $n^{-1}S_3(\lambda)$ is the sample covariance between the kth unfiltered sinusoid and the filtered interference; $n^{-1}S_4(\lambda)$ is the sample covariance between the filtered noise and the unfiltered interference; and $n^{-1}S_5(\lambda)$ is the covariance between the unfiltered and the filtered interferences. With this notation, the proof of Proposition 2 is outlined as follows.

Proof of (12). By Proposition 2 in [2], $U_n(\lambda)$ has the same asymptotic expression as in the right-hand side of (12) with $\Delta_k := 1$. Therefore, (12) is a direct consequence of the following:

$$\begin{aligned} S_1(\lambda) &= \mathcal{O}(\Delta_k^{-4}) + \mathcal{O}(\Delta_k^{-2}\delta^{-1}) \\ &\quad + (\lambda-\lambda_k)\,\{\mathcal{O}(\Delta_k^{-4}\delta^{-1}) + \mathcal{O}(\Delta_k^{-2}\delta^{-2}) + \mathcal{O}(\delta^{\varepsilon-3}n)\}, \end{aligned} \tag{27}$$

$$\begin{aligned} S_2(\lambda) &= \mathcal{O}(\delta^{-1}\sqrt{n\log n}) + \mathcal{O}(\delta^{-3/2}\sqrt{\log n}) \\ &\quad + (\lambda-\lambda_k)\,\{\mathcal{O}(\delta^{-2}\sqrt{n\log n}) + \mathcal{O}(\delta^{-5/2}\sqrt{\log n})\}, \end{aligned} \tag{28}$$

$$S_3(\lambda) = \text{same as the right-hand side of (27)}, \tag{29}$$

$$S_4(\lambda) = \text{same as the right-hand side of (28)}, \tag{30}$$

$$\begin{aligned} S_5(\lambda) &= n\eta^{-1}\xi_k + \mathcal{O}(\Delta_k^{-4}\delta^2 n) + \mathcal{O}(\Delta_k^{-2}\delta^{-1}) \\ &\quad + (\lambda-\lambda_k)\,\{\mathcal{O}(\Delta_k^{-2}\delta^{-2}) + \mathcal{O}(\Delta_k^{-4}n) + \mathcal{O}(\delta^{\varepsilon-3}n)\}. \end{aligned} \tag{31}$$

where ξ_k is defined in Proposition 2. A proof of these expressions is given in [29]. Note that one only needs to prove (27), (28), and (31) because (29) and (30) can be easily derived from these results by observing the symmetry in their definitions.

Proof of (13). Let $\Phi_n'(\lambda) := \sum_{i=1}^5 S_i(\lambda)$, so that $\Phi_n(\lambda) = U_n(\lambda) + \Phi_n'(\lambda)$. The TSE of $\Phi_n'(\lambda)$ at λ' can be expressed as $\Phi_n'(\lambda) - \Phi_n'(\lambda') = (\lambda - \lambda')\sum_{i=1}^5 \dot{S}_i(\lambda^*)$, where λ^* is between λ and λ'. It can be shown [29] that

$$\begin{aligned} \dot{S}_1(\lambda^*) &= \mathcal{O}(\Delta_k^{-4}\delta^{-1}) + \mathcal{O}(\Delta_k^{-2}\delta^{-2}) + \mathcal{O}(\delta^{\varepsilon-3}n), \\ \dot{S}_2(\lambda^*) &= \mathcal{O}(\delta^{-2}\sqrt{n\log n}) + \mathcal{O}(\delta^{-5/2}\sqrt{\log n}), \\ \dot{S}_5(\lambda^*) &= \mathcal{O}(\Delta_k^{-2}\delta^{-2}) + \mathcal{O}(\Delta_k^{-4}n) + \mathcal{O}(\delta^{\varepsilon-3}n). \end{aligned}$$

Moreover, $\dot{S}_3(\lambda^*)$ has the same expression as $\dot{S}_1(\lambda^*)$ and $\dot{S}_4(\lambda^*)$ has the same expression as $\dot{S}_2(\lambda^*)$. The proof is complete upon noting that $U_n(\lambda) - U_n(\lambda')$, by Proposition 2 in [2], has the same asymptotic expression as in the right-hand side of (13) with $\Delta_k := 1$.

C. Proof of Proposition 3

Consider (26). According to Proposition 3 in [2],

$$U_n(\lambda_k) = W_{n1} + W_{n2} + \mathcal{O}_P(\delta^2 n) + \mathcal{O}_P(n^{1/2}) + \mathcal{O}_P(\delta^{-1}).$$

This, combined with (26) implies that it suffices to show that

$$S_i(\lambda_k) = \begin{cases} \mathcal{O}_P(\Delta_k^{-4}) + \mathcal{O}_P(\Delta_k^{-2}\delta^{-1}) & \text{if } i = 1,3, \\ \mathcal{O}_P(\Delta_k^{-2} n^{1/2}) + \mathcal{O}_P(\Delta_k^{-2}\delta^{-1/2}) & \text{if } i = 2,4, \\ n\eta^{-1}\xi_k + \mathcal{O}_P(\Delta_k^{-4}\delta^2 n) + \mathcal{O}_P(\Delta_k^{-2}\delta^{-1}) & \text{if } i = 5. \end{cases}$$

A proof of these expressions is given in [29].

APPENDIX II

SOME USEFUL EXPRESSIONS

Let $c_k(t) := \beta_k \cos(\omega_k t + \phi_k)$, $s_k(t) := \beta_k \sin(\omega_k t + \phi_k)$, $f_{t\ell}(\lambda) := \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\cos(j\omega_\ell)$, $g_{t\ell}(\lambda) := \sum_{j=0}^{t-1} \eta^{j-1} \sin(j\lambda)\sin(j\omega_\ell)$. Then, the following assertions are true (see [29] for proofs).

Lemma 1 For $r = 0,1,2$ and $\ell \neq k$, if $\Delta_{\ell k}^{-1}\delta \to 0$ as $n \to \infty$, both $f_{t\ell}^{(r)}(\omega_k)$ and $g_{t\ell}^{(r)}(\omega_k)$ can be expressed as $\mathcal{O}(\Delta_{\ell k}^{-2(r+1)}) + \sum_{j=0}^{r} \mathcal{O}(\Delta_{\ell k}^{-2(r-j+1)} t^j \eta^t)$, which holds uniformly in t and η.

Lemma 2 Assume that $\delta \to 0$, $n(1-\delta)^n = \mathcal{O}(1)$, and $\Delta^{-1}\delta \to 0$ as $n \to \infty$.

(a) For $r = 0,1,2$, $f_{t\ell}^{(r)}(\lambda) = \mathcal{O}(\delta^{-r-1})$, $\quad g_{t\ell}^{(r)}(\lambda) = \mathcal{O}(\delta^{-r-1})$ uniformly in t, λ, and η.

(b) For $r = 0,1,2$, if $\lambda - \omega_k = \mathcal{O}(\delta^q)$ for some $q > 0$, then $f_{t\ell}^{(r)}(\lambda) = f_{t\ell}^{(r)}(\omega_k) + \mathcal{O}(\delta^{q-r-2})$ and $g_{t\ell}^{(r)}(\lambda) = g_{t\ell}^{(r)}(\omega_k) + \mathcal{O}(\delta^{q-r-2})$ uniformly in t, η, and λ.

(c) For $\ell \neq k$ and $\lambda \in \Lambda_{nk}$, if $\Delta_{\ell k}^{-1}\delta \to 0$ as $n \to \infty$, then $f_{t\ell}(\lambda)$ and $g_{t\ell}(\lambda)$ can be expressed as $\mathcal{O}(\Delta_{\ell k}^{-2}) + \mathcal{O}(\delta^{\varepsilon-2})$, which holds uniformly in t, η, and λ.

Lemma 3 Assume that $\delta \to 0$, $n(1-\delta)^n = \mathcal{O}(1)$, and $\Delta_k^{-1}\delta \to 0$ as $n \to \infty$. Then, for $\ell' \neq k$, the following expressions are true.

(a)

$$\sum_{t=1}^{n} c_{\ell'}(t)\, c_\ell(t)\, f_{t\ell}(\lambda_k) \;=\; \begin{cases} n\eta^{-1}\xi_{\ell k} + \mathcal{O}(\Delta_{\ell k}^{-4}\delta^2 n) + \mathcal{O}(\Delta_{\ell k}^{-2}\delta^{-1}) & \text{if } \ell' = \ell \neq k, \\ \min\{\mathcal{O}(\Delta_{\ell k}^{-2} n), \mathcal{O}(\Delta_{\ell k}^{-2}\Delta_{\ell'\ell}^{-1})\} + \mathcal{O}(\Delta_{\ell k}^{-2}\delta^{-1}) & \text{if } \ell' \neq \ell \neq k, \\ \mathcal{O}(\Delta_{\ell' k}^{-4}) + \mathcal{O}(\delta^{-1}) & \text{if } \ell' \neq \ell = k, \end{cases} \tag{32}$$

$$\sum_{t=1}^{n} c_{\ell'}(t)\, s_\ell(t)\, g_{t\ell}(\lambda_k) \;=\; \begin{cases} \mathcal{O}(\Delta_{\ell k}^{-2}\delta^{-1}) & \text{if } \ell' = \ell \neq k, \\ \text{same as in (32)} & \text{if } \ell' \neq \ell \neq k, \\ \mathcal{O}(\Delta_{\ell' k}^{-4}) + \mathcal{O}(\Delta_{\ell' k}^{-2}\delta^{-1}) & \text{if } \ell' \neq \ell = k, \end{cases}$$

where $\xi_{\ell k} := \frac{1}{8}\beta_\ell^2\{\cot(\frac{1}{2}\omega_{k\ell}^-) + \cot(\frac{1}{2}\omega_{k\ell}^+)\}$. In addition,

$$\sum_{t=1}^{n} c_k^2(t)\, f_{tk}(\lambda_k) = \mathcal{O}(\delta^2 n) + \mathcal{O}(\delta^{-1}).$$

(b)

$$\begin{aligned} \sum_{t=1}^{n} c_{\ell'}(t)\, c_\ell(t)\, \dot{f}_{t\ell}(\lambda_k) &= \mathcal{O}(\delta^{-1} n) \\ &\quad + \begin{cases} \mathcal{O}(\Delta_{\ell k}^{-4} n) + \mathcal{O}(\Delta_{\ell k}^{-2}\delta^{-2}) & \text{if } \ell \neq k, \\ \mathcal{O}(\Delta_{\ell' k}^{-4}\delta^{-1}) + \mathcal{O}(\Delta_{\ell' k}^{-2}\delta^{-2}) & \text{if } \ell = k, \end{cases} \end{aligned} \tag{33}$$

$$\begin{aligned} \sum_{t=1}^{n} c_{\ell'}(t)\, s_\ell(t)\, \dot{g}_{t\ell}(\lambda_k) &= \mathcal{O}(\delta^{-1} n) \\ &\quad + \begin{cases} \mathcal{O}(\Delta_{\ell k}^{-4}\delta^{-1}) + \mathcal{O}(\Delta_{\ell k}^{-2}\delta^{-2}) & \text{if } \ell' = \ell \neq k, \\ \text{same as in (33)} & \text{if } \ell' \neq \ell \neq k, \\ \mathcal{O}(\delta^{-2}) & \text{if } \ell' \neq \ell = k. \end{cases} \end{aligned}$$

(c)

$$\sum_{t=1}^{n} c_\ell(t)\, c_{\ell'}(t)\, f_{t\ell}(\lambda_k)\, \dot{f}_{t\ell'}(\lambda_k) = \mathcal{O}(\Delta_{\ell' k}^{-4} n) + \begin{cases} \mathcal{O}(\Delta_{\ell' k}^{-4}\Delta_{\ell k}^{-2} n) + \mathcal{O}(\Delta_{\ell' k}^{-2}\Delta_{\ell k}^{-2}\delta^{-2}) + \mathcal{O}(\Delta_{\ell k}^{-2}\delta^{-1} n) & \text{if } \ell \neq k, \\ \mathcal{O}(\Delta_{\ell' k}^{-4}\delta^{-1}) + \mathcal{O}(\Delta_{\ell' k}^{-2}\delta^{-2}) + \mathcal{O}(\delta^{-1} n) & \text{if } \ell = k, \end{cases} \tag{34}$$

$$\sum_{t=1}^{n} s_\ell(t)\, c_{\ell'}(t)\, g_{t\ell}(\lambda_k)\, \dot{f}_{t\ell'}(\lambda_k) = \mathcal{O}(\Delta_{\ell' k}^{-4} n) + \begin{cases} \mathcal{O}(\Delta_{\ell k}^{-6}\delta^{-1}) + \mathcal{O}(\Delta_{\ell k}^{-4}\delta^{-2}) + \mathcal{O}(\Delta_{\ell k}^{-2}\delta^{-1} n) & \text{if } \ell' = \ell \neq k, \\ \text{same as in (34)} & \text{if } \ell' \neq \ell \neq k, \\ \mathcal{O}(\Delta_{\ell' k}^{-6}\delta^{-1}) + \mathcal{O}(\Delta_{\ell' k}^{-4}\delta^{-2}) + \mathcal{O}(\Delta_{\ell' k}^{-2}\delta^{-3}) + \mathcal{O}(\delta^{-2} n) & \text{if } \ell' \neq \ell = k, \end{cases} \tag{35}$$

$$\sum_{t=1}^{n} c_\ell(t)\, s_{\ell'}(t)\, f_{t\ell}(\lambda_k)\, \dot{g}_{t\ell'}(\lambda_k) = \begin{cases} \text{same as in (35)} & \text{if } \ell' = \ell \neq k, \\ \text{same as in (34)} & \text{if } \ell' \neq \ell \neq k \text{ or } \ell' \neq \ell = k, \end{cases}$$

$$\sum_{t=1}^{n} s_\ell(t)\, s_{\ell'}(t)\, g_{t\ell}(\lambda_k)\, \dot{g}_{t\ell'}(\lambda_k) = \begin{cases} \text{same as in (34)} & \text{if } \ell \neq k, \\ \text{same as in (35)} & \text{if } \ell = k. \end{cases}$$

The next lemma is also instrumental.

Lemma 4 [27] As $n \to \infty$, the following expressions are true.

(a) $\max_\omega |\sum_{t=1}^{n} t^r \eta^t \varepsilon_t \exp(it\omega)| = \mathcal{O}(\delta^{-r-1/2}\sqrt{\log n})$ almost surely for any integer $r \geq 0$.

(b) $|\sum_{t=1}^{n} \varepsilon_t \exp(it\omega) \sum_{j=1}^{t} j^r \eta^j \exp(ij\omega')| = \mathcal{O}(\delta^{-r-1}\sqrt{n\log n}) + \mathcal{O}(\delta^{-r-3/2}\sqrt{\log n})$ almost surely and uniformly in ω and ω' for $r = 0, 1, 2$.

(c) $\sum_{t=1}^{n} \varepsilon_t \exp(it\omega) = \mathcal{O}_P(n^{1/2})$ and $\max_{1\leq t\leq n} |\sum_{j=1}^{t} j^r \eta^j \varepsilon_j \sin(j\omega)| = \mathcal{O}_P(\delta^{-r-1/2})$ for any given ω and any integer $r \geq 0$.

References

[1] J. A. Rice and M. Rosenblatt, “On frequency estimation,” *Biometrika*, vol. 75, pp. 477–484, 1988.

[2] K. S. Song and T. H. Li, “A statistically and computationally efficient method for frequency estimation,” *Stochastic Process. Appl.*, vol. 86, pp. 29–47, 2000.

[3] E. J. Hannan, “Nonlinear time series regression,” *J. Appl. Prob.*, vol. 8, pp. 767–780, 1971.

[4] A. M. Walker, “On the estimation of a harmonic component in a time series with stationary independent residuals,” *Biometrika*, vol. 58, pp. 21–36, 1971.

[5] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Englewood Cliffs, NJ: Prentice-Hall, 1988.

[6] P. Stoica and A. Nehorai, “Statistical analysis of two nonlinear least-squares estimators of sine-wave parameters in the colored-noise case,” *Circuits, Systems, Signal Process.*, vol. 8, pp. 3–15, 1989.

[7] C. R. Rao and L. C. Zhao, “Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals,” *IEEE Trans. Signal Process.*, vol. 41, pp. 1461–1464, 1993.

[8] D. C. Rife and R. R. Boorstyn, “Single-tone parameter estimation from discrete-time observations,” *IEEE Trans. Inform. Theory*, vol. 20, pp. 591–598, 1974.

[9] P. Stoica, R. L. Moses, B. Friedlander, and T. Söderström, “Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 378–392, 1989.

[10] S. M. Kay, “Accurate frequency estimation at low signal-to-noise ratio,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, pp. 540–547, 1984.

[11] A. Nehorai, “A minimum parameter adaptive notch filter with constrained poles and zeros,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, pp. 983–996, 1985.

[12] M. V. Dragoševiȁ and S. S. Stankoviȁ, “A generalized least squares method for frequency estimation,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 805–819, 1989.

[13] B. G. Quinn and J. M. Fernandes, “A fast efficient technique for the estimation of frequency,” *Biometrika*, vol. 78, pp. 489–497, 1991.

[14] T. H. Li and B. Kedem, “Iterative filtering for multiple frequency estimation,” *IEEE Trans. Signal Processs.*, vol. 42, pp. 1120–1132, 1994.

[15] B. Kedem and S. Yakowitz, “Practical aspects of a fast algorithm for frequency detection,” *IEEE Trans. Commun.*, vol. 42, pp. 2760–2767, 1994.

[16] J. A. Chambers and A. G. Constantinides, “Frequency tracking using constrained adaptive notch filters synthesised from allpass sections,” *IEE Proc. Part F*, vol. 137, pp. 475–481, 1990.

[17] J. F. Chicharo and T. S. Ng, “Gradient-based adaptive IIR notch filtering for frequency estimation,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol 38, pp. 769–777, 1990.

[18] B. S. Chen, T. Y. Yang, and B. H. Lin, “Adaptive notch filter by direct frequency estimation,” *Signal Process.*, vol. 27, pp. 161–176, 1992.

[19] T. H. Li and B. Kedem, “Tracking abrupt frequency changes,” *J. Time Series Anal.*, vol. 19, pp. 69–82, 1998.

[20] T. H. Li and B. Kedem, “Asymptotic analysis of a multiple frequency estimation method,” *J. Multivariate Anal.*, vol. 46, pp. 214–236, 1993.

[21] T. Kwan and K. Martin, “Adaptive detection and enhancement of multiple sinusoids using a cascade IIR filter,” *IEEE Trans. Circuits, Systems*, vol 36, pp. 937–946, 1989.

[22] T. H. Li and B. Kedem, “Strong consistency of the contraction mapping method for frequency estimation,” *IEEE Trans. Inform. Theory*, vol 39, pp. 989–998, 1993.

[23] S. He and B. Kedem, “Higher-order crossings of an almost periodic random sequence in noise,” *IEEE Trans. Inform. Theory*, vol. 35, pp. 360–370, 1989.

[24] S. Yakowitz, “Some contribution to a frequency location method due to He and Kedem,” *IEEE Trans. Inform. Theory*, vol. 37, pp. 1177–1182, 1991.

[25] B. Kedem, *Time Series Analysis by Higher-Order Crossings*, Piscataway, NJ: IEEE Press, 1994.

[26] T. H. Li, B. Kedem, and S. Yakowitz, “Asymptotic normality of the sample autocovariances with an application in frequency estimation,” *Stochastic Process. Appl.*, vol. 52, pp. 329–349, 1994.

[27] K. S. Song and T. H. Li, “Asymptotic theory of a frequency estimator,” Tech. Rep. M917, Statistics Dept., Florida State Univ., Tallahassee, 1997. Electronic version available at `http://www.research.ibm.com/people/t/thl`.

[28] P. Stoica and T. Söderström, “Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies,” *IEEE Trans. Signal Process.*, vol. 39, pp. 1836–1847, 1991.

[29] T. H. Li and K. S. Song, “Asymptotic analysis of a contraction mapping algorithm for multiple frequency estimation,” Tech. Rep. M958, Statistics Dept., Florida State Univ., Tallahassee, 2001. Electronic version available at `http://www.research.ibm.com/people/t/thl`.

[30] T. H. Li, “A fast algorithm for efficient estimation of frequencies,” *Signal Processing IX: Theories and Applications*, S. Theodoridis, I. Pitas, A. Stouraitis, and N. Kalouptsidis, Eds. Patras, Greece: Typorama Editions, vol. 1, pp. 65–68, 1998.

[31] J. Stoer and R. Bulirsch, *Introduction to Numerical Analysis*, New York: Springer-Verlag, 1980.

[32] E. L. Lehmann, *Theory of Point Estimation*, New York: Wiley, 1983.

Ta-Hsin Li (S'89, M'92) received the B.A. degree in mathematics from Beijing Institute of Posts and Telecommunications, Beijing, China, in 1982, the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1984, the M.A. degree in statistics from the American University, Washington, DC, in 1989, and the Ph.D. degree in applied mathematics from the University of Maryland at College Park in 1992.

From 1984 to 1987, he was Instructor of Electrical Engineering at Tsinghua University, Beijing, China. From 1992 to 1997, he was Assistant Professor and Associate Professor of Statistics at Texas A&M University, College Station. In 1998, he joined the University of California at Santa Barbara as Associate Professor of Statistics. Since 1999, he has been with IBM T. J. Watson Research Center at Yorktown Heights, NY. His current research interests include time series analysis, spectral and wavelet analysis, statistical signal processing in communications and biomedical applications, and spatial-temporal data analysis in environmental sciences.

Dr. Li is a member of Phi Kappa Phi, the American Statistical Association, and the IEEE Societies of Signal Processing and Information Theory. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

Kai-Sheng Song received the Ph.D. degree in statistics from the University of California, Davis, in 1993.

During 1993–1994, he was a Visiting Assistant Professor in the Department of Statistics at Purdue University, West Lafayette, IN, and during 1994–1995, at Texas A&M University, College Station. In 1995, he joined the faculty of the Florida State University, Tallahassee, where he is presently an Associate Professor of Statistics. His current research interests include information theory, signal processing, nonparametrics, and quantum computing.

Dr. Song is the recipient of several achievement awards including the Julius R. Blum Memorial Award in 1992 from the University of California, Davis, and the University Teaching Award in 2000 from the Florida State University, Tallahassee.