RC23479 Revised (W0507-137) July 19, 2005
Mathematics

IBM Research Report

On Convergence and Bias Correction of a Joint Estimation Algorithm for Multiple Sinusoidal Frequencies

Kai-Sheng Song
Department of Statistics
Florida State University
Tallahassee, FL 32306-4330

Ta-Hsin Li
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

On Convergence and Bias Correction of a Joint Estimation Algorithm for Multiple Sinusoidal Frequencies

(Complete Version)

Kai-Sheng Song

Department of Statistics

Florida State University

Tallahassee, FL 32306-4330

E-mail: kssong@stat.fsu.edu

Ta-Hsin Li

Department of Mathematical Sciences

IBM T. J. Watson Research Center

Yorktown Heights, NY 10598-0218

E-mail: thl@us.ibm.com

July 14, 2005

Abstract

Twenty years ago Kay (1984) proposed an iterative filtering algorithm (IFA) for jointly estimating the frequencies of multiple complex sinusoids from noisy observations. It is based on the fact that the noiseless signal is an autoregressive (AR) process, so the frequency estimation problem can be reformulated as the problem of estimating the AR coefficients. By iterating the cycle of AR coefficient estimation and AR filtering, IFA provides a computationally simple procedure yet capable of accurate frequency estimation especially at low signal-to-noise ratio (SNR). However, the convergence of IFA has not been established beyond simulation and a very special case of a single frequency and infinite sample size. This paper provides a statistical analysis of the algorithm and makes several important contributions: (a) It shows that the poles of the AR filter must be reduced via an extra shrinkage parameter in order to accommodate poor initial values and avoid being trapped into false solutions. (b) It shows that the AR estimates in each iteration must be bias-corrected in order to produce a more accurate frequency estimator; a closed-form expression is provided for bias correction. (c) It shows that for a sufficiently large sample size, the resulting algorithm, called new IFA, or NIFA, converges to the desired fixed point which constitutes a consistent frequency estimator. Numerical examples, including a real data example in radar applications, are provided to demonstrate the findings. It is shown in particular that the shrinkage parameter not only controls the estimation accuracy but also determines the requirements of initial values. It is also shown that the proposed bias-correction method considerably improves the estimation accuracy, especially for high SNR.

Abbreviated Title. Multiple Frequency Estimation

Keywords. Adaptive filter; Convergence; Contractive mapping; Doppler; Fixed point; Frequency estimation; Mixed spectrum; Signal processing; Sinusoid.

1. INTRODUCTION

As an alternative to the Gaussian maximum likelihood method, the iterative filtering algorithm (IFA) was proposed by Kay (1984) for estimating the frequencies of multiple complex sinusoids from noisy observations

$$y_t := x_t + \varepsilon_t, \quad x_t := \sum_{k=1}^{p} \beta_k e^{i(\omega_k t + \phi_k)} \quad (t = 1, \ldots, n), \tag{1}$$

where $p \geq 1$ is a known integer, $\beta_k > 0$, $\omega_k := 2\pi f_k \in (-\pi, \pi) \setminus \{0\}$, and $\phi_k \in (-\pi, \pi]$ are unknown constants, and $\{\varepsilon_t\}$ is a zero-mean complex-valued white noise process with unknown variance σ^2. IFA is based on the fact that the noiseless signal $\{x_t\}$ is a special autoregressive (AR) process of order p satisfying $x_t + \sum_{k=1}^{p} a_k x_{t-k} = 0$, where the a_k uniquely determine, and are determined by, the frequencies ω_k such that

$$1 + \sum_{k=1}^{p} a_k z^{-k} = \prod_{k=1}^{p} (1 - e^{i\omega_k} z^{-1}), \tag{2}$$

so the frequency estimation problem can be reformulated as the problem of estimating

$$\mathbf{a} := [a_1, \ldots, a_p]^T.$$

To estimate the AR parameter $\mathbf{a}$, IFA iterates a cycle of estimation and filtering: It starts with an initial estimate $\hat{\mathbf{a}} := [\hat{a}_1, \ldots, \hat{a}_p]^T$ obtained from $\{y_t\}$ and uses it to construct an AR filter which is applied to $\{y_t\}$ to produce

$$\tilde{y}_t = -\sum_{k=1}^{p} \hat{a}_k \tilde{y}_{t-k} + y_t \quad (t = 1, \ldots, n), \tag{3}$$

where $\tilde{y}_{-1} := \tilde{y}_0 := 0$. Then it re-estimates the AR parameter from the filtered time series $\{\tilde{y}_t\}$ and uses the new estimate to filter the original data $\{y_t\}$ in the same way as (3) to produce a new filtered time series. This cycle is repeated until a stopping criterion is satisfied. Li and Kedem (1994) generalized the basic idea of IFA by including other parametric filters and by regarding the cycle of estimation and filtering as a fixed-point iteration. The resulting procedure is called parametric filtering. IFA is simple computationally yet capable of

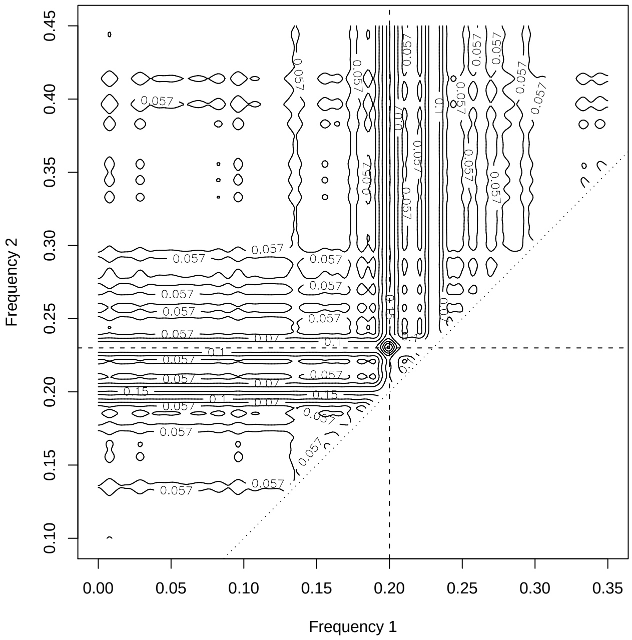

Figure 1: *Contour plot of the Gaussian likelihood, as a function of the normalized frequencies* f_1 *and* f_2*, for two complex sinusoids in complex Gaussian white noise (*$n = 100$ *and SNR* $= 0$ *dB). Dashed lines indicate the location of the true frequencies (*$f_1 = 0.2$ *and* $f_2 = 0.23$*); dotted diagonal line shows the boundary* $f_1 = f_2$.

providing accurate frequency estimates. However, the convergence of IFA has not been established beyond simulation and a very special case of a single frequency and infinite sample size.

The Gaussian maximum likelihood method (also known as nonlinear least squares) has been applied to the problem of frequency estimation (Walker 1971; Rife and Boorstyn 1976; Abatzoglou 1985; Rice and Rosenblatt 1988; Stoica, Moses, Friedlander, and Söderström 1989; Stoica and Nehorai 1989; Van Hamme 1991; Starer and Nehorai 1992; White 1993; Shaw 1995). Although it produces a statistically efficient estimator that attains the Cramér-Rao lower bound (CRLB) when $\{\varepsilon_t\}$ is Gaussian, the surface of the Gaussian likelihood function comprises numerous local extrema, as shown in Fig. 1, so that an initial value of precision $\mathcal{O}(n^{-1})$ is usually required in order for standard optimization algorithms, such as Newton's method, to converge to the desired solution. Since the n-point discrete Fourier transform (DFT) only produces an

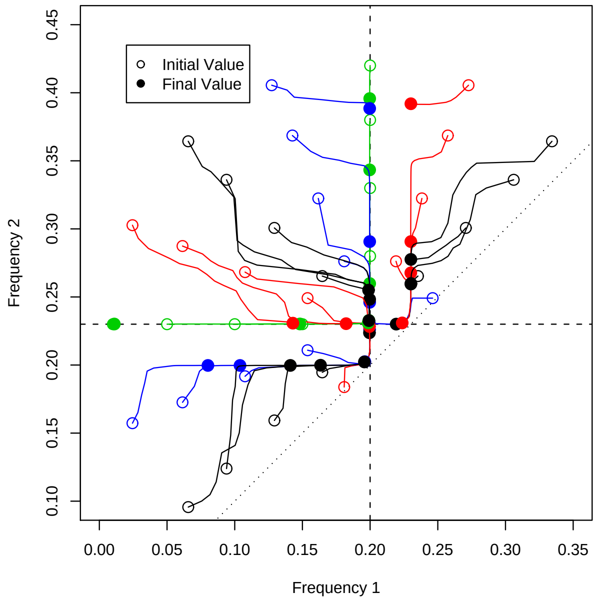

Figure 2: *Trajectory of the frequency estimates produced by IFA for the same data as used in Fig. 1. Open circles represent initial values; solid points represent final estimates after 15 iterations; intermediate estimates appear as lines that link the open circles with the solids points.*

estimator of accuracy $\mathcal{O}(n^{-1})$, interpolation techniques have been proposed to obtain improved initial values from DFT, but convergence with such initial values is still not guaranteed mathematically. IFA suffers from the same initial value problem, as shown in Fig. 2.

In this paper, we provide a rigorous statistical analysis of IFA in the case of $p = 2$ and make several important contributions:

(a) It is shown that the poles of the AR filter must be reduced via an extra shrinkage parameter in order to accommodate poor initial values and avoid being trapped into spurious solutions.

(b) It is shown that the AR estimates in each iteration must be bias-corrected in order to produce a more accurate frequency estimator; a simple closed-form expression is derived for bias correction.

(c) It is shown that with probability tending to unity as the sample size grows, the resulting algorithm, which we call the *new* IFA, or NIFA, converges to the desired fixed-point which constitutes a consistent frequency estimator.

Numerical examples are provided to illustrate these findings. The results can be regarded as an extension of the earlier work of Li and Kedem (1994), Li and Song (2001; 2002), and Song and Li (1997; 2000). Unlike the algorithms discussed there, the algorithm in this paper estimates the frequencies jointly as a multivariate problem rather than sequentially as a sequence of univariate problems (one frequency at a time). The benefit of the joint estimation approach is the relaxation of the frequency separation requirement (Kay 1984).

It is worth pointing out that by cascading the AR fitting with the AR filtering a notch filter can be obtained. It can be implemented as an adaptive filter for tracking time-varying frequencies. This is an advantage of the present approach that the DFT approach does not have. The results in this paper remain valid for the steady-state performance of the notch filtering algorithm.

2. NEW ITERATIVE FILTERING ALGORITHM (NIFA)

Let us assume in the remainder of the paper that $\{y_t\}$ is given by (1) with $p=2$ and $\omega_1 < \omega_2$. Under this assumption, it is easy to show from (2) that $\mathbf{a} = [a_1, a_2]^T$ takes the form

$$a_1 := -\left(e^{i\omega_1} + e^{i\omega_2}\right), \quad a_2 := e^{i(\omega_1+\omega_2)}. \tag{4}$$

For any admissible variable $\boldsymbol{\alpha} := [\alpha_1, \alpha_2]^T$, which will be defined later, let $\{y_t(\boldsymbol{\alpha})\}$ denote the filtered time series given by

$$y_t(\boldsymbol{\alpha}) = -\sum_{k=1}^{2} \alpha_k \eta^k y_{t-k}(\boldsymbol{\alpha}) + y_t \quad (t = 1, \ldots, n) \tag{5}$$

with $y_{-1}(\boldsymbol{\alpha}) := y_0(\boldsymbol{\alpha}) := 0$, where $\eta \in (0,1)$ is the shrinkage parameter that contracts the poles of the filter towards the origin and thus stabilizes the filter.

Given the filtered data $\{y_t(\boldsymbol{\alpha})\}$, we estimate $\mathbf{a}$ by the method of least squares (LS), i.e., by seeking $\hat{\mathbf{a}}(\boldsymbol{\alpha})$

that minimizes $\|\mathbf{y}(\boldsymbol{\alpha})+\mathbf{Y}(\boldsymbol{\alpha})\,\mathbf{D}\,\hat{\mathbf{a}}(\alpha)\|^2$, where

$$\mathbf{Y}(\boldsymbol{\alpha}) := \begin{bmatrix} y_2(\boldsymbol{\alpha}) & y_1(\boldsymbol{\alpha}) \\ \vdots & \vdots \\ y_{n-1}(\boldsymbol{\alpha}) & y_{n-2}(\boldsymbol{\alpha}) \end{bmatrix}, \quad \mathbf{y}(\boldsymbol{\alpha}) := \begin{bmatrix} y_3(\boldsymbol{\alpha}) \\ \vdots \\ y_n(\boldsymbol{\alpha}) \end{bmatrix}, \quad \mathbf{D} := \mathrm{diag}(\eta, \eta^2).$$

This gives rise to an AR estimator

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) := -\mathbf{D}^{-1}\{\mathbf{Y}^H(\boldsymbol{\alpha})\,\mathbf{Y}(\boldsymbol{\alpha})\}^{-1}\mathbf{Y}^H(\boldsymbol{\alpha})\,\mathbf{y}(\boldsymbol{\alpha}), \tag{6}$$

where superscript H stands for Hermitian transpose. Note that the role of $\mathbf{D}$ is to compensate for the shrinkage parameter η.

Unlike the true AR parameter $\mathbf{a}$, the AR estimator $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ does not necessarily correspond to a polynomial of the form (2) which has unit roots. To impose this constraint on each root of the polynomial of $\hat{\mathbf{a}}(\boldsymbol{\alpha})$, one can simply reset its modulus to unity while retaining its angle, thus projecting it on the unit circle. The resulting AR estimator is denoted by $\boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}))$, where $\boldsymbol{\psi}(\cdot)$ represents the unit-root (UR) projector. The UR projection not only stabilizes the AR filter but also eliminates the redundancy in the AR reparameterization of the frequency estimation problem: it reduces the number of free parameters from four (the real and imaginary parts of the AR coefficients) to two (the angles of the roots), which is identical to the number of unknown frequencies. As in regression problems, using the minimal number of free parameters is helpful in reducing the statistical variability of the estimation, especially when the sample size is small or the signal-to-noise ratio (SNR) is low.

With the AR estimator $\boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}))$ so defined, one can apply the general procedure of parametric filtering proposed by Li and Kedem (1994) and seek a fixed point of the mapping $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}))$ by the so-called fixed-point iteration

$$\boldsymbol{\alpha}_m = \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}_{m-1})) \quad (m = 1, 2, \ldots). \tag{7}$$

This algorithm can be decomposed into repeated cycles of AR filtering, LS estimation, and UR projection.

The IFA of Kay (1984) can be regarded as a special case of (7) with $\eta = 1$, although, strictly speaking, it employs Burg's estimator rather than the LS estimator and it does not impose the UR constraint. Fig. 2 shows that with $\eta = 1$ the iteration in (7) may converge to spurious fixed points if the initial values are not near the desired solution. This problem can be overcome, as will be seen later, by choosing $\eta < 1$.

A careful analysis shows that in the case of $\eta < 1$ the mapping $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ contains a bias term at $\mathbf{a}$ that can be expressed as $\mathbf{b} := [b_1, b_2]^T$ where

$$b_k := \frac{(1-\eta) i^k (e^{i\omega_2} - e^{i\omega_1})^k \sin^{2-k}(\omega_2 - \omega_1)}{\eta^k \{1 - \cos(\omega_2 - \omega_1)\}} \quad (k = 1, 2). \tag{8}$$

By subtracting it from $\hat{\mathbf{a}}(\boldsymbol{\alpha})$, a new mapping $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b})$ is formed. Theorem 1 in the next section establishes that with probability tending to unity, the new mapping is contractive in a neighborhood of $\mathbf{a}$ and therefore has a unique fixed point that can be obtained as the limiting value of the fixed-point iteration $\boldsymbol{\alpha}_m := \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}_{m-1}) - \mathbf{b})$ under fairly mild initial conditions. However, this procedure is not yet practical because the bias given by (8) depends on the true frequencies. One way of making it practical is to substitute $\mathbf{b}$ with an estimate $\hat{\mathbf{b}}$ obtained from an initial estimate of $\mathbf{a}$ (or equivalently, of the frequencies). Theorem 2 establishes that the convergence assertions in Theorem 1 remain valid for the resulting mapping $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \hat{\mathbf{b}})$ as long as the initial estimate of $\mathbf{a}$ is sufficiently accurate.

Alternatively, the bias can be re-estimated in each iteration using the estimate of $\mathbf{a}$ from the previous iteration. This gives rise to our final NIFA algorithm

$$\boldsymbol{\alpha}_m := \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}_{m-1}) - \mathbf{b}(\boldsymbol{\alpha}_{m-1})) \quad (m = 1, 2, \ldots), \tag{9}$$

where $\mathbf{b}(\boldsymbol{\alpha}) := [b_1(\boldsymbol{\alpha}), b_2(\boldsymbol{\alpha})]^T$ is defined in the same way as $\mathbf{b}$ by (8) except that ω_1 and ω_2 are replaced by λ_1 and λ_2 which are the angles of the roots of $1 + \alpha_1 z^{-1} + \alpha_2 z^{-2}$ satisfying $\lambda_1 \le \lambda_2$. Let $\hat{\mathbf{a}} := [\hat{a}_1, \hat{a}_2]^T$ be the limiting value of $\{\boldsymbol{\alpha}_m\}$ in (9) as $m \to \infty$. Then, the final NIFA frequency estimator is given by $\hat{\boldsymbol{\omega}} := [\hat{\omega}_1, \hat{\omega}_2]^T$, where $\hat{\omega}_1 = 2\pi \hat{f}_1$ and $\hat{\omega}_2 = 2\pi \hat{f}_2$ are defined as the angles of the roots of $1 + \hat{a}_1 z^{-1} + \hat{a}_2 z^{-2}$ satisfying $\hat{\omega}_1 \le \hat{\omega}_2$. Similarly, one can define the intermediate frequency estimates $\hat{\boldsymbol{\omega}}_m$ $(m = 0, 1, 2, \ldots)$ in

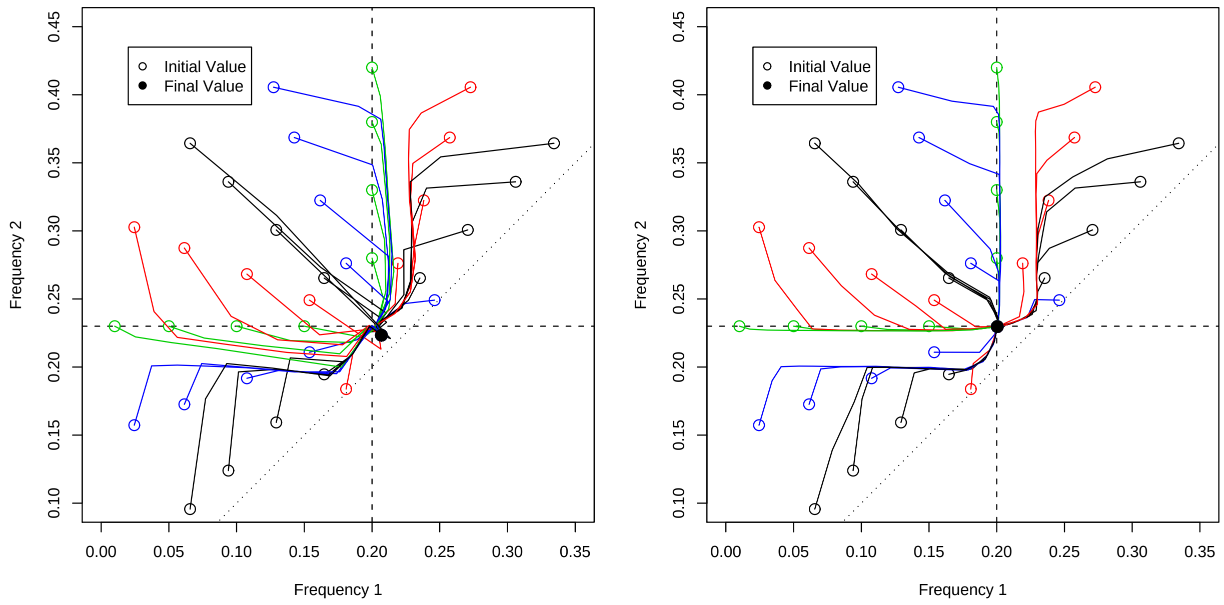

Figure 3: *Trajectory of the frequency estimates produced by NIFA in (9) with different values of η and 15 iterations for the same data as used in Figs. 1–2. Left, $\eta = 0.92$; right, $\eta = 0.96$.*

terms of the roots of the polynomial associated with $\boldsymbol{\alpha}_m$. By definition, $\hat{\boldsymbol{\omega}}_m \to \hat{\boldsymbol{\omega}}$ as $m \to \infty$.

Fig. 3 shows the trajectory of the frequency estimates $\{\hat{\boldsymbol{\omega}}_m\}$ produced by NIFA in (9) with $\eta = 0.92$ and $\eta = 0.96$ for the same data as used in Fig. 2. Fig. 4 shows the corresponding AR estimates $\{\boldsymbol{\alpha}_m\}$ for $\eta = 0.96$. As can be seen, the spurious fixed points in Fig. 2 no longer exist in Fig. 3 where all initial values lead to a single fixed point even if they are far away from the true frequencies. This implies that the initial requirement of NIFA, with $\eta < 1$, is much less stringent than that of IFA where $\eta = 1$. Note that in general η should not be made too close to unity either, otherwise (e.g., $\eta = 0.99$, not shown), the problem of spurious fixed points may appear again, just like when $\eta = 1$ in IFA.

The shrinkage parameter η plays a vital role not only in determining the requirement of initial values, but also in determining the accuracy of the final estimator. Fig. 3 shows that the fixed point of NIFA with the larger $\eta = 0.96$ is much closer to the true frequencies than that with the smaller $\eta = 0.92$ where the estimates are pushed towards a single value in the vicinity of $\frac{1}{2}(f_1 + f_2)$.

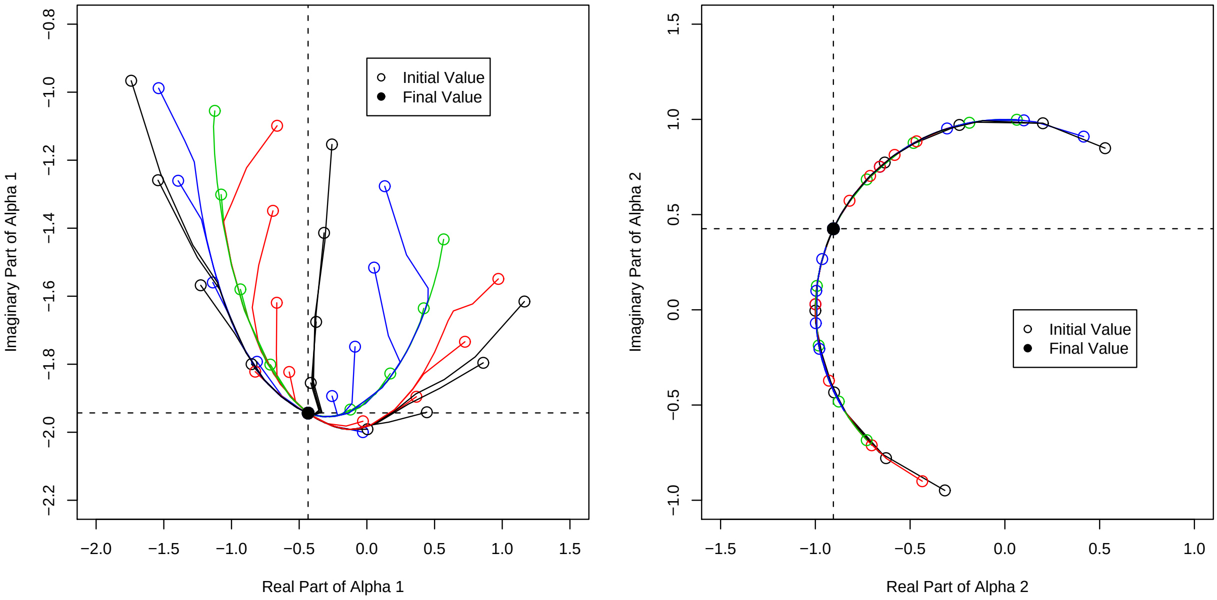

Figure 4: *Trajectory of the AR estimates* $\boldsymbol{\alpha}_m := [\alpha_{1m}, \alpha_{2m}]^T$ *which correspond to the frequency estimates in the right panel of Fig. 3.*

These results indicate that while accommodating poor initial values requires a smaller shrinkage parameter, producing an accurate final estimate demands a larger shrinkage parameter. In practice, the conflicting requirements on the shrinkage parameter can be satisfied by employing a sequence of increasing values $0 < \eta_1 < \eta_2 < \cdots < \eta_q < 1$, where the estimate produced by NIFA with $\eta = \eta_l$ after m_l iterations can be used as the initial value for NIFA with $\eta = \eta_{l+1}$ $(l = 1, \ldots, q-1)$. In this way, one can start with a poor initial value and still ends up with an accurate final estimate.

Table 1: Choice of Shrinkage Parameter

n	25	50	100	200	400
η_1	0.830	0.900	0.960	0.970	0.980
η_2	0.835	0.975	0.985	0.995	0.998

With two values of η given by Table 1 for each sample size, Fig. 5 shows the mean-square error (MSE)

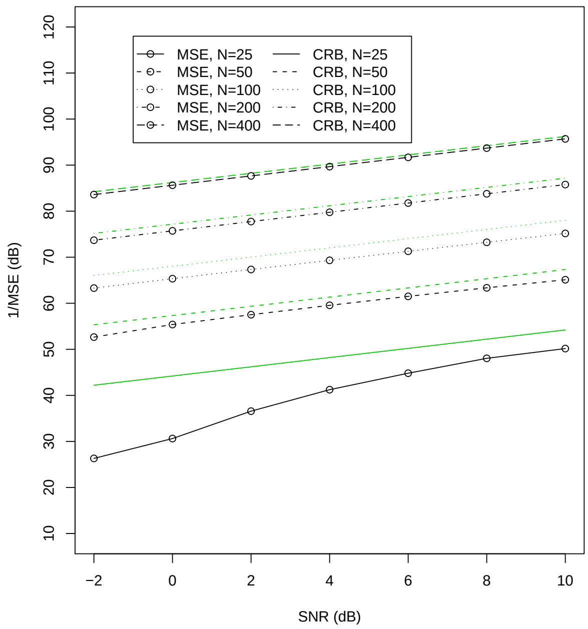

Figure 5: *Reciprocal MSE of the estimator for f_1 by NIFA in (9). Lines without circle represent the CRLB under the assumption of complex Gaussian white noise. Results are based on 1,000 Monte Carlo runs. The true frequencies are $f_1 = 0.2$ and $f_2 = 0.23$ (the phases are $\phi_1 = 0$ and $\phi_2 = \pi/2$). The simulated noise is complex Gaussian.*

of the estimator $\hat{f}_1$ by NIFA and the corresponding CRLB for various sample sizes and SNRs. (Similar results, not shown, are obtained for $\hat{f}_2$.) In all cases, the LS estimator of **a** from the unfiltered data, also known as Prony's estimator, is employed to initialize NIFA. The algorithm stops after 15 iterations: 6 with η_1 followed by 9 with η_2.

As can be seen from Fig. 5, the accuracy of the NIFA estimator is very close to the CRLB, except for the case of $n = 25$ where the frequency separation is less than $1/n$. Since Prony's estimator is bias and inconsistent (Li and Kedem 1994), the results in Fig. 5 suggest that NIFA with a proper choice of η is able to accommodate poor initial values of accuracy $\mathcal{O}(1)$ and still manages to converge to a final estimate which is nearly optimal.

The benefit of the bias-correction (BC) technique in (9) as compared to (7) can be appreciated by examining the results displayed in Figs. 6 and 7. Fig. 6 shows that whereas both algorithms converge from all

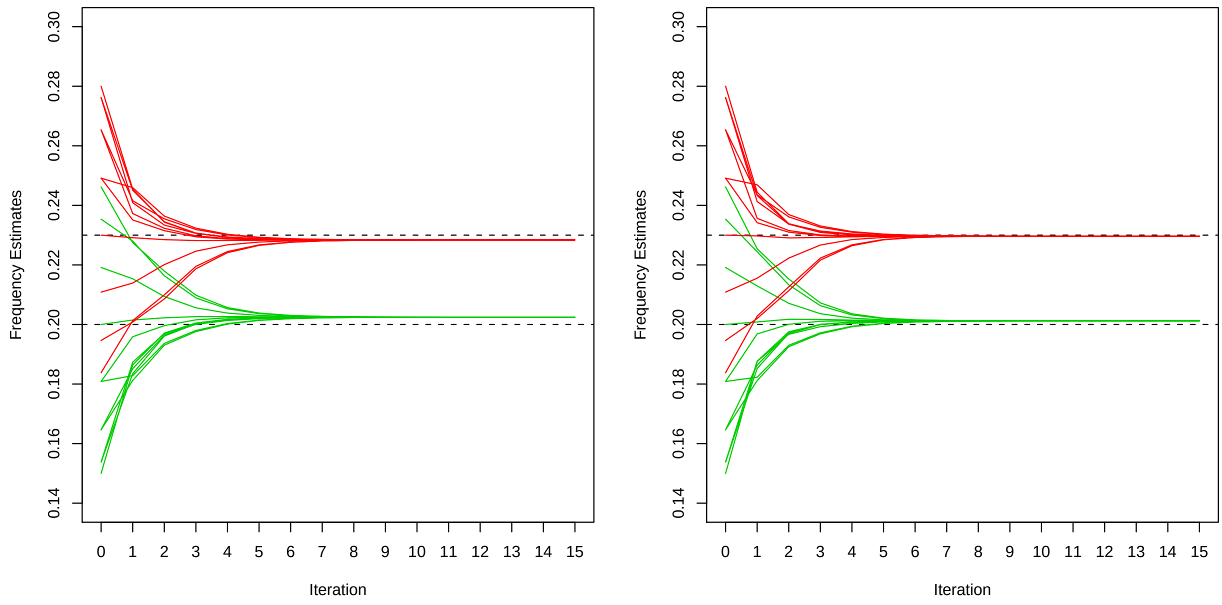

Figure 6: *Frequency estimates produced by NIFA from different initial values with and without using the BC technique. Left, NIFA without BC in (9); right, NIFA with BC in (7), both with* η *=0.975 and for the same time series of length* $n = 50$ *and SNR* $=$ *10 dB. Green lines represent estimates for* f_1 *and red lines for* f_2*. Dashed lines represent the true frequencies.*

initial values, the final estimates by NIFA with BC have a smaller bias than NIFA without BC. Fig. 7 further shows that the bias of NIFA without BC is more than 15 dB larger than that of NIFA with BC. Although the variance is approximately the same for both and is very close to the CRLB for all SNRs, the bias of NIFA without BC dominates the variance and becomes greater than the CRLB as the SNR increases beyond 2 dB. As a result, at SNR = 10 dB, the MSE, which equals the sum of variance and squared bias, of NIFA without BC exceeds the CRLB by approximately 12 dB as compared to 2 dB for NIFA with BC.

3. MAIN THEOREMS

Eq. (4) defines a one-to-one mapping from $\boldsymbol{\omega} := [\omega_1, \omega_2]^T$ to $\mathbf{a} = [a_1, a_2]^T$ which will be denoted by $\boldsymbol{\omega} \mapsto \boldsymbol{\phi}(\boldsymbol{\omega}) := \mathbf{a}$. Let Λ denote the set of $\boldsymbol{\lambda} := [\lambda_1, \lambda_2]^T$ with $-\pi < \lambda_1 < \lambda_2 \leq \pi$ and let $\mathcal{A}$ denote the set of

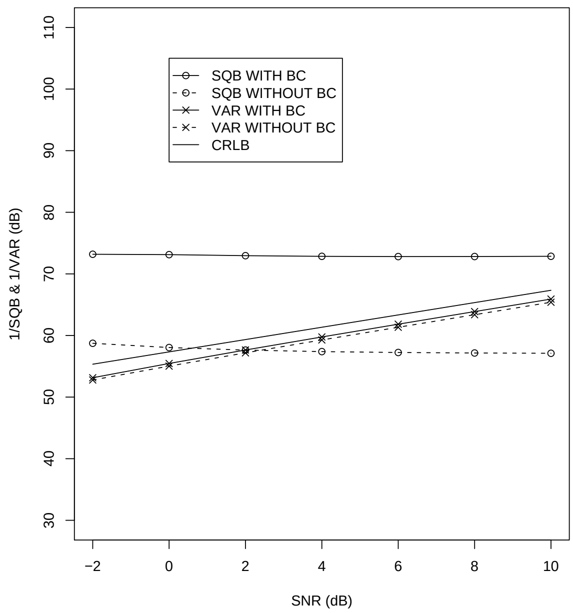

Figure 7: *Reciprocal squared bias (SQB) and variance (VAR) of the estimates for f_1 produced with and without using the BC technique. Solid lines with circles or crosses, NIFA with BC in (9); dashed lines with circles or crosses, NIFA without BC in (7); lines with circles, SQB; lines with crosses, VAR; solid line without circles or crosses, CRLB. True frequencies are employed as initial values. Shrinkage parameters in Table 1 for $n = 50$ are used. Results are based on 1,000 Monte Carlo runs, each of sample size $n = 50$. True frequencies and phases are the same as in Fig. 5.*

$\boldsymbol{\alpha} := [\alpha_1, \alpha_2]^T$ such that $\boldsymbol{\alpha} = \boldsymbol{\phi}(\boldsymbol{\lambda})$ for some $\boldsymbol{\lambda} \in \Lambda$, i.e., $\mathcal{A} := \boldsymbol{\phi}(\Lambda)$. Let $\mathcal{A}_\delta$ denote a closed subset of $\mathcal{A}$ such that $\mathcal{A}_\delta := \{\boldsymbol{\alpha} \in \mathcal{A}\colon \|\boldsymbol{\alpha} - \mathbf{a}\| \le \kappa\delta^\varepsilon\}$, where $\kappa > 0$ and $\varepsilon \in (1, \frac{3}{2})$ are constants and $\delta := 1 - \eta \in (0,1)$ depends on n such that $\delta \to 0$ as $n \to \infty$.

Theorem 1 *Let $\mathcal{A}_\delta$ be the neighborhood of $\mathbf{a}$ defined above and assume that $n(1-\delta)^n = \mathcal{O}(1)$ and $n\delta^\varepsilon \to \infty$ as $n \to \infty$. Then, with probability tending to unity, the mapping $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b})$ is contractive in $\mathcal{A}_\delta$, with a contraction factor of the form $\mathcal{O}_P(\delta^{\varepsilon-1})$, and therefore has a unique fixed point $\hat{\mathbf{a}} \in \mathcal{A}_\delta$. Furthermore, with probability tending to unity, the sequence $\boldsymbol{\alpha}_m := \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}_{m-1}) - \mathbf{b})$ $(m = 1, 2, \ldots)$ converges to $\hat{\mathbf{a}}$ as $m \to \infty$ for any initial value $\boldsymbol{\alpha}_0 \in \mathcal{A}_\delta$.*

The required accuracy for $\boldsymbol{\alpha}_0$ in Theorem 1 can be expressed as $\mathcal{O}(\delta^{\varepsilon})$. Since $n\delta^{\varepsilon} \to \infty$, it means that the initial accuracy can be lower than $\mathcal{O}(n^{-1})$. This is in contrast with the Gaussian maximum likelihood method which requires that the initial accuracy be higher than $\mathcal{O}(n^{-1})$.

Note that the requirement $n(1-\delta)^n = \mathcal{O}(1)$ implies $n\delta \to \infty$, so δ cannot approach zero faster than n^{-1}. The requirement $n\delta^{\varepsilon} \to \infty$ further asserts that δ should not approach zero faster than $n^{-1/\varepsilon}$. Both requirements explain why in our numerical examples η cannot be too close to unity in order to avoid converging to spurious fixed points when initial values are poor.

Theorem 2 *Assume that the conditions in Theorem 1 are satisfied. In addition, assume that there exists an initial estimator* $\hat{\boldsymbol{\omega}}_0$ *such that* $\|\hat{\boldsymbol{\omega}}_0 - \boldsymbol{\omega}\| = o_P(\delta^{\varepsilon-1})$. *Let* $\hat{\mathbf{b}}$ *be defined in the same way as* $\mathbf{b}$ *by* (8) *using* $\hat{\boldsymbol{\omega}}_0$ *instead of* $\boldsymbol{\omega}$. *Then, the assertions in Theorem 1 remain valid for the mapping* $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \hat{\mathbf{b}})$ *and the sequence* $\boldsymbol{\alpha}_m := \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}_{m-1}) - \hat{\mathbf{b}})$ $(m = 1, 2, \ldots)$.

The condition $\|\hat{\boldsymbol{\omega}}_0 - \boldsymbol{\omega}\| = o_P(\delta^{\varepsilon-1})$ can be easily satisfied by a wide range of initial estimators readily available in practice. For example, it is satisfied by any initial estimator $\hat{\boldsymbol{\omega}}_0$ such that $\|\hat{\boldsymbol{\omega}}_0 - \boldsymbol{\omega}\| = \mathcal{O}_P(\delta^{\gamma})$ for some $\gamma > \varepsilon - 1$. In particular, it is satisfied by all $\sqrt{n}$-consistent estimators such as those produced by Pisarenko's method (Pisarenko 1973; Sakai 1984) or the singular-value-decomposition-based methods (Stoica and Söderström 1991).

Theorem 3 *Assume that the conditions in Theorem 1 are satisfied. Then the assertions in Theorem 1 remain valid for the mapping* $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha}))$ *and the sequence* $\{\boldsymbol{\alpha}_m\}$ *defined by* (9).

Theorem 4 *Assume that the conditions in Theorem 1 are satisfied. Let* $\hat{\boldsymbol{\omega}}$ *be the frequency estimator corresponding to the fixed point* $\hat{\mathbf{a}}$ *of the mapping* $\boldsymbol{\alpha} \mapsto \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha}))$ *in* $\mathcal{A}_\delta$. *Then, for any constant* $\beta \le 3/2$ *such that* $n^{-1}\delta^{-\beta} = \mathcal{O}(1)$ *as* $n \to \infty$, $\delta^{-\beta}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega})$ *is uniformly tight, i.e.,* $\delta^{-\beta}(\hat{\boldsymbol{\omega}} - \boldsymbol{\omega}) = \mathcal{O}_P(1)$. *In other words,* $\hat{\boldsymbol{\omega}}$ *is at least* $\delta^{-\beta}$*-consistent for estimating* $\boldsymbol{\omega}$.

If η is chosen such that $\delta = 1-\eta = \mathcal{O}(n^{-\nu})$ for some $0 < \nu < \varepsilon^{-1} < 1$, then $n\delta^{\varepsilon} = \mathcal{O}(n^{1-\nu\varepsilon}) \to \infty$, so the conditions in Theorem 1 are satisfied. With this choice, according to Theorem 4, the NIFA estimator $\hat{\boldsymbol{\omega}}$ is at least $n^{\nu\beta}$-consistent for any $\beta \le \min(\nu^{-1}, 3/2)$, as it satisfies $n^{-1}\delta^{-\beta} = \mathcal{O}(n^{-1+\nu\beta}) = \mathcal{O}(1)$. The required accuracy of initial values can be expressed as $\mathcal{O}(n^{-\nu\varepsilon})$.

Now consider the special case where $\nu = 2/3$ and $\beta = \nu^{-1} = 3/2$. In this case, Theorem 4 implies that $\hat{\boldsymbol{\omega}}$ is at least n-consistent. By setting $\varepsilon = 1^{+}$, the initial requirement becomes nearly $\mathcal{O}(n^{-2/3})$. This means that it suffices to use a slightly better than $n^{2/3}$-consistent estimator as the initial value in order to obtain the n-consistent final estimator $\hat{\boldsymbol{\omega}}$. Such initial values can be produced by NIFA itself, with the choice of $\nu = (4/9)^{+}$. Indeed, with this choice, Theorem 4 guarantees a better than $n^{2/3}$ rate of consistency by taking $\beta = 3/2$ so that $\nu\beta = (2/3)^{+}$. To obtain this estimator, the initial values are required to be slightly more accurate than $\mathcal{O}(n^{-4/9})$, which can be satisfied by all $\sqrt{n}$-consistent estimators.

As can be seen, by applying NIFA twice, once with a smaller η and once with a large η, one is able to improve the accuracy of the frequency estimator from $\mathcal{O}(n^{-4/9})$ to $\mathcal{O}(n^{-1})$ or better. The convergence of NIFA in both cases are guaranteed by Theorem 4.

Note that one can determine the number of iterations needed to achieve the final accuracy ensured by Theorem 4 from the contraction factor, denoted by θ_n, of the mapping. Indeed, since after m iterations one can write $\|\boldsymbol{\alpha}_m - \hat{\mathbf{a}}\| \le \theta_n^m \|\boldsymbol{\alpha}_0 - \hat{\mathbf{a}}\|$, and since $\boldsymbol{\alpha}_0 - \hat{\mathbf{a}} = \mathcal{O}_P(\delta^{\varepsilon})$ and $\theta_n = \mathcal{O}_P(\delta^{\varepsilon-1})$, it follows that

$$\|\boldsymbol{\alpha}_m - \hat{\mathbf{a}}\| = \mathcal{O}_P\big(\delta^{m(\varepsilon-1)+\varepsilon}\big).$$

Therefore, in order to make this quantity the same order of magnitude as the final estimation error of the form $\mathcal{O}_P(\delta^{\beta})$, it suffices that $m \ge (\beta-1)/(\varepsilon-1)$. With $\beta = 3/2$ in particular, it becomes

$$m \ge \tfrac{1}{2}(\varepsilon-1)^{-1}.$$

This expression suggests, with no surprise, that the number of iterations depends on the accuracy of initial values: More iterations are needed when the initial values are poor so that ε has to be near unity, but for good initial values, ε can be made near $3/2$, so a few iterations are enough to bring them sufficiently close to the final estimator.

Finally we note that Theorem 4 does not exclude the possibility of a higher rate of consistency, which is what "at least" means. In fact, by subtracting a more accurate bias term, which now depends on the amplitude as well as the frequency of the sinusoids and needs to be estimated iteratively as before, one can remove the condition $n^{-1}\delta^{-\beta} = \mathcal{O}(1)$, so that ν in the above discussion can take on values such as 1^-. This, coupled with $\beta = 3/2$, gives rise to a rate of consistency which is arbitrarily close to the optimal rate $n^{3/2}$.

4. EXAMPLE AND DISCUSSION

4.1 Real Data Example

The dataset in this example is the "Sea Clutter + Target Data File" #283, taken from a large database of high-resolution radar measurements collected in November 1993 with the McMaster IPIX Radar overlooking the Atlantic Ocean from a cliff-top in Dartmouth, Nova Scotia, Canada (see http://soma.ece.mcmaster.ca/ipix). The dataset is about 2 minutes long and has a weak target in one of its range bins. The target is a spherical block of styrofoam wrapped with wire mesh. It has a diameter of one meter. The objective of this exercise is to estimate the Doppler frequency of the target from the radar data. For a target of p scatterers moving at a constant speed within the observation time window, the radar data (after preprocessing) can be modeled by (1), where the real and imaginary parts represent the in-phase and quadrature components and the noise is attributed to the backscatters from the ocean surface (sea clutters) (Lewis, Kretschmer, and Shelton 1986).

Fig. 8 shows the frequency estimation results for two 128-point (128 milliseconds long) segments of the radar data from the 10th range bin (2685 meters). As can be seen, for Segment 1 (left panel) where the signal is strong, the frequency estimates coincide and appear at the spectral peak of the target; for Segment 2 (right panel) where the signal is weak, the frequency estimates split, with one locked at the spectral peak of the

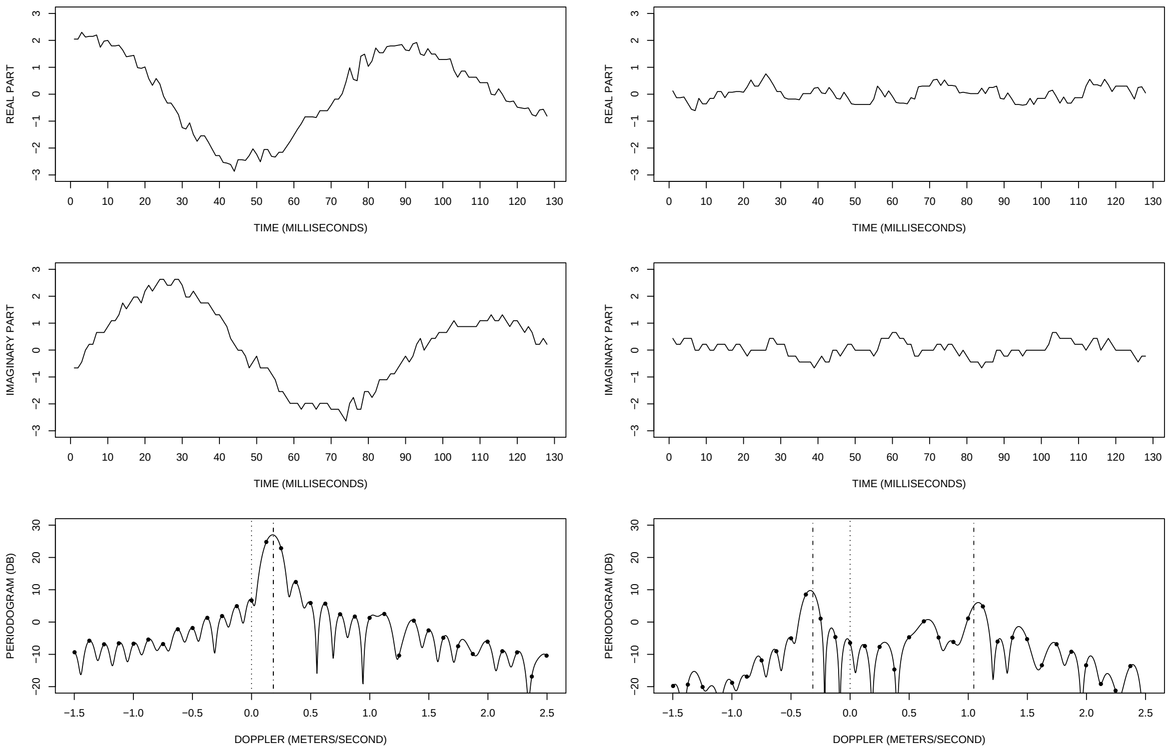

Figure 8: *Frequency Estimates and Periodograms for Two 128-Point Segments of Radar Data. Left panel, Segment 1 (strong signal); Right panel, Segment 2 (weak signal). Top, real part of the data; Middle, imaginary part of the data; Bottom, frequency estimates superimposed on the periodogram: dots, periodogram ordinates; solid line, interpolated periodogram; dotted line, initial values of NIFA; dash-dotted line, final frequency estimates produced by NIFA after 20 iterations (5 with* $\eta = 0.9$ *followed by 15 with* $\eta = 0.95$*).*

target (dash-dotted line on the left) and the other locked at the largest spectral peak of the noise (dash-dotted line on the right). According to the estimation, in Segment 1 the target is moving towards the radar at an estimated speed of 0.17 meters per second, whereas in Segment 2 the target is moving away from the radar at an estimated speed of 0.31 meters per second.

4.2 Automatic Selection of Shrinkage Parameter

In practice, the shrinkage parameter η must be fully specified in advance. Although one may treat η as a tuning parameter and experiment with different values until a satisfactory result is obtained, one may also wish to automate this process based solely on the data. A simple way of doing so is to regard the frequency estimates from NIFA as functions of η, regress the observations on the complex sinusoids with the estimated frequencies in place of the true frequencies, and then minimize the resulting mean-squared error (MSE) as a univariate function of η.

Fig. 9 shows 100 independent realizations of the regression MSE as a function of η where the frequency estimates are obtained from NIFA with Prony's estimator as the initial value. As can be seen, the MSE starts from large values when η is far from unity, decreases as η moves towards unity, reaches its minimum when η enters the interval between 0.98 and 0.99 (approximately), and, in some cases, begins to rise when η further increases. The minimum values are very close to the ideal MSE values (MSE0) obtained using the true frequencies, resulting in a near zero excess MSE defined as the difference between the MSE curve and the MSE0 (which is a constant). Note that the best values of η appear to lie within a relatively large interval, indicating that the frequency estimates are not overly sensitive to the choice of η once it is in the correct neighborhood.

4.3 Generalization

The proposed method can be extended easily to the general case of $p \geq 1$. Indeed, for $p \geq 1$, the filtered data $\{y_t(\boldsymbol{\alpha})\}$ can be obtained from

$$y_t(\boldsymbol{\alpha}) := -\sum_{k=1}^{p} \alpha_k \eta^k y_{t-k}(\boldsymbol{\alpha}) + y_t \quad (t = 1, \ldots, n) \tag{10}$$

with $y_{1-p}(\boldsymbol{\alpha}) = \cdots = y_0(\boldsymbol{\alpha}) = 0$ and $\boldsymbol{\alpha} := [\alpha_1, \ldots, \alpha_p]^T$. Given $\{y_t(\boldsymbol{\alpha})\}$, the AR parameter $\mathbf{a} := [a_1, \ldots, a_p]^T$ is estimated by the LS method that seeks $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ to minimize $\|\mathbf{y}(\boldsymbol{\alpha}) + \mathbf{Y}(\boldsymbol{\alpha})\,\mathbf{D}\,\hat{\mathbf{a}}(\alpha)\|^2$, where

$$\mathbf{Y}(\boldsymbol{\alpha}) := \begin{bmatrix} y_p(\boldsymbol{\alpha}) & \cdots & y_1(\boldsymbol{\alpha}) \\ \vdots & & \vdots \\ y_{n-1}(\boldsymbol{\alpha}) & \cdots & y_{n-p}(\boldsymbol{\alpha}) \end{bmatrix}, \quad \mathbf{y}(\boldsymbol{\alpha}) := \begin{bmatrix} y_{p+1}(\boldsymbol{\alpha}) \\ \vdots \\ y_n(\boldsymbol{\alpha}) \end{bmatrix}, \quad \mathbf{D} := \mathrm{diag}(\eta, \ldots, \eta^p).$$

This gives rise to the LS mapping of the form (6).

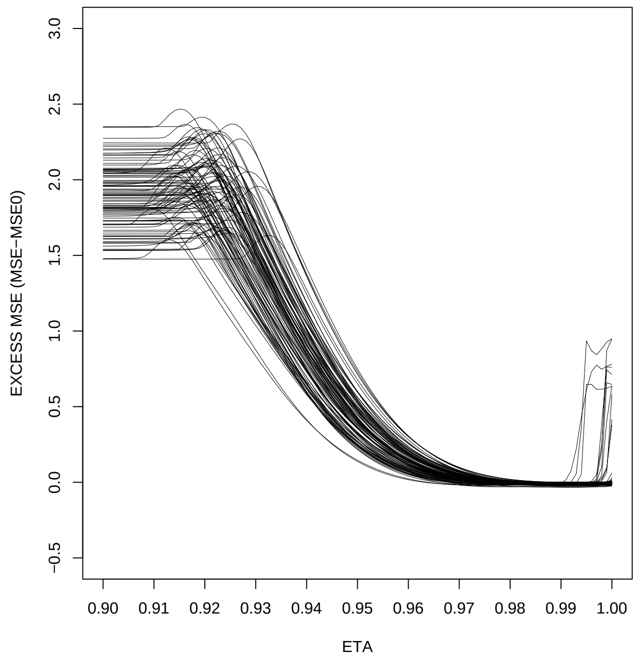

Figure 9: *Mean-Squared Error of Complex Sinusoidal Regression with Estimated Frequencies as a Function of η From 100 Independent Monte Carlo Trials. For each trial, the MSE of complex sinusoidal regression using the true frequencies (denoted by MSE0) is subtracted for easy comparison. Prony's estimator is used to initialize NIFA. Model parameters are the same as in Fig. 3.*

To analyze this mapping, we first need to express $\{y_t(\boldsymbol{\alpha})\}$ in terms of $\{y_t\}$. To this end, let $\boldsymbol{\lambda} := [\lambda_1, \cdots, \lambda_p]^T := \boldsymbol{\phi}^{-1}(\boldsymbol{\alpha})$ and $\boldsymbol{\zeta} := [\zeta_1, \cdots, \zeta_p]^T := [e^{i\lambda_1}, \cdots, e^{i\lambda_p}]^T$, where $\mathbf{a} = \boldsymbol{\phi}(\boldsymbol{\omega})$ is the one-to-one mapping defined by (2). Then, the transfer function of the filter in (10) can be expressed as

$$H(z) := \sum_{j=0}^{\infty} h_j(\boldsymbol{\lambda}) z^{-j} = \left(1 + \sum_{k=1}^{p} \alpha_k \eta^k z^{-k}\right)^{-1} = \left\{\prod_{k=1}^{p} \left(1 - \zeta_k \eta z^{-1}\right)\right\}^{-1}.$$

where $h_0(\boldsymbol{\lambda}) = 1$ and for $j \geq 1$, by the Cauchy integral theorem and the residue theorem,

$$\begin{aligned} h_j(\boldsymbol{\lambda}) &= \frac{1}{2\pi i} \oint_{|z|=1} H(z) z^{j-1} dz = \sum_{k=1}^{p} \mathrm{Res}\{H(z) z^{j-1}\}_{z=\zeta_k \eta} \\ &= \sum_{k=1}^{p} \lim_{z \to \zeta_k \eta} (z - \zeta_k \eta) H(z) z^{j-1} = \eta^j \sum_{k=1}^{p} \frac{\zeta_k^{j+p-1}}{\prod_{k \neq k'} (\zeta_k - \zeta_{k'})}, \end{aligned}$$

where $\prod_{k\neq k'}(\zeta_k - \zeta_{k'})$ is defined as 1 when $p = 1$. Given this expression, (10) can be rewritten as

$$y_t(\boldsymbol{\alpha}) = \sum_{j=0}^{t-1} h_j(\boldsymbol{\lambda})\, y_{t-j}. \tag{11}$$

In the expression of $h_j(\boldsymbol{\lambda})$, the decaying factor η^j does not depend on p. This suggests that the rate of convergence of the NIFA frequency estimator in the case of $p \geq 1$ would be the same as in the case of $p = 2$.

Moreover, some algebra shows that the general entry of the Hermitian matrix $\mathbf{Y}^H(\boldsymbol{\alpha})\mathbf{Y}(\boldsymbol{\alpha})$ is given by $\hat{r}_{jk}(\boldsymbol{\alpha}) := \sum_{t=p-j+1}^{n-j} y_t^*(\boldsymbol{\alpha})\, y_{t+j-k}(\boldsymbol{\alpha})$ $(1 \leq k \leq j \leq p)$. With the explicit expression (11), we can apply the same techniques developed for the case of $p = 2$ to evaluate the exact order of magnitudes of $\hat{r}_{jk}(\boldsymbol{\alpha})$ as well as other terms. For $p = 1$, this leads to

$$\hat{a}_1(\alpha) = -\frac{\sum_{t=1}^{n-1} y_t^*(\alpha)\, y_{t+1}(\alpha)}{\eta \sum_{t=1}^{n-1} |y_t(\alpha)|^2}, \quad b_1 = -\frac{(1-\eta)e^{i\omega_1}}{\eta},$$

where

$$y_t(\alpha) = -\alpha\eta\, y_{t-1}(\alpha) + y_t = \sum_{j=1}^{t-1} h_j(\lambda)\, y_{t-j},$$

with $y_0(\alpha) := 0$, $h_j(\lambda) := (\eta\zeta)^j$, and $\zeta := e^{i\lambda} := -\alpha$.

A major mathematical challenge in the general case of $p > 2$ is the inversion of the p-by-p matrix $\mathbf{Y}^H(\boldsymbol{\alpha})\mathbf{Y}(\boldsymbol{\alpha})$, which makes direct evaluation of the elements in $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ computationally very complicated. Although the lack of simple inversion formula currently prevents us from identifying the general leading bias term, the basic ideas and mathematical arguments used in our analysis for the case of $p = 2$ would carry over to the general case of $p > 2$.

In this paper, the bias in the LS mapping is corrected by subtracting the leading term of the bias in an asymptotic expansion. An alternative method of bias correction was proposed by Li and Kedem (1994). Instead of subtracting the bias, which depends on the unknown frequencies, the method of Li and Kedem (1994) relies on reparameterization of the filter so that the autocovariances of the filtered noise satisfy a set of equations called the parameterization property. The method was successfully applied to an AR filter for estimating the frequencies of real sinusoids. The earlier work of Li and Song (2001; 2003) and Song and Li (1997; 2000) was devoted to analyzing the resulting estimators. Unfortunately, for complex sinusoids and the AR filter in (10), this method does not lead to a useful solution because of a fundamental difference between real and complex sinusoids, which motivates us to use the bias subtraction method discussed in the present article.

More specifically, let $\{h_j(\boldsymbol{\alpha})\}_{j=0}^{\infty}$ be a linear filter such that $\sum|h_j(\boldsymbol{\alpha})| < \infty$. Applying this filter to $\{y_t\}$ in (1) yields the filtered process

$$y_t(\boldsymbol{\alpha}) := \sum_{j=0}^{\infty} h_j(\boldsymbol{\alpha})\, y_{t-j} = x_t(\boldsymbol{\alpha}) + \varepsilon_t(\boldsymbol{\alpha}),$$

where $x_t(\boldsymbol{\alpha})$ and $\varepsilon_t(\boldsymbol{\alpha})$ are the filtered signal and the filtered noise, respectively. It is not difficult to show that $x_t(\boldsymbol{\alpha})$ remains a sum of p complex sinusoids with the same frequencies as those of x_t and hence satisfies the same AR equation as x_t does. Let $\mathbf{R}_\varepsilon(\boldsymbol{\alpha})$ denote the p-by-p autocovariance matrix of $\{\varepsilon_t(\boldsymbol{\alpha})\}$ with the (j,k)th entry $r^{\varepsilon}_{j-k}(\boldsymbol{\alpha}) := E\{\varepsilon_{t+j-k}(\boldsymbol{\alpha})\varepsilon_t^*(\boldsymbol{\alpha})\}$ $(j,k = 0,1,\ldots,p-1)$, and let $\mathbf{r}_\varepsilon(\boldsymbol{\alpha}) := [r^{\varepsilon}_{-1}(\boldsymbol{\alpha}),\ldots,r^{\varepsilon}_{-p+1}(\boldsymbol{\alpha})]^T$. Then, the parameterization property can be written as

$$\boldsymbol{\alpha} = -\mathbf{R}_\varepsilon^{-1}(\boldsymbol{\alpha})\,\mathbf{r}_\varepsilon(\boldsymbol{\alpha}). \tag{12}$$

As shown by Li and Kedem (1994), any filter satisfying (12) will lead to an asymptotically unbiased frequency estimator as the fixed point of the LS mapping $\{\mathbf{Y}^H(\boldsymbol{\alpha})\,\mathbf{Y}(\boldsymbol{\alpha})\}^{-1}\mathbf{Y}^H(\boldsymbol{\alpha})\,\mathbf{y}(\boldsymbol{\alpha})$. Unfortunately, the AR filter (10) cannot be reparameterized to satisfy (12). To see this, let $\theta_k := \theta_k(\boldsymbol{\alpha},\eta)$ be any reparameterization so that (10) becomes $y_t(\boldsymbol{\alpha}) = -\sum_{k=1}^{p}\theta_k y_{t-k}(\boldsymbol{\alpha}) + y_t$. Then, it is easy to see that the filtered noise $\{\varepsilon_t(\boldsymbol{\alpha})\}$ with the new parameters is an AR(p) process with $\{\theta_k\}$ as the AR coefficients. For this AR process, the Yule-Walker equations take the form $\mathbf{R}_\varepsilon(\boldsymbol{\alpha})\,\boldsymbol{\theta} = -\mathbf{r}_\varepsilon(\boldsymbol{\alpha})$, where $\boldsymbol{\theta} := \boldsymbol{\theta}(\boldsymbol{\alpha},\eta) := [\theta_1(\boldsymbol{\alpha},\eta),\ldots,\theta_p(\boldsymbol{\alpha},\eta)]^T$. Therefore, the only way to make this filter satisfy (12) is to choose $\boldsymbol{\theta} = \boldsymbol{\alpha}$, which corresponds to the AR filter (10) with shrinkage parameter $\eta = 1$. This means that for $\eta < 1$, the AR filter (10) cannot be reparameterized to satisfy (12).

A closer look reveals that the reason why the reparameterization method works for real sinusoids is that $x_t := \sum_{k=1}^{p} A_k \cos(\omega_k t + \phi_k)$ satisfies a *symmetric* AR equation $\sum_{j=0}^{2p} a_j x_{t-j} = 0$, where $a_{2p-j} = a_j$ for $j = 0, 1, \ldots, p-1$ and $a_0 = 1$. As a result, the AR filter can be constrained by the symmetry. On the contrast, for complex sinusoids, the corresponding AR equation is not symmetric, so the AR filter cannot be constrained. A real sinusoid equals the sum of two complex sinusoids which are conjugate pairs. Therefore, it is not surprising that the complex sinusoid model is more general than the real sinusoid model in the sense that estimation methods developed under the complex sinusoid model are directly applicable to the real sinusoid model, but not vise versa.

5. CONCLUDING REMARKS

In this article, we have presented a new iterative filtering algorithm (NIFA) for joint estimation of the frequencies of multiple complex sinusoids from noisy observations. The algorithm is based on the idea of repeating the cycle of least-squares (LS) estimation and autoregressive (AR) filtering to form a fixed-point iteration. The AR filter is endowed with a bandwidth shrinkage parameter which is shown to control not only the accuracy of the final frequency estimates but also the accuracy required for the initial values. With a proper choice of the shrinkage parameter, which can be an increasing sequence, the algorithm achieves nearly global convergence with satisfactory estimation accuracy, and thereby provides a more practical alternative to the Gaussian maximum likelihood method which has stringent requirements on initial values. In our statistical analysis, it is shown that the mapping formed by the composition of LS estimation and AR filtering contains a bias, for which an explicit asymptotic expression is derived in the case of one and two sinusoids, that can be corrected to produce more accurate frequency estimates. It is shown that the bias-corrected iteration converges to the desired fixed-point which forms a consistent frequency estimator. Derivation of a closed-form expression for the bias in the general case of more than two sinusoids remains an open problem for future research.

APPENDIX I: PROOFS

For any $\boldsymbol{\alpha} := [\alpha_1, \alpha_2]^T \in \mathcal{A}_\delta$, let $\boldsymbol{\lambda} := [\lambda_1, \lambda_2]^T := \boldsymbol{\phi}^{-1}(\boldsymbol{\alpha})$ and $\boldsymbol{\zeta} := [\zeta_1, \zeta_2]^T := [e^{i\lambda_1}, e^{i\lambda_2}]^T$. Because $\boldsymbol{\phi}^{-1}(\boldsymbol{\alpha})$ is continuously differentiable in a neighborhood of $\mathbf{a}$, and $\boldsymbol{\phi}(\boldsymbol{\lambda})$ in a neighborhood of $\boldsymbol{\omega}$, there exist constants $c_1 > 0$ and $c_2 > 0$ such that $\boldsymbol{\alpha} \in \mathcal{A}_\delta$ implies $\|\boldsymbol{\lambda} - \boldsymbol{\omega}\| \leq c_1 \delta^\varepsilon$ and $\|\boldsymbol{\lambda} - \boldsymbol{\omega}\| \leq c_2 \delta^\varepsilon$ implies $\boldsymbol{\alpha} \in \mathcal{A}_\delta$.

A key technique in the proof of Theorems 1–3 is to apply the well-known fixed-point theorem (Stoer and Bulirsch 2001) to the corresponding random mappings. For example, to establish Theorem 1, it suffices to show that with probability tending to unity as $n \to \infty$,

(a) $\|\boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b}) - \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}') - \mathbf{b})\| \leq \theta_n \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|$ for any $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathcal{A}_\delta$, where $\theta_n \in (0,1)$ is the contraction factor which may depend on n but not on $\boldsymbol{\alpha}$ or $\boldsymbol{\alpha}'$, and

(b) $\|\boldsymbol{\psi}(\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b}) - \mathbf{a}\| \leq (1 - \theta_n)\kappa\delta^\varepsilon$.

Since $\hat{\mathbf{a}}(\boldsymbol{\alpha}) = \mathcal{O}_P(1)$ uniformly in $\boldsymbol{\alpha} \in \mathcal{A}_\delta$, and since $\hat{\mathbf{a}}(\boldsymbol{\alpha})$ has distinct roots with probability tending to unity as $n \to \infty$, which can be proved by using (13), (14), together with the fact that $\boldsymbol{\alpha} \in \mathcal{A}_\delta$ implies $\|\boldsymbol{\lambda} - \boldsymbol{\omega}\| = \mathcal{O}(\delta^\varepsilon)$, it follows from Lemma 2(c) that there exists $\kappa_n = \mathcal{O}_P(1)$ such that

$$\|\boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b}) - \boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}') - \mathbf{b})\| \leq \kappa_n \|\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \hat{\mathbf{a}}(\boldsymbol{\alpha}')\|.$$

Therefore, one can establish (a) by proving

$$\|\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \hat{\mathbf{a}}(\boldsymbol{\alpha}')\| \leq c_n \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| \tag{13}$$

for some $c_n = \mathcal{o}_P(1)$, because it leads to (a) with $\theta_n := c_n \kappa_n = \mathcal{O}_P(c_n) = \mathcal{o}_P(1)$, so that $\theta_n \in (0,1)$ with probability tending to unity as $n \to \infty$. Similarly, one can establish (b) by proving

$$\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b} - \mathbf{a} = \mathcal{o}_P(\delta^\varepsilon), \tag{14}$$

as it leads to $\|\boldsymbol{\psi}(\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b}) - \mathbf{a}\| \leq d_n := \kappa_n \|\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b} - \mathbf{a}\|$, where $\delta^{-\varepsilon} d_n = \mathcal{o}_P(1) \to 0$, so that $d_n \leq (1 - \theta_n)\kappa\delta^\varepsilon$ with probability tending to unity as $n \to \infty$.

Before proving the theorems, we first provide some useful propositions which constitute the basis of the proof. The proof of the propositions is outlined in the appendix.

Given $\boldsymbol{\alpha} \in \mathcal{A}_\delta$, it follows from (5) that

$$y_t(\boldsymbol{\alpha}) = \sum_{l=0}^{t-1} h_l(\boldsymbol{\lambda}) y_{t-l},$$

where

$$h_l(\boldsymbol{\lambda}) := \eta^l G_l(\boldsymbol{\lambda})/G_0(\boldsymbol{\lambda}), \quad G_l(\boldsymbol{\lambda}) := \zeta_2^{l+1} - \zeta_1^{l+1}.$$

For $k = 0, 1, 2$, let

$$V_{k,n}(\boldsymbol{\alpha}) \;:=\; |G_0(\boldsymbol{\lambda})|^2 \, \Phi_{k,n}(\boldsymbol{\alpha}), \tag{15}$$

$$U_{k,n}(\boldsymbol{\alpha}) \;:=\; G_0(\boldsymbol{\lambda}) \, \Psi_{k,n}(\boldsymbol{\alpha}), \tag{16}$$

where

$$\Phi_{k,n}(\boldsymbol{\alpha}) \;:=\; \sum_{t=1}^{n-2} y_{t,t+k}(\boldsymbol{\alpha}), \quad \Psi_{k,n}(\boldsymbol{\alpha}) \;:=\; \sum_{t=1}^{n-2} \tilde{y}_{t,t+k}(\boldsymbol{\alpha}), \tag{17}$$

$$y_{j,l}(\boldsymbol{\alpha}) \;:=\; y_j(\boldsymbol{\alpha}) y_l^*(\boldsymbol{\alpha}), \quad \tilde{y}_{j,l}(\boldsymbol{\alpha}) \;:=\; y_j(\boldsymbol{\alpha}) y_l^*. \tag{18}$$

The asterisk $*$ denotes complex conjugate. Let the lag-k sample autocorrelation coefficient of $\{y_t(\boldsymbol{\alpha})\}$ be defined as

$$\rho_{k,n}(\boldsymbol{\alpha}) := \Phi_{k,n}(\boldsymbol{\alpha})/\Phi_{0,n}(\boldsymbol{\alpha}) = V_{k,n}(\boldsymbol{\alpha})/V_{0,n}(\boldsymbol{\alpha}). \tag{19}$$

Moreover, define

$$\Omega_n(\boldsymbol{\alpha}) := \{1-|\rho_{1,n}(\boldsymbol{\alpha})|^2\}\,\Phi_{0,n}(\boldsymbol{\alpha}), \tag{20}$$

$$P_{1,n}(\boldsymbol{\alpha}) := \Psi^*_{1,n}(\boldsymbol{\alpha})-\rho_{1,n}(\boldsymbol{\alpha})\,\Psi^*_{2,n}(\boldsymbol{\alpha}), \tag{21}$$

$$P_{2,n}(\boldsymbol{\alpha}) := \Psi^*_{2,n}(\boldsymbol{\alpha})-\rho^*_{1,n}(\boldsymbol{\alpha})\,\Psi^*_{1,n}(\boldsymbol{\alpha}), \tag{22}$$

$$Q_{1,n}(\boldsymbol{\alpha}) := \alpha_2\eta\, y^*_{n-2,n-1}(\boldsymbol{\alpha})+\eta^{-1}\{y^*_{n-1,n}(\boldsymbol{\alpha})-y^*_{1,2}(\boldsymbol{\alpha})\}, \tag{23}$$

$$\begin{aligned} Q_{2,n}(\boldsymbol{\alpha}) &:= \rho^*_{1,n}(\boldsymbol{\alpha})\,\{\alpha_2 y_{n-2,n-1}(\boldsymbol{\alpha})+\eta^{-2}[y^*_{n-1,n}(\boldsymbol{\alpha})-y^*_{1,2}(\boldsymbol{\alpha})]\} \\ &\quad -\eta^{-2}\rho^*_{2,n}(\boldsymbol{\alpha})\,\{y_{n-1,n-1}(\boldsymbol{\alpha})-y_{1,1}(\boldsymbol{\alpha})\}, \end{aligned} \tag{24}$$

$$\begin{aligned} R_{1,n}(\boldsymbol{\alpha}) &:= -\frac{Q_{1,n}(\boldsymbol{\alpha})}{\Omega_n(\boldsymbol{\alpha})}-\frac{\alpha_1\{y_{n-1,n-1}(\boldsymbol{\alpha})-y_{1,1}(\boldsymbol{\alpha})\}}{\Omega_n(\boldsymbol{\alpha})+y_{n-1,n-1}(\boldsymbol{\alpha})-y_{1,1}(\boldsymbol{\alpha})} \\ &\quad +\frac{\{\eta^{-1}P_{1,n}(\boldsymbol{\alpha})+Q_{1,n}(\boldsymbol{\alpha})\}\,\{y_{n-1,n-1}(\boldsymbol{\alpha})-y_{1,1}(\boldsymbol{\alpha})\}}{\Omega_n(\boldsymbol{\alpha})\,\{\Omega_n(\boldsymbol{\alpha})+y_{n-1,n-1}(\boldsymbol{\alpha})-y_{1,1}(\boldsymbol{\alpha})\}}, \end{aligned} \tag{25}$$

$$\begin{aligned} R_{2,n}(\boldsymbol{\alpha}) &:= \frac{Q_{2,n}(\boldsymbol{\alpha})}{\Omega_n(\boldsymbol{\alpha})}-\frac{\alpha_2\{y_{n-1,n-1}(\boldsymbol{\alpha})-y_{1,1}(\boldsymbol{\alpha})\}}{\Omega_n(\boldsymbol{\alpha})+y_{n-1,n-1}(\boldsymbol{\alpha})-y_{1,1}(\boldsymbol{\alpha})} \\ &\quad +\frac{\{\eta^{-2}P_{2,n}(\boldsymbol{\alpha})-Q_{2,n}(\boldsymbol{\alpha})\}\,\{y_{n-1,n-1}(\boldsymbol{\alpha})-y_{1,1}(\boldsymbol{\alpha})\}}{\Omega_n(\boldsymbol{\alpha})\,\{\Omega_n(\boldsymbol{\alpha})+y_{n-1,n-1}(\boldsymbol{\alpha})-y_{1,1}(\boldsymbol{\alpha})\}}. \end{aligned} \tag{26}$$

With this notation, our propositions can be stated as follows.

Proposition 1 *Let* $P_{j,n}(\boldsymbol{\alpha})$ *and* $R_{j,n}(\boldsymbol{\alpha})$ $(j=1,2)$ *be defined by* (21), (22), (25), *and* (26). *Let* $\Omega_n(\boldsymbol{\alpha})$ *be defined by* (20). *Then,*

$$\hat{\mathbf{a}}(\boldsymbol{\alpha})=\boldsymbol{\alpha}-\Omega_n^{-1}(\boldsymbol{\alpha})\,\mathbf{D}^{-1}\mathbf{p}_n(\boldsymbol{\alpha})+\mathbf{r}_n(\boldsymbol{\alpha}),$$

where $\mathbf{p}_n(\boldsymbol{\alpha}):=[P_{1,n}(\boldsymbol{\alpha}),P_{2,n}(\boldsymbol{\alpha})]^T$ *and* $\mathbf{r}_n(\boldsymbol{\alpha}):=[R_{1,n}(\boldsymbol{\alpha}),R_{2,n}(\boldsymbol{\alpha})]^T$.

Proposition 2 *Let* $V_{k,n}$ $(k=0,1,2)$ *be defined by* (15). *If* $n(1-\delta)^n=\mathcal{O}(1)$ *and* $n\delta^{\varepsilon}\to\infty$ *as* $n\to\infty$, *then*

$$\begin{aligned} V_{k,n}(\boldsymbol{\alpha}) &= C_k\,(n-2)\delta^{-2}+\mathcal{O}_P(\delta^{-3})+\mathcal{O}_P(n\delta^{-3/2}) \\ &\quad +[\mathcal{O}_P(n\delta^{\varepsilon-4}),\,\mathcal{O}_P(n\delta^{\varepsilon-4})]\,(\boldsymbol{\lambda}-\boldsymbol{\omega}), \\ V_{k,n}(\boldsymbol{\alpha})-V_{k,n}(\boldsymbol{\alpha}') &= [\mathcal{O}_P(n\delta^{\varepsilon-4}),\,\mathcal{O}_P(n\delta^{\varepsilon-4})]\,(\boldsymbol{\lambda}-\boldsymbol{\lambda}'), \end{aligned}$$

uniformly in $\boldsymbol{\alpha},\boldsymbol{\alpha}' \in \mathcal{A}_\delta$, *where* $\boldsymbol{\alpha}' := \boldsymbol{\phi}(\boldsymbol{\lambda}')$, $C_k := \sum_{j=1}^2 C_{jk}$, *and* $C_{jk} := \beta_j^2 e^{-ik\omega_j}$ $(j=1,2)$.

Proposition 3 *Let* $U_{k,n}(\boldsymbol{\alpha})$ $(k=0,1,2)$ *be defined by* (16). *Then, under the conditions of Proposition 2,*

$$\begin{aligned} U_{k,n}(\boldsymbol{\alpha}) &= (C_{2,k-1}-C_{1,k-1})(n-2)\delta^{-1} + \mathcal{O}_P(n\delta^{-1/2}) + \mathcal{O}_P(\delta^{-2}) \\ &\quad + (n-2)\delta^{-2}\,[-iC_{1,k-1}+\mathcal{O}_P(\delta^{\varepsilon-1}),\, iC_{2,k-1}+\mathcal{O}_P(\delta^{\varepsilon-1})]\,(\boldsymbol{\lambda}-\boldsymbol{\omega}), \end{aligned}$$

$$\begin{aligned} & U_{k,n}(\boldsymbol{\alpha}) - U_{k,n}(\boldsymbol{\alpha}') \\ &= (n-2)\delta^{-2}\,[-iC_{1,k-1}+\mathcal{O}_P(\delta^{\varepsilon-1}),\, iC_{2,k-1}+\mathcal{O}_P(\delta^{\varepsilon-1})]\,(\boldsymbol{\lambda}-\boldsymbol{\lambda}'), \end{aligned}$$

uniformly in $\boldsymbol{\alpha},\boldsymbol{\alpha}' \in \mathcal{A}_\delta$.

Proposition 4 *Let* $\rho_{k,n}(\boldsymbol{\alpha})$ $(k=0,1,2)$ *be defined by* (19). *Then, under the conditions of Proposition 2,*

$$\rho_{k,n}(\boldsymbol{\alpha}) = C_k C_0^{-1}\{1+\mathcal{O}_P(n^{-1}\delta^{-1})+\mathcal{O}_P(\delta^{1/2})\} + [\mathcal{O}_P(\delta^{\varepsilon-2}),\, \mathcal{O}_P(\delta^{\varepsilon-2})]\,(\boldsymbol{\lambda}-\boldsymbol{\omega}),$$

$$\rho_{k,n}(\boldsymbol{\alpha}) - \rho_{k,n}(\boldsymbol{\alpha}') = [\mathcal{O}_P(\delta^{\varepsilon-2}),\, \mathcal{O}_P(\delta^{\varepsilon-2})]\,(\boldsymbol{\lambda}-\boldsymbol{\lambda}'),$$

uniformly in $\boldsymbol{\alpha},\boldsymbol{\alpha}' \in \mathcal{A}_\delta$.

Now let us begin with the proof of the theorems.

A. Proof of Theorem 1

Recall that Theorem 1 can be proved by establishing (13) and (14). Consider (14) first. It follows from Proposition 1 that

$$\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b} - \mathbf{a} = -\mathbf{b} - \mathbf{D}^{-1}\Omega_n^{-1}(\mathbf{a})\,\mathbf{p}_n(\mathbf{a}) + \mathbf{r}_n(\mathbf{a}). \tag{27}$$

Some algebra shows that

$$\Omega_n^{-1}(\mathbf{a})\,P_{1,n}(\mathbf{a}) = \frac{G_0(\boldsymbol{\omega})\,\{U_{1,n}^*(\mathbf{a}) - \rho_{1,n}(\mathbf{a})\,U_{2,n}^*(\mathbf{a})\}}{\{1-|\rho_{1,n}(\mathbf{a})|^2\}\,V_{0,n}(\mathbf{a})}.$$

From Proposition 3, we obtain

$$U^*_{k,n}(\mathbf{a}) = (C^*_{2,k-1} - C^*_{1,k-1})\,(n-2)\delta^{-1}\{1 + \mathcal{O}_P(n^{-1}\delta^{-1}) + \mathcal{O}_P(\delta^{1/2})\}.$$

This, coupled with Proposition 4, implies

$$\begin{aligned} & U^*_{1,n}(\mathbf{a}) - \rho_{1,n}(\mathbf{a})\,U^*_{2,n}(\mathbf{a}) \\ = \quad & (n-2)\delta^{-1}C_0^{-1}\left\{\sum_{k=0}^{1}(-1)^k C_k(C^*_{2k} - C^*_{1k}) + \mathcal{O}_P(n^{-1}\delta^{-1}) + \mathcal{O}_P(\delta^{1/2})\right\}. \end{aligned}$$

Again by Proposition 4, we obtain

$$1 - |\rho_{1,n}(\mathbf{a})|^2 = (1 - |C_1|^2 C_0^{-2})\,\{1 + \mathcal{O}_P(n^{-1}\delta^{-1}) + \mathcal{O}_P(\delta^{1/2})\}.$$

This, combined with an application of Proposition 2, leads to

$$(1 - |\rho_{1,n}(\mathbf{a})|^2)\,V_{0,n}(\mathbf{a}) = (C_0^2 - |C_1|^2)C_0^{-1}(n-2)\delta^{-2}\,\{1 + \mathcal{O}_P(n^{-1}\delta^{-1}) + \mathcal{O}_P(\delta^{1/2})\}.$$

Moreover, direct computation shows that

$$\frac{G_0(\boldsymbol{\omega})\,\delta\,\sum_{k=0}^{1}(-1)^k C_k(C^*_{2k} - C^*_{1k})}{C_0^2 - |C_1|^2} = -b_1\eta\,.$$

Combining these expressions leads to

$$\Omega_n^{-1}(\mathbf{a})\,P_{1,n}(\mathbf{a}) = -b_1\eta + \mathcal{O}_P(n^{-1}) + \mathcal{O}_P(\delta^{3/2}). \tag{28}$$

Similar calculations show that

$$\Omega_n^{-1}(\mathbf{a})\,P_{2,n}(\mathbf{a}) = \frac{G_0(\boldsymbol{\omega})\,\{U^*_{2,n}(\mathbf{a}) - \rho^*_{1,n}(\mathbf{a})\,U^*_{1,n}(\mathbf{a})\}}{\{1 - |\rho_{1,n}(\mathbf{a})|^2\}\,V_{0,n}(\mathbf{a})},$$

and

$$\frac{G_0(\boldsymbol{\omega})\,\delta\,\{C_0(C^*_{2,1} - C^*_{1,1}) - C^*_1(C^*_{2,0} - C^*_{1,0})\}}{C_0^2 - |C_1|^2} = -b_2\eta^2.$$

Therefore, we obtain

$$\Omega_n^{-1}(\mathbf{a})\,P_{2,n}(\mathbf{a}) = -b_2\eta^2 + \mathcal{O}_P(n^{-1}) + \mathcal{O}_P(\delta^{3/2}). \tag{29}$$

Finally, it can be shown that both $R_{1,n}(\mathbf{a})$ and $R_{2,n}(\mathbf{a})$ are of smaller order than $\mathcal{O}_P(n^{-1}) + \mathcal{O}_P(\delta^{3/2})$. Under the conditions of Theorem 1, $\mathcal{O}_P(n^{-1}) = o_P(\delta^{\varepsilon})$ and $\mathcal{O}_P(\delta^{3/2}) = o_P(\delta^{\varepsilon})$. Therefore, the proof of (14) is complete by plugging these expressions together with (28) and (29) into (27).

To show (13), we note from Proposition 1 that for any $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathcal{A}_{\delta}$,

$$\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \hat{\mathbf{a}}(\boldsymbol{\alpha}') = (\boldsymbol{\alpha} - \boldsymbol{\alpha}') + \mathbf{D}^{-1}\{\Omega_n^{-1}(\boldsymbol{\alpha}')\,\mathbf{p}_n(\boldsymbol{\alpha}') - \Omega_n^{-1}(\boldsymbol{\alpha})\,\mathbf{p}_n(\boldsymbol{\alpha})\} + \mathbf{r}_n(\boldsymbol{\alpha}) - \mathbf{r}_n(\boldsymbol{\alpha}'). \tag{30}$$

To evaluate the second term, let

$$\begin{aligned}
\overline{V}_{0,n}(\boldsymbol{\alpha}) &= \{1 - |\rho_{1,n}(\boldsymbol{\alpha})|^2\} V_{0,n}(\boldsymbol{\alpha}), \\
\Delta U^*_{j,n}(\boldsymbol{\alpha}', \boldsymbol{\alpha}) &= U^*_{j,n}(\boldsymbol{\alpha}') - U^*_{j,n}(\boldsymbol{\alpha}) \quad (j = 1, 2), \\
\Delta G_0(\boldsymbol{\lambda}', \boldsymbol{\lambda}) &= G_0(\boldsymbol{\lambda}') - G_0(\boldsymbol{\lambda}), \\
\Delta \rho^*_{1,n}(\boldsymbol{\alpha}, \boldsymbol{\alpha}') &= \rho^*_{1,n}(\boldsymbol{\alpha}) - \rho^*_{1,n}(\boldsymbol{\alpha}'), \\
\Delta \overline{V}_{0,n}(\boldsymbol{\alpha}, \boldsymbol{\alpha}') &= \overline{V}_{0,n}(\boldsymbol{\alpha}) - \overline{V}_{0,n}(\boldsymbol{\alpha}').
\end{aligned}$$

Then, it follows from (21) and (20) that

$$\Omega_n^{-1}(\boldsymbol{\alpha}')\,P_{1,n}(\boldsymbol{\alpha}') - \Omega_n^{-1}(\boldsymbol{\alpha})\,P_{1,n}(\boldsymbol{\alpha}) = \sum_{k=1}^{4} A_k(\boldsymbol{\alpha}, \boldsymbol{\alpha}'),$$

where

$$\begin{aligned}
A_1(\boldsymbol{\alpha}, \boldsymbol{\alpha}') &:= \frac{G_0(\boldsymbol{\lambda}')\,\{\Delta U^*_{1,n}(\boldsymbol{\alpha}', \boldsymbol{\alpha}) - \rho_{1,n}(\boldsymbol{\alpha}')\,\Delta U^*_{2,n}(\boldsymbol{\alpha}', \boldsymbol{\alpha})\}}{\overline{V}_{0,n}(\boldsymbol{\alpha}')}, \\
A_2(\boldsymbol{\alpha}, \boldsymbol{\alpha}') &:= \frac{\Delta G_0(\boldsymbol{\lambda}', \boldsymbol{\lambda})\,\{U^*_{1,n}(\boldsymbol{\alpha}) - \rho_{1,n}(\boldsymbol{\alpha})\,U^*_{2,n}(\boldsymbol{\alpha})\}}{\overline{V}_{0,n}(\boldsymbol{\alpha}')}, \\
A_3(\boldsymbol{\alpha}, \boldsymbol{\alpha}') &:= \frac{G_0(\boldsymbol{\lambda}')\,\Delta\rho_{1,n}(\boldsymbol{\alpha}, \boldsymbol{\alpha}')\,U^*_{2,n}(\boldsymbol{\alpha})}{\overline{V}_{0,n}(\boldsymbol{\alpha}')}, \\
A_4(\boldsymbol{\alpha}, \boldsymbol{\alpha}') &:= \frac{G_0(\boldsymbol{\lambda})\,\{U^*_{1,n}(\boldsymbol{\alpha}) - \rho_{1,n}(\boldsymbol{\alpha})\,U^*_{2,n}(\boldsymbol{\alpha})\}\,\Delta\overline{V}_{0,n}(\boldsymbol{\alpha}, \boldsymbol{\alpha}')}{\overline{V}_{0,n}(\boldsymbol{\alpha})\,\overline{V}_{0,n}(\boldsymbol{\alpha}')}.
\end{aligned}$$

According to Proposition 3,

$$\begin{aligned}
& G_0(\boldsymbol{\lambda}')\,\Delta U^*_{k,n}(\boldsymbol{\alpha}', \boldsymbol{\alpha}) \\
&= (n-2)\delta^{-2}\,[-C^*_{1,k-1} + \mathcal{O}_P(\delta^{\varepsilon-1}), C^*_{2,k-1} + \mathcal{O}_P(\delta^{\varepsilon-1})]\,G_0(\boldsymbol{\lambda}')\,i\,(\boldsymbol{\lambda} - \boldsymbol{\lambda}').
\end{aligned}$$

It follows from Lemma 2 and the Taylor series expansion of $e^{-i\lambda_j'}$ at ω_j that

$$G_0(\boldsymbol{\lambda}')\, i\,(\boldsymbol{\lambda}-\boldsymbol{\lambda}') = \begin{bmatrix} 1 & e^{-i\omega_1} \\ -1 & -e^{-i\omega_2} \end{bmatrix} (\boldsymbol{\alpha}-\boldsymbol{\alpha}') + \mathcal{O}(\|\boldsymbol{\alpha}-\boldsymbol{\alpha}'\|^2).$$

By noting the definition of C_k and C_{jk}, we obtain

$$\begin{aligned} G_0(\boldsymbol{\lambda}')\,\Delta U_{1,n}^*(\boldsymbol{\alpha}',\boldsymbol{\alpha}) &= (n-2)\delta^{-2}[C_0(1+\mathcal{O}_P(\delta^{\varepsilon-1})), C_1(1+\mathcal{O}_P(\delta^{\varepsilon-1}))]\,(\boldsymbol{\alpha}'-\boldsymbol{\alpha}), \\ G_0(\boldsymbol{\lambda}')\,\Delta U_{2,n}^*(\boldsymbol{\alpha}',\boldsymbol{\alpha}) &= (n-2)\delta^{-2}[C_1^*(1+\mathcal{O}_P(\delta^{\varepsilon-1})), C_0(1+\mathcal{O}_P(\delta^{\varepsilon-1}))]\,(\boldsymbol{\alpha}'-\boldsymbol{\alpha}). \end{aligned}$$

Moreover, it follows from Proposition 4 that

$$\rho_{k,n}(\boldsymbol{\alpha}) = C_k C_0^{-1}\{1+\mathcal{O}_P(n^{-1}\delta^{-1})+\mathcal{O}_P(\delta^{1/2})+\mathcal{O}_P(\delta^{2\varepsilon-2})\}.$$

Therefore,

$$\begin{aligned} &G_0(\boldsymbol{\lambda}')\,\rho_{1,n}(\boldsymbol{\alpha}')\,\Delta U_{2,n}^*(\boldsymbol{\alpha}',\boldsymbol{\alpha}) \\ &\quad = (n-2)\delta^{-2}\,[|C_1|^2 C_0^{-1}(1+\mathcal{O}_P(\delta^{\varepsilon-1})), C_1(1+\mathcal{O}_P(\delta^{\varepsilon-1}))]\,(\boldsymbol{\alpha}'-\boldsymbol{\alpha}). \end{aligned}$$

Combining these expressions leads to

$$\begin{aligned} &G_0(\boldsymbol{\lambda}')\,[\Delta U_{1,n}^*(\boldsymbol{\alpha}',\boldsymbol{\alpha}) - \rho_{1,n}(\boldsymbol{\alpha}')\,\Delta U_{2,n}^*(\boldsymbol{\alpha}',\boldsymbol{\alpha})] \\ &\quad = (n-2)\delta^{-2}[(|C_1|^2-C_0^2)\,C_0^{-1}(1+\mathcal{O}_P(\delta^{\varepsilon-1})), \mathcal{O}_P(\delta^{\varepsilon-1})]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}'). \end{aligned}$$

Furthermore, it follows from Proposition 2 and Proposition 4 that

$$\overline{V}_{0,n}(\boldsymbol{\alpha}') = (n-2)\delta^{-2}(C_0^2-|C_1|^2)C_0^{-1}\{1+\mathcal{O}_P(n^{-1}\delta^{-1})+\mathcal{O}_P(\delta^{1/2})\}.$$

Combining these results yields

$$A_1(\boldsymbol{\alpha},\boldsymbol{\alpha}') = [-(1+\mathcal{O}_P(\delta^{\varepsilon-1})), \mathcal{O}_P(\delta^{\varepsilon-1})]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}').$$

Again, by Proposition 2, Proposition 4, and Lemma 2,

$$U_{1,n}^*(\boldsymbol{\alpha}) - \rho_{1,n}(\boldsymbol{\alpha})\,U_{2,n}^*(\boldsymbol{\alpha}) = \mathcal{O}_P(n\delta^{-1}),$$

$$\Delta\rho_{1,n}(\boldsymbol{\alpha},\boldsymbol{\alpha}') = [\mathcal{O}_P(\delta^{\varepsilon-2}),\, \mathcal{O}_P(\delta^{\varepsilon-2})]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}').$$

Using the Taylor expansion and Lemma 2, we can show that

$$\Delta G_0(\boldsymbol{\lambda}',\boldsymbol{\lambda}) = \mathcal{O}(1)\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}').$$

Therefore, we obtain

$$\begin{aligned} A_2(\boldsymbol{\alpha},\boldsymbol{\alpha}') &= [\mathcal{O}_P(\delta),\, \mathcal{O}_P(\delta)]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}'),\\ A_3(\boldsymbol{\alpha},\boldsymbol{\alpha}') &= [\mathcal{O}_P(\delta^{\varepsilon-1}),\, \mathcal{O}_P(\delta^{\varepsilon-1})]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}'). \end{aligned}$$

Similarly, by expressing $\Delta\overline{V}_{0,n}(\boldsymbol{\alpha},\boldsymbol{\alpha}')$ in terms of $V_{0,n}(\boldsymbol{\alpha})-V_{0,n}(\boldsymbol{\alpha}')$ and $\Delta\rho_{1,n}(\boldsymbol{\alpha},\boldsymbol{\alpha}')$, one can show that

$$A_4(\boldsymbol{\alpha},\boldsymbol{\alpha}') = [\mathcal{O}_P(\delta^{\varepsilon-1}),\, \mathcal{O}_P(\delta^{\varepsilon-1})]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}').$$

Combining these expressions for the A_k's leads to

$$\Omega_n^{-1}(\boldsymbol{\alpha}')\,P_{1,n}(\boldsymbol{\alpha}') - \Omega_n^{-1}(\boldsymbol{\alpha})\,P_{1,n}(\boldsymbol{\alpha}) = [-(1+\mathcal{O}_P(\delta^{\varepsilon-1})),\, \mathcal{O}_P(\delta^{\varepsilon-1})]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}'). \tag{31}$$

Using the same technique, we write

$$\Omega_n^{-1}(\boldsymbol{\alpha}')\,P_{2,n}(\boldsymbol{\alpha}') - \Omega_n^{-1}(\boldsymbol{\alpha})\,P_{2,n}(\boldsymbol{\alpha}) = \sum_{k=1}^{4} B_k(\boldsymbol{\alpha},\boldsymbol{\alpha}'),$$

where

$$\begin{aligned} B_1(\boldsymbol{\alpha},\boldsymbol{\alpha}') &:= \frac{G_0(\boldsymbol{\lambda}')\,\{\Delta U^*_{2,n}(\boldsymbol{\alpha}',\boldsymbol{\alpha}) - \rho^*_{1,n}(\boldsymbol{\alpha}')\,\Delta U^*_{1,n}(\boldsymbol{\alpha}',\boldsymbol{\alpha})\}}{\overline{V}_{0,n}(\boldsymbol{\alpha}')},\\ B_2(\boldsymbol{\alpha},\boldsymbol{\alpha}') &:= \frac{\Delta G_0(\boldsymbol{\lambda}',\boldsymbol{\lambda})\,\{U^*_{2,n}(\boldsymbol{\alpha}) - \rho^*_{1,n}(\boldsymbol{\alpha})\,U^*_{1,n}(\boldsymbol{\alpha})\}}{\overline{V}_{0,n}(\boldsymbol{\alpha}')},\\ B_3(\boldsymbol{\alpha},\boldsymbol{\alpha}') &:= \frac{G_0(\boldsymbol{\lambda}')\,\Delta\rho^*_{1,n}(\boldsymbol{\alpha}',\boldsymbol{\alpha})\,U^*_{1,n}(\boldsymbol{\alpha})}{\overline{V}_{0,n}(\boldsymbol{\alpha}')},\\ B_4(\boldsymbol{\alpha},\boldsymbol{\alpha}') &:= \frac{G_0(\boldsymbol{\lambda})\,\{U^*_{2,n}(\boldsymbol{\alpha}) - \rho^*_{1,n}(\boldsymbol{\alpha})\,U^*_{1,n}(\boldsymbol{\alpha})\}\,\Delta\overline{V}_{0,n}(\boldsymbol{\alpha},\boldsymbol{\alpha}')}{\overline{V}_{0,n}(\boldsymbol{\alpha})\,\overline{V}_{0,n}(\boldsymbol{\alpha}')}. \end{aligned}$$

Since

$$\begin{aligned} &G_0(\boldsymbol{\lambda}')\,\rho^*_{1,n}(\boldsymbol{\alpha}')\,\Delta U^*_{1,n}(\boldsymbol{\alpha}',\boldsymbol{\alpha})\\ &= (n-2)\delta^{-2}[C_1^*(1+\mathcal{O}_P(\delta^{\varepsilon-1})),\, |C_1|^2 C_0^{-1}(1+\mathcal{O}_P(\delta^{\varepsilon-1}))]\,(\boldsymbol{\alpha}'-\boldsymbol{\alpha}), \end{aligned}$$

we obtain

$$\begin{aligned} & G_0(\boldsymbol{\lambda}')\,[\Delta U^*_{2,n}(\boldsymbol{\alpha}',\boldsymbol{\alpha}) - \rho^*_{1,n}(\boldsymbol{\alpha}')\,\Delta U^*_{1,n}(\boldsymbol{\alpha}',\boldsymbol{\alpha})] \\ = \;& (n-2)\delta^{-2}\,[\mathcal{O}_P(\delta^{\varepsilon-1}),\,(|C_1|^2 - C_0^2)\,C_0^{-1}\big(1+\mathcal{O}_P(\delta^{\varepsilon-1})\big)]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}'). \end{aligned}$$

This leads to

$$B_1(\boldsymbol{\alpha},\boldsymbol{\alpha}') = [\mathcal{O}_P(\delta^{\varepsilon-1}),\,-(1+\mathcal{O}_P(\delta^{\varepsilon-1}))]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}').$$

Similarly, one can show that $B_k(\boldsymbol{\alpha},\boldsymbol{\alpha}')$, for $k = 2,3,4$, has the same asymptotic expression as $A_k(\boldsymbol{\alpha},\boldsymbol{\alpha}')$. Combining these results yields

$$\Omega_n^{-1}(\boldsymbol{\alpha}')\,P_{2,n}(\boldsymbol{\alpha}') - \Omega_n^{-1}(\boldsymbol{\alpha})\,P_{2,n}(\boldsymbol{\alpha}) = [\mathcal{O}_P(\delta^{\varepsilon-1}),\,-(1+\mathcal{O}_P(\delta^{\varepsilon-1}))]\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}'). \tag{32}$$

Finally, by a similar argument, it can be shown that

$$\mathbf{r}_n(\boldsymbol{\alpha}) - \mathbf{r}_n(\boldsymbol{\alpha}') = o_P(\delta^{\varepsilon-1})\,(\boldsymbol{\alpha}-\boldsymbol{\alpha}').$$

Combining this expression with (30)–(32) proves (13) with $c_n = \mathcal{O}_P(\delta^{\varepsilon-1}) \to 0$ as $n \to \infty$. □

B. Proof of Theorem 2

In the proof of of Theorem 1, we have established (13) and (14). Now, for any $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}' \in \mathcal{A}_\delta$, Eq. (13) implies $\|\{\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \hat{\mathbf{b}}\} - \{\hat{\mathbf{a}}(\boldsymbol{\alpha}') - \hat{\mathbf{b}}\}\| \le c_n\,\|\boldsymbol{\alpha}-\boldsymbol{\alpha}'\|$. It remains to show that the counterpart of (14), i.e., $\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \hat{\mathbf{b}} - \mathbf{a} = o_P(\delta^{\varepsilon})$. Since $\hat{\mathbf{a}}(\mathbf{a}) - \hat{\mathbf{b}} - \mathbf{a} = \{\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b} - \mathbf{a}\} + \mathbf{b} - \hat{\mathbf{b}}$, we obtain

$$\|\hat{\mathbf{a}}(\mathbf{a}) - \hat{\mathbf{b}} - \mathbf{a}\| \le \|\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b} - \mathbf{a}\| + \|\hat{\mathbf{b}} - \mathbf{b}\|.$$

Eq. (14) ensures $\|\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b} - \mathbf{a}\| = o_P(\delta^{\varepsilon})$. By using the Taylor expansion and the continuous mapping theorem, we obtain $\|\hat{\mathbf{b}} - \mathbf{b}\| = \mathcal{O}_P(\delta)\,\|\hat{\boldsymbol{\omega}}_0 - \boldsymbol{\omega}\|$. Combining these expressions with the assumption that $\|\hat{\boldsymbol{\omega}}_0 - \boldsymbol{\omega}\| = o_P(\delta^{\varepsilon-1})$ completes the proof. □

C. Proof of Theorem 3

Recall that we have established (13) and (14) under the same conditions. It follows from (13) that

$$\|\{\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha})\} - \{\hat{\mathbf{a}}(\boldsymbol{\alpha}') - \mathbf{b}(\boldsymbol{\alpha}')\}\| \le c_n \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| + \|\mathbf{b}(\boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha}')\|,$$

where $c_n = \mathcal{O}_P(\delta^{\varepsilon - 1})$. Using the Taylor expansion and the continuous mapping theorem, we obtain

$$\|\mathbf{b}(\boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha}')\| = \mathcal{O}(\delta)\, \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|.$$

Thus, the counterpart of (13) is proved with $c_n' := c_n + \mathcal{O}(\delta) = \mathcal{O}_P(\delta^{\varepsilon - 1})$. The counterpart of (14) follows immediately from (14) and the observation that $\mathbf{b}(\mathbf{a}) = \mathbf{b}$. □

D. Proof of Theorem 4

Since $\hat{\mathbf{a}}$ is the fixed point of $\boldsymbol{\psi}(\hat{\mathbf{a}}(\boldsymbol{\alpha}) - \mathbf{b}(\boldsymbol{\alpha}))$ in $\mathcal{A}_\delta$, we have

$$\hat{\mathbf{a}} - \mathbf{a} = \boldsymbol{\psi}(\hat{\mathbf{a}}(\hat{\mathbf{a}}) - \mathbf{b}(\hat{\mathbf{a}})) - \boldsymbol{\psi}(\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b}(\mathbf{a})) + \boldsymbol{\psi}(\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b}(\mathbf{a})) - \mathbf{a}.$$

Theorem 3 ensures that

$$\boldsymbol{\psi}(\hat{\mathbf{a}}(\hat{\mathbf{a}}) - \mathbf{b}(\hat{\mathbf{a}})) - \boldsymbol{\psi}(\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b}(\mathbf{a})) = o_P(1)\,(\hat{\mathbf{a}} - \mathbf{a}).$$

Therefore,

$$\hat{\mathbf{a}} - \mathbf{a} = \{1 + o_P(1)\}^{-1}\{\boldsymbol{\psi}(\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b}(\mathbf{a})) - \mathbf{a}\}.$$

By Lemma 2(c), we have

$$\boldsymbol{\psi}(\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b}(\mathbf{a})) - \mathbf{a} = \mathcal{O}_P(1)\,\{\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b}(\mathbf{a}) - \mathbf{a}\}.$$

Combining (27)–(29) with the fact that $\mathbf{b}(\mathbf{a}) = \mathbf{b}$ yields that

$$\hat{\mathbf{a}}(\mathbf{a}) - \mathbf{b}(\mathbf{a}) - \mathbf{a} = [\mathcal{O}_P(n^{-1}) + \mathcal{O}_P(\delta^{3/2}),\, \mathcal{O}_P(n^{-1}) + \mathcal{O}_P(\delta^{3/2})].$$

Therefore, under the conditions of Theorem 4,

$$\begin{aligned}\delta^{-\beta}(\hat{\mathbf{a}}-\mathbf{a}) &= [\mathcal{O}_P(n^{-1}\delta^{-\beta})+\mathcal{O}_P(\delta^{3/2-\beta}),\, \mathcal{O}_P(n^{-1}\delta^{-\beta})+\mathcal{O}_P(\delta^{3/2-\beta})] \\ &= [\mathcal{O}_P(1),\mathcal{O}_P(1)].\end{aligned}$$

An application of the Taylor expansion and the continuous mapping theorem to $\boldsymbol{\phi}^{-1}$ completes the proof. □

APPENDIX II: PROOF OF PROPOSITIONS

A. Proof of Proposition 1

It follows from the equation $y^*_{t+1}(\boldsymbol{\alpha})+\alpha^*_1\eta y^*_t(\boldsymbol{\alpha})+\alpha^*_2\eta^2 y^*_{t-1}(\boldsymbol{\alpha})=y^*_{t+1}$ that

$$\Phi_{1,n}(\boldsymbol{\alpha})+\alpha^*_1\eta\Phi_{0,n}(\boldsymbol{\alpha})+\alpha^*_2\eta^2\Phi_{-1,n}(\boldsymbol{\alpha})=\Psi_{1,n}(\boldsymbol{\alpha}).$$

By noting that $y_0(\boldsymbol{\alpha})=y_{-1}(\boldsymbol{\alpha})=0$, we have

$$\Phi_{-1,n}(\boldsymbol{\alpha})=\Phi^*_{1,n}(\boldsymbol{\alpha})-y^*_{n-2,n-1}(\boldsymbol{\alpha}).$$

Thus

$$\rho_{1,n}(\boldsymbol{\alpha})=-\alpha^*_1\eta-\alpha^*_2\eta^2\rho^*_{1,n}(\boldsymbol{\alpha})+\frac{\Psi_{1,n}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})}+\alpha^*_2\eta^2\frac{y^*_{n-2,n-1}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})}. \tag{A.1}$$

Similarly, because $y^*_{t+2}(\boldsymbol{\alpha})+\alpha^*_1\eta y^*_{t+1}(\boldsymbol{\alpha})+\alpha^*_2\eta^2 y^*_t(\boldsymbol{\alpha})=y^*_{t+2}$, we have

$$\Phi_{2,n}(\boldsymbol{\alpha})+\alpha^*_1\eta\Phi_{1,n}(\boldsymbol{\alpha})+\alpha^*_2\eta^2\Phi_{0,n}(\boldsymbol{\alpha})=\Psi_{2,n}(\boldsymbol{\alpha}).$$

Hence

$$\rho_{2,n}(\boldsymbol{\alpha})+\alpha^*_1\eta\rho_{1,n}(\boldsymbol{\alpha})+\alpha^*_2\eta^2=\frac{\Psi_{2,n}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})}.$$

This yields

$$\rho^*_{2,n}(\boldsymbol{\alpha})=-\alpha_1\eta\rho^*_{1,n}(\boldsymbol{\alpha})-\alpha_2\eta^2+\frac{\Psi^*_{2,n}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})}. \tag{A.2}$$

Combining (A.1) and (A.2) gives

$$\rho_{1,n}(\boldsymbol{\alpha})\,\rho^*_{2,n}(\boldsymbol{\alpha}) = -\alpha_1\eta\,|\rho_{1,n}(\boldsymbol{\alpha})|^2 - \alpha_2\eta^2\rho_{1,n}(\boldsymbol{\alpha}) + \rho_{1,n}(\boldsymbol{\alpha})\frac{\Psi^*_{2,n}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})}.$$

This, coupled with (A.1), leads to

$$\begin{aligned}&\rho^*_{1,n}(\boldsymbol{\alpha}) - \rho_{1,n}(\boldsymbol{\alpha})\rho^*_{2,n}(\boldsymbol{\alpha})\\ &= -\alpha_1\eta\{1-|\rho_{1,n}(\boldsymbol{\alpha})|^2\} + \frac{\Psi^*_{1,n}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})} - \rho_{1,n}(\boldsymbol{\alpha})\frac{\Psi^*_{2,n}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})} + \alpha_2\eta^2\frac{y^*_{n-2,n-1}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})}.\end{aligned} \tag{A.3}$$

Similarly,

$$\begin{aligned}&\rho^*_{2,n}(\boldsymbol{\alpha}) - \{\rho^*_{1,n}(\boldsymbol{\alpha})\}^2\\ &= -\alpha_2\eta^2\{1-|\rho_{1,n}(\boldsymbol{\alpha})|^2\} + \frac{P_{2,n}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})} - \frac{\alpha_2\eta^2\rho^*_{1,n}(\boldsymbol{\alpha})\,y_{n-2,n-1}(\boldsymbol{\alpha})}{\Phi_{0,n}(\boldsymbol{\alpha})}.\end{aligned} \tag{A.4}$$

Now, to evaluate $\hat{\mathbf{a}}(\boldsymbol{\alpha}) := [\hat{a}_1(\boldsymbol{\alpha}), \hat{a}_2(\boldsymbol{\alpha})]^T$, we first note that

$$\begin{aligned}\hat{a}_1(\boldsymbol{\alpha}) &= \eta^{-1}\det\{\mathbf{Y}^H(\boldsymbol{\alpha})\mathbf{Y}(\boldsymbol{\alpha})\}^{-1}\{[\rho_{1,n}(\boldsymbol{\alpha})\rho^*_{2,n}(\boldsymbol{\alpha}) - \rho^*_{1,n}(\boldsymbol{\alpha})]\,\Phi_{0,n}(\boldsymbol{\alpha})\\ &\quad + y^*_{1,2}(\boldsymbol{\alpha}) - y^*_{n-1,n}(\boldsymbol{\alpha})\}\end{aligned} \tag{A.5}$$

Direct calculation shows that

$$\det\{\mathbf{Y}^H(\boldsymbol{\alpha})\mathbf{Y}(\boldsymbol{\alpha})\} = \Omega_n(\boldsymbol{\alpha})\,\Phi_{0,n}(\boldsymbol{\alpha}) + \{y_{n-1,n-1}(\boldsymbol{\alpha}) - y_{1,1}(\boldsymbol{\alpha})\}\,\Phi_{0,n}(\boldsymbol{\alpha}). \tag{A.6}$$

Substituting (A.3) and (A.6) in (A.5) gives

$$\hat{a}_1(\boldsymbol{\alpha}) = \frac{\alpha_1\Omega_n(\boldsymbol{\alpha}) - \eta^{-1}P_{1,n}(\boldsymbol{\alpha}) - \alpha_2\eta y^*_{n-2,n-1}(\boldsymbol{\alpha}) - \eta^{-1}\{y^*_{n-1,n}(\boldsymbol{\alpha}) - y^*_{1,2}(\boldsymbol{\alpha})\}}{\Omega_n(\boldsymbol{\alpha}) + y_{n-1,n-1}(\boldsymbol{\alpha}) - y_{1,1}(\boldsymbol{\alpha})}.$$

This expression can be further simplified as

$$\hat{a}_1(\boldsymbol{\alpha}) = \alpha_1 - \frac{P_{1,n}(\boldsymbol{\alpha})}{\eta\Omega_n(\boldsymbol{\alpha})} + R_{1,n}(\boldsymbol{\alpha}). \tag{A.7}$$

Similarly, we obtain

$$\begin{aligned}\hat{a}_2(\boldsymbol{\alpha}) &= \frac{\{(\rho^*_{1,n}(\boldsymbol{\alpha}))^2 - \rho^*_{2,n}(\boldsymbol{\alpha})\}\,\Phi_{0,n}(\boldsymbol{\alpha}) - \rho^*_{2,n}(\boldsymbol{\alpha})\,\{y_{n-1,n-1}(\boldsymbol{\alpha}) - y_{1,1}(\boldsymbol{\alpha})\}}{\eta^2\{\Omega_n(\boldsymbol{\alpha}) + y_{n-1,n-1}(\boldsymbol{\alpha}) - y_{1,1}(\boldsymbol{\alpha})\}}\\ &\quad + \frac{\rho^*_{1,n}(\boldsymbol{\alpha})\,\{y^*_{n-1,n}(\boldsymbol{\alpha}) - y^*_{1,2}(\boldsymbol{\alpha})\}}{\eta^2\{\Omega_n(\boldsymbol{\alpha}) + y_{n-1,n-1}(\boldsymbol{\alpha}) - y_{1,1}(\boldsymbol{\alpha})\}}.\end{aligned}$$

Using (A.4), one can show that

$$\begin{aligned}\hat{a}_2(\boldsymbol{\alpha}) &= \alpha_2 - \frac{\eta^{-2}P_{2,n}(\boldsymbol{\alpha}) - Q_{2,n}(\boldsymbol{\alpha})}{\Omega_n(\boldsymbol{\alpha})} - \frac{\alpha_2\{y_{n-1,n-1}(\boldsymbol{\alpha}) - y_{1,1}(\boldsymbol{\alpha})\}}{\Omega_n(\boldsymbol{\alpha}) + y_{n-1,n-1}(\boldsymbol{\alpha}) - y_{1,1}(\boldsymbol{\alpha})} \\ &\quad + \frac{\{\eta^{-2}P_{2,n}(\boldsymbol{\alpha}) - Q_{2,n}(\boldsymbol{\alpha})\}\{y_{n-1,n-1}(\boldsymbol{\alpha}) - y_{1,1}(\boldsymbol{\alpha})\}}{\Omega_n(\boldsymbol{\alpha})\{\Omega_n(\boldsymbol{\alpha}) + y_{n-1,n-1}(\boldsymbol{\alpha}) - y_{1,1}(\boldsymbol{\alpha})\}}.\end{aligned}$$

This expression can be further simplified as

$$\hat{a}_2(\boldsymbol{\alpha}) = \alpha_2 - \frac{P_{2,n}(\boldsymbol{\alpha})}{\eta^2\Omega_n(\boldsymbol{\alpha})} + R_{2,n}(\boldsymbol{\alpha}). \tag{A.8}$$

Collecting (A.7) and (A.8) proves the assertion. □

B. Proof of Proposition 2

Let us introduce the following notation:

$$x_t(\boldsymbol{\alpha}) := \sum_{l=0}^{t-1} h_l(\boldsymbol{\lambda}) x_{t-l}, \tag{B.1}$$

$$\varepsilon_t(\boldsymbol{\alpha}) := \sum_{l=0}^{t-1} h_l(\boldsymbol{\lambda}) \varepsilon_{t-l}, \tag{B.2}$$

$$v_t(\boldsymbol{\lambda}) := \sum_{l=0}^{t-1} \eta^l G_l(\boldsymbol{\lambda}) \varepsilon_{t-l}, \tag{B.3}$$

$$u_{t,j}(\boldsymbol{\lambda}) := r_j z_j^t \sum_{l=0}^{t-1} \eta^l G_l(\boldsymbol{\lambda}) z_j^{-l} \qquad (j = 1, 2). \tag{B.4}$$

where $z_j := e^{i\omega_j}$, and $r_j := \beta_j e^{i\phi_j}$ $(j = 1, 2)$. Since $y_t(\boldsymbol{\alpha}) := \sum_{l=0}^{t-1} h_l(\boldsymbol{\lambda}) y_{t-l}$, we have $y_t(\boldsymbol{\alpha}) = x_t(\boldsymbol{\alpha}) + \varepsilon_t(\boldsymbol{\alpha})$.

Thus

$$V_{k,n}(\boldsymbol{\alpha}) = |G_0(\boldsymbol{\lambda})|^2 \Phi_{k,n}(\boldsymbol{\alpha}) := \sum_{j=1}^{4} T_j(\boldsymbol{\lambda}), \tag{B.5}$$

where

$$
\begin{aligned}
T_1(\boldsymbol{\lambda}) &:= |G_0(\boldsymbol{\lambda})|^2 \sum_{t=1}^{n-2} x_t(\boldsymbol{\alpha}) x_{t+k}^*(\boldsymbol{\alpha}), \\
T_2(\boldsymbol{\lambda}) &:= |G_0(\boldsymbol{\lambda})|^2 \sum_{t=1}^{n-2} \varepsilon_t(\boldsymbol{\alpha}) x_{t+k}^*(\boldsymbol{\alpha}), \\
T_3(\boldsymbol{\lambda}) &:= |G_0(\boldsymbol{\lambda})|^2 \sum_{t=1}^{n-2} x_t(\boldsymbol{\alpha}) \varepsilon_{t+k}^*(\boldsymbol{\alpha}), \\
T_4(\boldsymbol{\lambda}) &:= |G_0(\boldsymbol{\lambda})|^2 \sum_{t=1}^{n-2} \varepsilon_t(\boldsymbol{\alpha}) \varepsilon_{t+k}^*(\boldsymbol{\alpha}).
\end{aligned}
$$

The goal of the proof is to establish the following expressions:

$$
\begin{aligned}
T_1(\boldsymbol{\lambda}) &= C_k(n-2)\delta^{-2}\{1+\mathcal{O}(n^{-1}\delta^{-1})+\mathcal{O}(\delta)\} \\
&\quad + [\mathcal{O}(\delta^{-4})+\mathcal{O}(n\delta^{-2}), \mathcal{O}(\delta^{-4})+\mathcal{O}(n\delta^{-2})]\,(\boldsymbol{\lambda}-\boldsymbol{\omega}) \\
&\quad + \mathcal{O}(\delta^{-4}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2)+\mathcal{O}(\delta^{-5}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^3)+\mathcal{O}(\delta^{-6}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^4), \\
T_2(\boldsymbol{\lambda}) &= \mathcal{O}_P(n\delta^{-3/2}) \\
&\quad + [\mathcal{O}_P(n\delta^{-5/2}), \mathcal{O}_P(n\delta^{-5/2})]\,(\boldsymbol{\lambda}-\boldsymbol{\omega}) \\
&\quad + \mathcal{O}_P(n\delta^{-7/2}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2)+\mathcal{O}_P(n\delta^{-9/2}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^3)+\mathcal{O}_P(n\delta^{-11/2}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^4), \\
T_3(\boldsymbol{\lambda}) &= \text{same expression as } T_2(\boldsymbol{\lambda}), \\
T_4(\boldsymbol{\lambda}) &= \mathcal{O}_P(n\delta^{-1}) \\
&\quad + [\mathcal{O}_P(n\delta^{-2}), \mathcal{O}_P(n\delta^{-2})]\,(\boldsymbol{\lambda}-\boldsymbol{\omega}) \\
&\quad + \mathcal{O}_P(n\delta^{-3}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2)+\mathcal{O}_P(n\delta^{-4}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^3)+\mathcal{O}_P(n\delta^{-5}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^4).
\end{aligned}
$$

As can be seen, the assertion in Proposition 2 follows immediately from these expressions.

To simplify the evaluation of these expressions as functions of $\boldsymbol{\lambda}$, let us introduce some additional notation. For $j=1,2$, the first and second partial derivatives of a function f with respect to λ_j will be denoted by $D_j f$ and $D_j^2 f$, respectively. The mixed partial derivatives will be denoted by $D_{21}^2 f$ and $D_{12}^2 f$. The gradient (Jacobian) row vector evaluated at $\boldsymbol{\omega}$ will be denoted by $J(f(\boldsymbol{\omega})) := [D_1(f(\boldsymbol{\omega})), D_2(f(\boldsymbol{\omega}))]$.

The Hessian matrix of the second derivatives at $\tilde{\boldsymbol{\lambda}}$ will be denoted by $\boldsymbol{H}(f(\tilde{\boldsymbol{\lambda}}))$. By a slight abuse of notation, any intermediate point between two given vectors will be denoted by $\tilde{\boldsymbol{\lambda}}$, which may vary not only at different expressions, but also within the same expression (e.g., it may vary from the real part to the imaginary part of a complex-valued function). For any complex number z and real number $r > 0$, let $g_t(z,r) := (z - r^t z^{t+1})/(1 - zr)$, where t is any positive integer. Let $z_{jl} := z_j z_l^*$ and $\phi_{jl} := \phi_j - \phi_l$.

First, let us derive the expression for $T_1(\boldsymbol{\lambda})$. It follows from (B.1) and (B.4) that

$$x_t(\alpha) = G_0^{-1}(\boldsymbol{\lambda}) \sum_{j=1}^{2} u_{t,j}(\boldsymbol{\lambda}).$$

Thus

$$\begin{aligned} T_1(\boldsymbol{\lambda}) &= \sum_{t=1}^{n-2} \{ u_{t,1}(\boldsymbol{\lambda})\, u_{t+k,1}^*(\boldsymbol{\lambda}) + u_{t,2}(\boldsymbol{\lambda})\, u_{t+k,1}^*(\boldsymbol{\lambda}) \\ &\quad + u_{t,1}(\boldsymbol{\lambda})\, u_{t+k,2}^*(\boldsymbol{\lambda}) + u_{t,2}(\boldsymbol{\lambda})\, u_{t+k,2}^*(\boldsymbol{\lambda}) \}. \end{aligned}$$

For $j = 1,2$, we have the following Taylor expansion at $\boldsymbol{\omega}$:

$$u_{t,j}(\boldsymbol{\lambda}) = u_{t,j}(\boldsymbol{\omega}) + \boldsymbol{J}(u_{t,j}(\boldsymbol{\omega}))\,(\boldsymbol{\lambda} - \boldsymbol{\omega}) + \tfrac{1}{2}(\boldsymbol{\lambda} - \boldsymbol{\omega})^T \boldsymbol{H}(u_{t,j}(\tilde{\boldsymbol{\lambda}}))\,(\boldsymbol{\lambda} - \boldsymbol{\omega}).$$

Since $D_{jk}^2(G_l(\tilde{\boldsymbol{\lambda}})) = 0$ for $j \neq k$ and $D_j^2(G_l(\tilde{\boldsymbol{\lambda}})) = (-1)^{j+1}(l+1)^2 e^{i(l+1)\tilde{\lambda}_j}$, we have

$$\|\boldsymbol{H}(G_l(\tilde{\boldsymbol{\lambda}}))\| = \mathcal{O}((l+1)^2),$$

where $\|\boldsymbol{H}(G_l(\tilde{\boldsymbol{\lambda}}))\|$ denotes the matrix norm induced by the vector norm and the big $\mathcal{O}$ term is uniform in $\tilde{\boldsymbol{\lambda}}$. It follows from (B.4) that

$$\|\boldsymbol{H}(u_{t,j}(\tilde{\boldsymbol{\lambda}}))\| \le \beta_j \sum_{l=0}^{t-1} \eta^l \|\boldsymbol{H}(G_l(\tilde{\boldsymbol{\lambda}}))\| = \mathcal{O}\left(\sum_{l=0}^{\infty} \eta^l (l+1)^2 \right) = \mathcal{O}(\delta^{-3}).$$

Thus

$$u_{t,j}(\boldsymbol{\lambda}) = u_{t,j}(\boldsymbol{\omega}) + \boldsymbol{J}(u_{t,j}(\boldsymbol{\omega}))\,(\boldsymbol{\lambda} - \boldsymbol{\omega}) + \mathcal{O}(\delta^{-3}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2). \tag{B.6}$$

Using this expression, we obtain

$$
\begin{aligned}
& u_{t,j}(\boldsymbol{\lambda})\, u^*_{t+k,j}(\boldsymbol{\lambda}) \\
= \;& u_{t,j}(\boldsymbol{\omega})\, u^*_{t+k,j}(\boldsymbol{\omega}) + \{u_{t,j}(\boldsymbol{\omega})\, \boldsymbol{J}^*(u_{t+k,j}(\boldsymbol{\omega})) + u^*_{t+k,j}(\boldsymbol{\omega})\, \boldsymbol{J}(u_{t,j}(\boldsymbol{\omega}))\}\, (\boldsymbol{\lambda}-\boldsymbol{\omega}) \\
& + \boldsymbol{J}(u_{t,j}(\boldsymbol{\omega}))\, (\boldsymbol{\lambda}-\boldsymbol{\omega})\, \boldsymbol{J}^*(u_{t+k,j}(\boldsymbol{\omega}))\, (\boldsymbol{\lambda}-\boldsymbol{\omega}) + \mathcal{O}\big(u_{t,j}(\boldsymbol{\omega})\, \delta^{-3} \|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2\big) \\
& + \mathcal{O}\big(u^*_{t+k,j}(\boldsymbol{\omega})\, \delta^{-3} \|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2\big) + \mathcal{O}\big(\boldsymbol{J}(u_{t,j}(\boldsymbol{\omega}))\, (\boldsymbol{\lambda}-\boldsymbol{\omega}) \delta^{-3} \|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2\big) \\
& + \mathcal{O}\big(\boldsymbol{J}^*(u_{t+k,j}(\boldsymbol{\omega}))\, (\boldsymbol{\lambda}-\boldsymbol{\omega}) \delta^{-3} \|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2\big) + \mathcal{O}\big(\delta^{-6} \|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^4\big).
\end{aligned}
$$

Direct computation shows that

$$
u_{t,j}(\boldsymbol{\omega}) = r_j z_j^{t+1} \{-g_t(z_{1j}, \eta) + g_t(z_{2j}, \eta)\}. \tag{B.7}
$$

This, coupled with the fact that $g^*_t(z, r) = g_t(z^*, r)$ and $z^*_{jl} = z_{lj}$, implies

$$
\begin{aligned}
u_{t,j}(\boldsymbol{\omega})\, u^*_{t+k,j}(\boldsymbol{\omega}) \;=\;& C_{jk}\{g_t(z_{2j}, \eta)\, g^*_{t+k}(z_{j2}, \eta) - g_t(z_{1j}, \eta)\, g^*_{t+k}(z_{j2}, \eta) \\
& - g_t(z_{2j}, \eta)\, g^*_{t+k}(z_{j1}, \eta) + g_t(z_{1j}, \eta)\, g^*_{t+k}(z_{j1}, \eta)\}.
\end{aligned}
$$

By noting that $z_{jj} = 1$, we obtain

$$
\begin{aligned}
& \sum_{j=1}^{2} u_{t,j}(\boldsymbol{\omega}) u^*_{t+k,j}(\boldsymbol{\omega}) \\
= \;& C_k\, g_t(1, \eta) g_{t+k}(1, \eta) + C_{1k}\, g_t(z_{21}, \eta)\, g^*_{t+k}(z_{12}, \eta) + C_{2k}\, g_t(z_{12}, \eta)\, g^*_{t+k}(z_{21}, \eta) \\
& - \sum_{j=1}^{2} C_{jk}\{g_t(z_{1j}, \eta)\, g^*_{t+k}(z_{j2}, \eta) + g_t(z_{2j}, \eta)\, g^*_{t+k}(z_{j1}, \eta)\}.
\end{aligned}
$$

Note that $n\delta^{\varepsilon} \to \infty$ implies $n\delta \to \infty$. Note also that the leading term of $\sum_{t=1}^{n-2} C_k\, g_t(1, \eta)\, g_{t+k}(1, \eta)$ takes the form $C_k(n-2)\delta^{-2}$. Therefore,

$$
\sum_{t=1}^{n-2} \sum_{j=1}^{2} u_{t,j}(\boldsymbol{\omega})\, u^*_{t+k,j}(\boldsymbol{\omega}) = C_k(n-2)\delta^{-2}\{1 + \mathcal{O}(n^{-1}\delta^{-1}) + \mathcal{O}(\delta)\}. \tag{B.8}
$$

Using the same argument, one can show that

$$
D^*_{j'}(u_{t+k,j}(\boldsymbol{\omega})) = (-1)^{j'+1} i r^*_j z^*_{j'} z_j^{-(t+k)} \sum_{l=0}^{t+k-1} (l+1) \eta^l z^l_{jj'} \quad (j' = 1, 2). \tag{B.9}
$$

Thus,

$$
\begin{aligned}
u_{t,j}(\boldsymbol{\omega})D_{j'}^{*}(u_{t+k,j}(\boldsymbol{\omega})) &= (-1)^{j'+1}i\beta_j^2 z_{jj'}z_j^{-k}\sum_{m=1}^{2}(-1)^m g_t(z_{mj},\eta)\sum_{l=0}^{t+k-1}(l+1)\eta^l z_{jj'}^l, \\
u_{t+k,j}^{*}(\boldsymbol{\omega})D_{j'}(u_{t,j}(\boldsymbol{\omega})) &= (-1)^{j'}i\beta_j^2 z_{j'j}z_j^{-k}\sum_{m=1}^{2}(-1)^m g_{t+k}^{*}(z_{mj},\eta)\sum_{l=0}^{t-1}(l+1)\eta^l z_{j'j}^l.
\end{aligned}
$$

In the case of $j=j'$, we obtain

$$
\begin{aligned}
& u_{t,j}(\boldsymbol{\omega})D_{j}^{*}(u_{t+k,j}(\boldsymbol{\omega})) + u_{t+k,j}^{*}(\boldsymbol{\omega})D_{j}(u_{t,j}(\boldsymbol{\omega})) \\
&= (-1)^j iC_{jk}\Bigg\{\sum_{m=1}^{2}(-1)^m g_{t+k}^{*}(z_{mj},\eta)\sum_{l=0}^{t-1}(l+1)\eta^l \\
&\qquad -\sum_{m=1}^{2}(-1)^m g_t(z_{mj},\eta)\sum_{l=0}^{t+k-1}(l+1)\eta^l\Bigg\} \\
&= (-1)^j iC_{jk}\Bigg\{\sum_{m=1}^{2}(-1)^m [g_{t+k}^{*}(z_{mj},\eta)-g_t(z_{mj},\eta)]\sum_{l=0}^{t-1}(l+1)\eta^l \\
&\qquad -\sum_{m=1}^{2}(-1)^m g_t(z_{mj},\eta)\sum_{l=t}^{t+k-1}(l+1)\eta^l\Bigg\}.
\end{aligned}
$$

Direct calculation shows that

$$
\begin{aligned}
& g_{t+k}^{*}(z_{mj},\eta)-g_t(z_{mj},\eta) \\
&\quad = \frac{(z_{mj}^{*}-z_{mj})+\eta^t\{z_{mj}^t(z_{mj}-\eta)-\eta^k z_{mj}^{-(t+k)}(z_{mj}^{*}-\eta)\}}{|1-\eta z_{mj}|^2}.
\end{aligned}
$$

For $m=j$, we have $g_{t+k}^{*}(z_{jj},\eta)-g_t(z_{jj},\eta)=\eta^t(1-\eta^k)\delta^{-1}$, which, coupled with the formula

$$
\sum_{l=0}^{t-1}(l+1)\eta^l = (1-\eta^t-\delta t\eta^t)\delta^{-2},
$$

leads to

$$
\{g_{t+k}^{*}(z_{jj},\eta)-g_t(z_{jj},\eta)\}\sum_{l=0}^{t-1}(l+1)\eta^l = \eta^t(1-\eta^k)(1-\eta^t-\delta t\eta^t)\delta^{-3}.
$$

For $m\neq j$, $\{g_{t+k}^{*}(z_{mj},\eta)-g_t(z_{mj},\eta)\}\sum_{l=0}^{t-1}(l+1)\eta^l = \mathcal{O}(\delta^{-2})$ uniformly in t. Thus

$$
\sum_{t=1}^{n-2}\sum_{m=1}^{2}(-1)^m\{g_{t+k}^{*}(z_{mj},\eta)-g_t(z_{mj},\eta)\}\sum_{l=0}^{t-1}(l+1)\eta^l = \mathcal{O}(\delta^{-4})+\mathcal{O}(n\delta^{-2}).
$$

A similar argument with an application of the formula

$$\sum_{l=t}^{t+k-1}(l+1)\eta^l = \eta^t\{(1-\eta^k)(1+\delta t)-\delta k\eta^k\}\delta^{-2}$$

shows that

$$\sum_{t=1}^{n-2}\sum_{m=1}^{2}(-1)^m g_t(z_{mj},\eta)\sum_{l=t}^{t+k-1}(l+1)\eta^l = \mathcal{O}(\delta^{-4}).$$

Combining these results leads to

$$\sum_{t=1}^{n-2}\{u_{t,j}(\boldsymbol{\omega})\,D_j^*(u_{t+k,j}(\boldsymbol{\omega}))+u_{t+k,j}^*(\boldsymbol{\omega})\,D_j(u_{t,j}(\boldsymbol{\omega}))\} = \mathcal{O}(\delta^{-4})+\mathcal{O}(n\delta^{-2}).$$

For $j \neq j'$, the order of magnitude of the term

$$\sum_{t=1}^{n-2}\{u_{t,j}(\boldsymbol{\omega})\,D_{j'}^*(u_{t+k,j}(\boldsymbol{\omega}))+u_{t+k,j}^*(\boldsymbol{\omega})\,D_{j'}(u_{t,j}(\boldsymbol{\omega}))\}$$

does not exceed the order of $\mathcal{O}(n\delta^{-2})$. Thus

$$\begin{aligned}&\sum_{t=1}^{n-2}\left\{u_{t,j}(\boldsymbol{\omega})\,\boldsymbol{J}^*(u_{t+k,j}(\boldsymbol{\omega}))+u_{t+k,j}^*(\boldsymbol{\omega})\,\boldsymbol{J}(u_{t,j}(\boldsymbol{\omega}))\right\}\\ &\quad= [\mathcal{O}(\delta^{-4})+\mathcal{O}(n\delta^{-2}),\,\mathcal{O}(\delta^{-4})+\mathcal{O}(n\delta^{-2})].\end{aligned}$$

Finally, it follows from (B.7) and (B.9) that $u_{t,j}(\boldsymbol{\omega}) = \mathcal{O}(\delta^{-1})$ and $D_{j'}^*(u_{t+k,j}(\boldsymbol{\omega})) = \mathcal{O}(\delta^{-2})$ uniformly in t. Combining these results with (B.8) yields

$$\begin{aligned}&\sum_{t=1}^{n-2}u_{t,j}(\boldsymbol{\lambda})\,u_{t+k,j}^*(\boldsymbol{\lambda})\\ &\quad= C_k(n-2)\delta^{-2}\{1+\mathcal{O}(n^{-1}\delta^{-1})+\mathcal{O}(\delta)\}\\ &\qquad+[\mathcal{O}(\delta^{-4})+\mathcal{O}(n\delta^{-2}),\,\mathcal{O}(\delta^{-4})+\mathcal{O}(n\delta^{-2})]\,(\boldsymbol{\lambda}-\boldsymbol{\omega})\\ &\qquad+\mathcal{O}(\delta^{-4}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2)+\mathcal{O}(\delta^{-5}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^3)+\mathcal{O}(\delta^{-6}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^4).\end{aligned}$$

To evaluate $\sum_{t=1}^{n-2}u_{t,j}(\boldsymbol{\lambda})\,u_{t+k,s}^*(\boldsymbol{\lambda})$ in $T_1(\boldsymbol{\lambda})$ for $j \neq s$, it suffices to identify the magnitude of $u_{t,j}(\boldsymbol{\omega})u_{t+k,s}^*(\boldsymbol{\omega})$ and $u_{t,j}(\boldsymbol{\omega})D_{j'}^*(u_{t+k,s}(\boldsymbol{\omega}))+u_{t+k,s}^*(\boldsymbol{\omega})D_{j'}(u_{t,j}(\boldsymbol{\omega}))$ in view of the above analysis. It follows from (B.7) and

(B.9) that

$$u_{t,j}(\boldsymbol{\omega})\,u^*_{t+k,s}(\boldsymbol{\omega}) = r_j r_s^* z_{js}^{t+1} z_s^{-k} \sum_{m=1}^{2}(-1)^m g_t(z_{mj},\eta) \sum_{m=1}^{2}(-1)^m g^*_{t+k}(z_{ms},\eta),$$

$$u_{t,j}(\boldsymbol{\omega})\,D^*_{j'}(u_{t+k,s}(\boldsymbol{\omega})) = (-1)^{j'+1} i r_j r_s^* z^*_{j'} z_j z_{js}^t z_s^{-k} \sum_{m=1}^{2}(-1)^m g_t(z_{mj},\eta) \sum_{l=0}^{t+k-1}(l+1)\eta^l z^l_{sj'}.$$

Thus

$$\sum_{t=1}^{n-2} u_{t,j}(\boldsymbol{\omega})\,u^*_{t+k,s}(\boldsymbol{\omega}) = \mathcal{O}(\delta^{-2}), \quad \sum_{t=1}^{n-2} u_{t,j}(\boldsymbol{\omega})\,D^*_{j'}(u_{t+k,s}(\boldsymbol{\omega})) = \mathcal{O}(\delta^{-3}).$$

Similarly, we obtain

$$\sum_{t=1}^{n-2} u^*_{t+k,s}(\boldsymbol{\omega})\,D_{j'}(u_{t,j}(\boldsymbol{\omega})) = \mathcal{O}(\delta^{-3}).$$

Combining all these results yields the final expression for $T_1(\boldsymbol{\lambda})$.

For the other terms in (B.5), we only outline the proof for $T_2(\boldsymbol{\lambda})$ and $T_4(\boldsymbol{\lambda})$. The proof for $T_3(\boldsymbol{\lambda})$ is the same as that for $T_2(\boldsymbol{\lambda})$ because of the symmetry.

It follows from (B.2) and (B.3) that $\varepsilon_t(\alpha) = G_0^{-1}(\boldsymbol{\lambda}) v_t(\boldsymbol{\lambda})$. This, coupled with the expression of $x_t(\boldsymbol{\alpha})$, implies

$$T_2(\boldsymbol{\lambda}) = \sum_{t=1}^{n-2} v_t(\boldsymbol{\lambda}) \sum_{j=1}^{2} u^*_{t+k,j}(\boldsymbol{\lambda}).$$

The Taylor series expansion of $v_t(\boldsymbol{\lambda})$ at $\boldsymbol{\omega}$ takes the form

$$v_t(\boldsymbol{\lambda}) = v_t(\boldsymbol{\omega}) + \boldsymbol{J}(v_t(\boldsymbol{\omega}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega}) + \tfrac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\omega})^T \boldsymbol{H}(v_t(\tilde{\boldsymbol{\lambda}}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega}). \tag{B.10}$$

It follows from (B.10), (B.6), and Lemma 1 that

$$\begin{aligned}
& v_t(\boldsymbol{\lambda}) u^*_{t+k,j}(\boldsymbol{\lambda}) \\
= \;& v_t(\boldsymbol{\omega})\,u^*_{t+k,j}(\boldsymbol{\omega}) + \{v_t(\boldsymbol{\omega})\,\boldsymbol{J}^*(u_{t+k,j}(\boldsymbol{\omega})) + u^*_{t+k,j}(\boldsymbol{\omega})\,\boldsymbol{J}(v_t(\boldsymbol{\omega}))\}\,(\boldsymbol{\lambda}-\boldsymbol{\omega}) \\
& + \boldsymbol{J}(v_t(\boldsymbol{\omega}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega})\,\boldsymbol{J}^*(u_{t+k,j}(\boldsymbol{\omega}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega}) + \mathcal{O}_P\big(v_t(\boldsymbol{\omega})\delta^{-3}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2\big) \\
& + \mathcal{O}_P\big(u^*_{t+k,j}(\boldsymbol{\omega})\delta^{-5/2}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2\big) + \mathcal{O}_P\big(\boldsymbol{J}(v_t(\boldsymbol{\omega}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega})\delta^{-3}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2\big) \\
& + \mathcal{O}_P\big(\boldsymbol{J}^*(u_{t+k,j}(\boldsymbol{\omega}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega})\delta^{-5/2}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2\big) + \mathcal{O}_P(\delta^{-11/2}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^4).
\end{aligned}$$

Thus, from (B.7) and Lemma 1, we have

$$T_2(\boldsymbol{\omega}) = \sum_{t=1}^{n-2} \sum_{j=1}^{2} v_t(\boldsymbol{\omega}) u_{t+k,j}^*(\boldsymbol{\omega}) = \mathcal{O}_P(n\delta^{-3/2}).$$

Moreover, it follows again from Lemma 1, (B.7), and (B.9) that

$$\begin{aligned} D_{j'}(T_2(\boldsymbol{\omega})) &= \sum_{t=1}^{n-2} \sum_{j=1}^{2} \{v_t(\boldsymbol{\omega}) D_{j'}(u_{t+k,j}^*(\boldsymbol{\omega})) + u_{t+k,j}^*(\boldsymbol{\omega}) D_{j'}(v_t(\boldsymbol{\omega}))\} \\ &= \mathcal{O}_P(n\delta^{-5/2}) \quad (j' = 1,2). \end{aligned}$$

Combining these results leads to the final expression for $T_2(\boldsymbol{\lambda})$.

Finally, we note that $T_4(\boldsymbol{\lambda}) = \sum_{t=1}^{n-2} v_t(\boldsymbol{\lambda}) v_{t+k}^*(\boldsymbol{\lambda})$. By Lemma 1 and (B.10), we have

$$\begin{aligned} & v_t(\boldsymbol{\lambda}) v_{t+k}^*(\boldsymbol{\lambda}) \\ = \ & v_t(\boldsymbol{\omega}) v_{t+k}^*(\boldsymbol{\omega}) + \{v_t(\boldsymbol{\omega}) \boldsymbol{J}^*(v_{t+k}(\boldsymbol{\omega})) + v_{t+k}^*(\boldsymbol{\omega}) \boldsymbol{J}(v_t(\boldsymbol{\omega}))\} (\boldsymbol{\lambda} - \boldsymbol{\omega}) \\ & + \boldsymbol{J}(v_t(\boldsymbol{\omega})) (\boldsymbol{\lambda} - \boldsymbol{\omega}) \boldsymbol{J}^*(v_{t+k}(\boldsymbol{\omega})) (\boldsymbol{\lambda} - \boldsymbol{\omega}) + \mathcal{O}_P\big(v_t(\boldsymbol{\omega})\delta^{-5/2}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2\big) \\ & + \mathcal{O}_P\big(v_{t+k}^*(\boldsymbol{\omega})\delta^{-5/2}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2\big) + \mathcal{O}_P\big(\boldsymbol{J}(v_t(\boldsymbol{\omega})) (\boldsymbol{\lambda} - \boldsymbol{\omega})\delta^{-5/2}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2\big) \\ & + \mathcal{O}_P\big(\boldsymbol{J}^*(v_{t+k}(\boldsymbol{\omega})) (\boldsymbol{\lambda} - \boldsymbol{\omega})\delta^{-5/2}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2\big) + \mathcal{O}_P(\delta^{-5}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^4). \end{aligned}$$

Carrying out the same analysis as above leads to the final expression for $T_4(\boldsymbol{\lambda})$. The proof of Proposition 2 is complete. □

C. Proof of Proposition 3

It follows from (16), (B.2), and (B.3) that

$$U_{k,n}(\boldsymbol{\alpha}) = \sum_{j=1}^{4} S_j(\boldsymbol{\lambda}),$$

where

$$S_1(\boldsymbol{\lambda}) := \sum_{t=1}^{n-2}\sum_{j=1}^{2} u_{t,j}(\boldsymbol{\lambda}) x^*_{t+k}, \tag{C.1}$$

$$S_2(\boldsymbol{\lambda}) := \sum_{t=1}^{n-2} v_t(\boldsymbol{\lambda}) x^*_{t+k}, \tag{C.2}$$

$$S_3(\boldsymbol{\lambda}) := \sum_{t=1}^{n-2}\sum_{j=1}^{2} u_{t,j}(\boldsymbol{\lambda}) \varepsilon^*_{t+k}, \tag{C.3}$$

$$S_4(\boldsymbol{\lambda}) := \sum_{t=1}^{n-2} v_t(\boldsymbol{\lambda}) \varepsilon^*_{t+k}. \tag{C.4}$$

Therefore, to prove Proposition 3, it suffices to derive the following expressions:

$$\begin{aligned}
S_1(\boldsymbol{\lambda}) &= (C_{2,k-1} - C_{1,k-1})(n-2)\delta^{-1} + \mathcal{O}(\delta^{-2}) + \mathcal{O}(n) \\
&\quad + [-iC_{1,k-1}(n-2)\delta^{-2} + \mathcal{O}(\delta^{-3}) + \mathcal{O}(n), \\
&\quad\ iC_{2,k-1}(n-2)\delta^{-2} + \mathcal{O}(\delta^{-3}) + \mathcal{O}(n)]\,(\boldsymbol{\lambda} - \boldsymbol{\omega}) \\
&\quad + \mathcal{O}(n\delta^{-3}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2), \\
S_2(\boldsymbol{\lambda}) &= \mathcal{O}_P(n\delta^{-1/2}) + [\mathcal{O}_P(n\delta^{-3/2}), \mathcal{O}_P(n\delta^{-3/2})]\,(\boldsymbol{\lambda} - \boldsymbol{\omega}) \\
&\quad + \mathcal{O}_P(n\delta^{-5/2}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2), \\
S_3(\boldsymbol{\lambda}) &= \mathcal{O}_P(n^{1/2}\delta^{-1}) + [\mathcal{O}_P(n^{1/2}\delta^{-2}), \mathcal{O}_P(n^{1/2}\delta^{-2})]\,(\boldsymbol{\lambda} - \boldsymbol{\omega}) \\
&\quad + \mathcal{O}_P(n^{1/2}\delta^{-3}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2), \\
S_4(\boldsymbol{\lambda}) &= \mathcal{O}_P(n\delta^{-1/2}) + [\mathcal{O}_P(n\delta^{-3/2}), \mathcal{O}_P(n\delta^{-3/2})]\,(\boldsymbol{\lambda} - \boldsymbol{\omega}) \\
&\quad + \mathcal{O}_P(n\delta^{-5/2}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2).
\end{aligned}$$

Now let us derive these expressions one by one.

First, to evaluate $S_1(\boldsymbol{\lambda})$, we obtain using (B.6) that

$$S_1(\boldsymbol{\lambda}) = S_1(\boldsymbol{\omega}) + \sum_{t=1}^{n-2}\sum_{j=1}^{2} x^*_{t+k}\boldsymbol{J}(u_{t,j}(\boldsymbol{\omega}))\,(\boldsymbol{\lambda} - \boldsymbol{\omega}) + \mathcal{O}(n\delta^{-3}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2). \tag{C.5}$$

Direct computation shows that

$$\begin{aligned}\sum_{j=1}^{2} u_{t,j}(\boldsymbol{\omega}) x_{t+k}^{*} &= \sum_{j=1}^{2} C_{j,k-1} \sum_{m=1}^{2} (-1)^m g_t(z_{mj},\eta) \\ &\quad + \sum_{j\neq l} r_j r_l^* z_{jl}^t z_j z_l^{-k} \sum_{m=1}^{2} (-1)^m g_t(z_{mj},\eta).\end{aligned}$$

Since

$$\begin{aligned}&\sum_{j=1}^{2} C_{j,k-1} \sum_{m=1}^{2} (-1)^m g_t(z_{mj},\eta) \\ &= (C_{2,k-1}-C_{1,k-1})\, g_t(1,\eta) + C_{1,k-1}\, g_t(z_{21},\eta) - C_{2,k-1}\, g_t(z_{12},\eta),\end{aligned}$$

we obtain

$$\begin{aligned}&\sum_{t=1}^{n-2}\sum_{j=1}^{2} C_{j,k-1} \sum_{m=1}^{2} (-1)^m g_t(z_{mj},\eta) \\ &= (C_{2,k-1}-C_{1,k-1})(n-2)\delta^{-1} + \mathcal{O}(\delta^{-2}) + \mathcal{O}(n).\end{aligned}$$

The remaining terms in $S_1(\boldsymbol{\omega})$ is of smaller order than $\mathcal{O}(\delta^{-2})$. Thus,

$$S_1(\boldsymbol{\omega}) = (C_{2,k-1}-C_{1,k-1})(n-2)\delta^{-1} + \mathcal{O}(\delta^{-2}) + \mathcal{O}(n).$$

To evaluate the second term in (C.5), we apply the formula

$$D_{j'}(u_{t,j}(\boldsymbol{\omega})) = (-1)^{j'} i r_j z_{j'} z_j^t \sum_{l=0}^{t-1} (l+1)\eta^l z_{j'j}^l$$

and obtain

$$x_{t+k}^{*} D_{j'}(u_{t,j}(\boldsymbol{\omega})) = (-1)^{j'} i \sum_{m=1}^{2} r_j r_m^* z_{j'} z_{jm}^t z_m^{-k} \sum_{l=0}^{t-1} (l+1)\eta^l z_{j'j}^l.$$

This leads to

$$\begin{aligned}x_{t+k}^{*} \sum_{j=1}^{2} D_1(u_{t,j}(\boldsymbol{\omega})) &= -i(C_{1,k-1} + r_1 r_2^* z_1 z_{12}^t z_2^{-k})\,(1-\eta^t-\delta t\eta^t)\delta^{-2} \\ &\quad - i\Big(C_{2,k} z_1 + r_1^* r_2 z_{21}^t z_1^{-(k-1)}\Big) \sum_{l=0}^{t-1} (l+1)\eta^l z_{12}^l, \\ x_{t+k}^{*} \sum_{j=1}^{2} D_2(u_{t,j}(\boldsymbol{\omega})) &= i(C_{2,k-1} + r_1^* r_2 z_2 z_{21}^t z_1^{-k})\,(1-\eta^t-\delta t\eta^t)\delta^{-2} \\ &\quad + i\Big(C_{1,k} z_2 + r_1 r_2^* z_{12}^t z_2^{-(k-1)}\Big) \sum_{l=0}^{t-1} (l+1)\eta^l z_{21}^l,\end{aligned}$$

hence

$$\begin{aligned}\sum_{t=1}^{n-2} x_{t+k}^* \sum_{j=1}^{2} D_1(u_{t,j}(\boldsymbol{\omega})) &= -iC_{1,k-1}(n-2)\delta^{-2} + \mathcal{O}(\delta^{-3}) + \mathcal{O}(n) \\ \sum_{t=1}^{n-2} x_{t+k}^* \sum_{j=1}^{2} D_2(u_{t,j}(\boldsymbol{\omega})) &= iC_{2,k-1}(n-2)\delta^{-2} + \mathcal{O}(\delta^{-3}) + \mathcal{O}(n).\end{aligned}$$

Substituting these results in (C.5) yields the final expression for $S_1(\boldsymbol{\lambda})$.

Next, consider $S_2(\boldsymbol{\lambda})$, which has the following Taylor expansion:

$$S_2(\boldsymbol{\lambda}) = S_2(\boldsymbol{\omega}) + \boldsymbol{J}(S_2(\boldsymbol{\omega}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega}) + \tfrac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\omega})^T \boldsymbol{H}(S_2(\tilde{\boldsymbol{\lambda}}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega}). \tag{C.6}$$

Using the Cauchy-Schwarz inequality, we obtain

$$|S_2(\boldsymbol{\omega})|^2 \le \sum_{t=1}^{n-2} |v_t(\boldsymbol{\omega})|^2 \sum_{t=1}^{n-2} |x_{t+k}|^2.$$

By Lemma 1, $v_t(\boldsymbol{\omega}) = \mathcal{O}_P(\delta^{-1/2})$ uniformly in t. Thus

$$\sum_{t=1}^{n-2} |v_t(\boldsymbol{\omega})|^2 = \mathcal{O}_P(n\delta^{-1}).$$

This together with the observation that $\sum_{t=1}^{n-2} |x_{t+k}|^2 = \mathcal{O}(n)$ yields

$$S_{k,2}(\boldsymbol{\omega}) = \mathcal{O}_P(n\delta^{-1/2}).$$

To evaluate the second term in (C.6), we note that

$$\boldsymbol{J}(S_2(\boldsymbol{\omega})) = \sum_{t=1}^{n-2} x_{t+k}^* \boldsymbol{J}(v_t(\boldsymbol{\omega})).$$

Moreover, by Lemma 1, $D_j(v_t(\boldsymbol{\omega})) = \mathcal{O}_P(\delta^{-3/2})$ $(j = 1,2)$ uniformly in t. This, coupled with an application of the Cauchy-Schwarz inequality, gives

$$\boldsymbol{J}(S_2(\boldsymbol{\omega})) = \mathcal{O}_P(n\delta^{-3/2}).$$

A similar argument leads to

$$(\boldsymbol{\lambda}-\boldsymbol{\omega})^T \boldsymbol{H}(S_2(\tilde{\boldsymbol{\lambda}}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega}) = \mathcal{O}_P(n\delta^{-5/2}\|\boldsymbol{\lambda}-\boldsymbol{\omega}\|^2).$$

Substituting these expressions in (C.6) gives rise to the final expression for $S_2(\boldsymbol{\lambda})$.

Similarly, the Taylor expansion of $S_3(\boldsymbol{\lambda})$ at $\boldsymbol{\omega}$ takes the form

$$S_3(\boldsymbol{\lambda}) = S_3(\boldsymbol{\omega}) + \boldsymbol{J}(S_3(\boldsymbol{\omega}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega}) + \tfrac{1}{2}(\boldsymbol{\lambda}-\boldsymbol{\omega})^T \boldsymbol{H}(S_3(\tilde{\boldsymbol{\lambda}}))\,(\boldsymbol{\lambda}-\boldsymbol{\omega}). \tag{C.7}$$

The first term in this expansion can be rewritten as

$$S_3(\boldsymbol{\omega}) = \sum_{j=1}^{2}\sum_{t=1}^{n-2} \varepsilon_{t+k}^* u_{t,j}(\boldsymbol{\omega}).$$

Since $\{\varepsilon_t\}$ is a zero-mean white noise process, it follows from (B.7) that

$$E\left|\sum_{t=1}^{n-2} \varepsilon_{t+k}^* u_{t,j}(\boldsymbol{\omega})\right|^2 = \sigma^2\beta_j^2 \sum_{t=1}^{n-2}\left|\sum_{m=1}^{2}(-1)^m g_t(z_{mj},\eta)\right|^2 = \mathcal{O}(n\delta^{-2}).$$

Using this result and Chebyshev's inequality, one can show that

$$\sum_{t=1}^{n-2} \varepsilon_{t+k}^* u_{t,j}(\boldsymbol{\omega}) = \mathcal{O}_P(n^{1/2}\delta^{-1}).$$

Thus,

$$S_3(\boldsymbol{\omega}) = \mathcal{O}_P(n^{1/2}\delta^{-1}).$$

To evaluate the second term in (C.7), we first observe that

$$D_{j'}(S_3(\boldsymbol{\omega})) = \sum_{j=1}^{2}\sum_{t=1}^{n-2} \varepsilon_{t+k}^* D_{j'}(u_{t,j}(\boldsymbol{\omega})) \quad (j'=1,2).$$

Using a similar argument together with the expression of $D_{j'}(u_{t,j}(\boldsymbol{\omega}))$, we obtain

$$\begin{aligned} E\left|\sum_{t=1}^{n-2} \varepsilon_{t+k}^* D_{j'}(u_{t,j}(\boldsymbol{\omega}))\right|^2 &= \sigma^2\beta_j^2 \sum_{t=1}^{n-2}\left(\sum_{l=0}^{t-1}(l+1)\eta^l\right)^2 \\ &= \sigma^2\beta_j^2 \sum_{t=1}^{n-2}(1-\eta^t-\delta t\eta^t)^2\delta^{-4} \\ &= \mathcal{O}(n\delta^{-4}). \end{aligned}$$

An application of Chebyshev's inequality gives

$$\boldsymbol{J}(S_3(\boldsymbol{\omega})) = \mathcal{O}_P(n^{1/2}\delta^{-2}).$$

Combining these results with an application of Lemma 1 to the third term in (C.7) yields the final expression for $S_3(\boldsymbol{\lambda})$.

Finally, the expression for $S_4(\boldsymbol{\lambda})$ can be derived by the same method that led to the expression for $S_2(\boldsymbol{\lambda})$, together with the observation that $\sum_{t=1}^{n-2} |\varepsilon_{t+k}|^2 = \mathcal{O}_P(n)$. This completes the proof. □

D. Proof of Proposition 4

The first part of Proposition 4 follows from Proposition 2 and the identity $(\chi+\varepsilon)^{-1} = \chi^{-1} - \varepsilon[\chi(\chi+\varepsilon)]^{-1}$. To prove the second part of Proposition 4, we express $\rho_{k,n}(\boldsymbol{\alpha}) - \rho_{k,n}(\boldsymbol{\alpha}')$ as

$$\rho_{k,n}(\boldsymbol{\alpha}) - \rho_{k,n}(\boldsymbol{\alpha}') = \frac{V_{k,n}(\boldsymbol{\alpha}) - V_{k,n}(\boldsymbol{\alpha}')}{V_{0,n}(\boldsymbol{\alpha})} - \frac{\{V_{0,n}(\boldsymbol{\alpha}) - V_{0,n}(\boldsymbol{\alpha}')\}V_{k,n}(\boldsymbol{\alpha}')}{V_{0,n}(\boldsymbol{\alpha})V_{0,n}(\boldsymbol{\alpha}')}.$$

An application of Proposition 2 completes the proof. □

E. Some Technical Lemmas

Lemma 1 *Let v_t be defined by (B.3). Then, uniformly in both t and $\boldsymbol{\lambda}$, we have $v_t(\boldsymbol{\lambda}) = \mathcal{O}_P(\delta^{-1/2})$ and $D_j v_t(\boldsymbol{\lambda}) = \mathcal{O}_P(\delta^{-3/2})$ for $j = 1, 2$. Furthermore, uniformly in t, $\boldsymbol{\lambda}$, and $\tilde{\boldsymbol{\lambda}}$, we have $(\boldsymbol{\lambda} - \boldsymbol{\omega})^T \boldsymbol{H}(v_t(\tilde{\boldsymbol{\lambda}}))\,(\boldsymbol{\lambda} - \boldsymbol{\omega}) = \mathcal{O}_P(\delta^{-5/2}\|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^2)$.*

Proof. We provide a proof only for the second assertion because a similar argument can be used to prove the first one. To show the second assertion, we observe that

$$\boldsymbol{H}(v_t(\tilde{\boldsymbol{\lambda}})) = \sum_{l=0}^{t-1} \eta^l \varepsilon_{t-1} \boldsymbol{H}(G_l(\tilde{\boldsymbol{\lambda}}))$$

Since $\{\varepsilon_t\}$ is a zero-mean white noise process with variance σ^2, we obtain

$$\begin{aligned} & E\{|(\boldsymbol{\lambda} - \boldsymbol{\omega})^T \boldsymbol{H}(v_t(\tilde{\boldsymbol{\lambda}}))\,(\boldsymbol{\lambda} - \boldsymbol{\omega})|^2\} \\ &= \sigma^2 \sum_{l=0}^{t-1} \eta^{2l} |(\boldsymbol{\lambda} - \boldsymbol{\omega})^T \boldsymbol{H}(G_j(\tilde{\boldsymbol{\lambda}}))\,(\boldsymbol{\lambda} - \boldsymbol{\omega})|^2 \\ &= \mathcal{O}\left(\sum_{l=0}^{t-1} \eta^{2l} (l+1)^4 \|\boldsymbol{\lambda} - \boldsymbol{\omega}\|^4\right), \end{aligned}$$

where the last equality follows from the fact that $\|\boldsymbol{H}(G_l(\tilde{\boldsymbol{\lambda}}))\| = \mathcal{O}((l+1)^2)$ uniformly in $\tilde{\boldsymbol{\lambda}}$. Since

$$\sum_{l=0}^{t-1} \eta^{2l}(l+1)^4 \leq \sum_{l=0}^{\infty} \eta^{2l}(l+1)^4 = \mathcal{O}(\delta^{-5})$$

uniformly in t, an application of Chebyshev's inequality completes the proof. □

Let $\mathfrak{C}^2$ denote the two-dimensional complex space. For any $\boldsymbol{\alpha} := [\alpha_1, \alpha_2]^T \in \mathfrak{C}^2$, let $\zeta_1 := \rho_1 e^{i\lambda_1}$ and $\zeta_2 := \rho_2 e^{i\lambda_2}$ be the roots of the quadratic polynomial

$$1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} = (1 - \zeta_1 z^{-1})(1 - \zeta_2 z^{-1})$$

where $\rho_1 \geq 0$, $\rho_2 \geq 0$, and $-\pi < \lambda_1 \leq \lambda_2 \leq \pi$. This equation defines a one-to-one mapping between $\boldsymbol{\alpha}$ and $\boldsymbol{\zeta} := [\zeta_1, \zeta_2]^T$ which we denote by $\boldsymbol{\alpha} = \boldsymbol{\xi}(\boldsymbol{\zeta})$. In the special case where $\boldsymbol{\alpha} \in \mathcal{A}$ so that $\boldsymbol{\zeta} = \boldsymbol{\mu}(\boldsymbol{\lambda}) := [e^{i\lambda_1}, e^{i\lambda_2}]^T$ consists of unit roots, the mapping reduces to $\boldsymbol{\alpha} = \boldsymbol{\xi}(\boldsymbol{\mu}(\boldsymbol{\lambda})) = \boldsymbol{\phi}(\boldsymbol{\lambda})$. Note that the roots are distinct if and only if $|\alpha_1^2 - 4\alpha_2| > 0$. Note also that the UR projection $\boldsymbol{\psi}(\boldsymbol{\alpha}) := [\psi_1(\boldsymbol{\alpha}), \psi_2(\boldsymbol{\alpha})]^T$ can be expressed as

$$\psi_1(\boldsymbol{\alpha}) = -(\zeta_1/|\zeta_1| + \zeta_2/|\zeta_2|), \quad \psi_2(\boldsymbol{\alpha}) = \zeta_1\zeta_2/|\zeta_1\zeta_2|$$

for any $\boldsymbol{\alpha} \in \mathfrak{C}^2$ such that $\alpha_2 = \zeta_1\zeta_2 \neq 0$.

Lemma 2 *Let $\mathcal{B}$ be a closed and bounded domain in $\mathfrak{C}^2$ such that $|\alpha_1^2 - 4\alpha_2| \geq c_1$ for all $\boldsymbol{\alpha} \in \mathcal{B}$, where $c_1 > 0$ is a constant. Let $\mathcal{D}$ be a subset of $\mathcal{B}$ such that $|\alpha_2| \geq c_2$ for all $\boldsymbol{\alpha} \in \mathcal{D}$, where $c_2 > 0$ is a constant. Then, the following assertions are true.*

(a) There exist constants $\kappa_1 > 0$ and $\kappa_2 > 0$ such that $\kappa_1\|\boldsymbol{\zeta} - \boldsymbol{\zeta}'\| \leq \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| \leq \kappa_2\|\boldsymbol{\zeta} - \boldsymbol{\zeta}'\|$ for all $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathcal{B}$, where $\boldsymbol{\zeta} := \boldsymbol{\xi}^{-1}(\boldsymbol{\alpha})$ and $\boldsymbol{\zeta}' := \boldsymbol{\xi}^{-1}(\boldsymbol{\alpha}')$.

(b) For all $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathcal{A} \cap \mathcal{B}$, so that $\boldsymbol{\zeta} = \boldsymbol{\mu}(\boldsymbol{\lambda})$ and $\boldsymbol{\zeta}' = \boldsymbol{\mu}(\boldsymbol{\lambda}')$ are unit roots, one can write

$$\kappa_1\|\boldsymbol{\lambda} - \boldsymbol{\lambda}'\| \leq \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| \leq \kappa_2\|\boldsymbol{\lambda} - \boldsymbol{\lambda}'\|, \quad \boldsymbol{\alpha} - \boldsymbol{\alpha}' = \frac{\partial \boldsymbol{\phi}(\boldsymbol{\lambda}')}{\partial \boldsymbol{\lambda}^T}(\boldsymbol{\lambda} - \boldsymbol{\lambda}') + \mathcal{O}(\|\boldsymbol{\lambda} - \boldsymbol{\lambda}'\|^2),$$

where κ_1 and κ_2 are the same constants as in part (a).

(c) There exists a constant $\kappa_3 > 0$ such that $\|\boldsymbol{\psi}(\boldsymbol{\alpha}) - \boldsymbol{\psi}(\boldsymbol{\alpha}')\| \leq \kappa_3 \|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|$ for all $\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathcal{D}$.

Proof. Since $\alpha_1 = -(\zeta_1 + \zeta_2)$ and $\alpha_2 = \zeta_1\zeta_2$, it is easy to show that

$$\boldsymbol{\alpha} - \boldsymbol{\alpha}' = \mathbf{Z}(\boldsymbol{\zeta} - \boldsymbol{\zeta}')$$

where

$$\mathbf{Z} := \begin{bmatrix} -1 & -1 \\ \zeta_2 & \zeta_1' \end{bmatrix}.$$

Since $\|\mathbf{Z}\|$ is bounded above by some constant $\kappa_2 > 0$, we obtain $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\| \leq \kappa_2 \|\boldsymbol{\zeta} - \boldsymbol{\zeta}'\|$. On the other hand, since the roots can be expressed as

$$\zeta_j = \tfrac{1}{2}\big\{-\alpha_1 + |\Delta|^{1/2}\rho e^{i(j-1)\pi}\big\} \quad (j = 1, 2),$$

where $\Delta := \alpha_1^2 - 4\alpha_2$ and $\rho := (\Delta/|\Delta|)^{1/2}$, it follows that

$$2(\zeta_j - \zeta_j') = -(\alpha_1 - \alpha_1') + \big(|\Delta|^{1/2} - |\Delta'|^{1/2}\big)\rho e^{i(j-1)\pi} + (\rho - \rho')|\Delta'|^{1/2} e^{i(j-1)\pi}. \tag{E.1}$$

Note that

$$|\Delta|^{1/2} - |\Delta'|^{1/2} = \frac{|\Delta|^2 - |\Delta'|^2}{(|\Delta|^{1/2} + |\Delta'|^{1/2})\,(|\Delta| + |\Delta'|)}.$$

and

$$|\Delta|^2 = |\alpha_1|^4 - 4\alpha_1^2\alpha_2^* - 4\alpha_1^{*2}\alpha_2 + 16|\alpha_2|^2.$$

It is easy to show that

$$|\alpha_1|^4 - |\alpha_1'|^4 = (\alpha_1 - \alpha_1')(\alpha_1 + \alpha_1')\alpha_1^{*2} + (\alpha_1 - \alpha_1')^*(\alpha_1 + \alpha_1')^*\alpha_1\alpha_1'.$$

Therefore we can write $|\alpha_1|^4 - |\alpha_1'|^4 = \mathcal{O}(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|)$. By applying a similar argument to the other terms, we obtain $|\Delta|^2 - |\Delta'|^2 = \mathcal{O}(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|)$. Combining this result with the assumption that $|\Delta| \geq c$ and $|\Delta'| \geq c$ yields

$$|\Delta|^{1/2} - |\Delta'|^{1/2} = \mathcal{O}(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|).$$

Similarly, we can show that

$$\rho - \rho' = \mathcal{O}(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|).$$

Substituting these results in (E.1) yields $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}'\| = \mathcal{O}(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}\|)$, which proves part (a). Part (c) can be shown similarly by first establishing that $\|\boldsymbol{\psi}(\boldsymbol{\alpha}) - \boldsymbol{\psi}(\boldsymbol{\alpha}')\| = \mathcal{O}(\|\boldsymbol{\zeta} - \boldsymbol{\zeta}'\|)$ and then applying the assertion in part (a) that $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}'\| = \mathcal{O}(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}'\|)$.

For part (b), one can show that when $\boldsymbol{\zeta} = \boldsymbol{\mu}(\boldsymbol{\lambda})$ and $\boldsymbol{\zeta}' = \boldsymbol{\mu}(\boldsymbol{\lambda}')$, there is a Taylor series expansion

$$\boldsymbol{\zeta} - \boldsymbol{\zeta}' = \frac{\partial \boldsymbol{\mu}(\tilde{\boldsymbol{\lambda}})}{\partial \boldsymbol{\lambda}^T}(\boldsymbol{\lambda} - \boldsymbol{\lambda}') = \mathrm{diag}\left\{ie^{i\tilde{\lambda}_1}, ie^{i\tilde{\lambda}_2}\right\}(\boldsymbol{\lambda} - \boldsymbol{\lambda}'),$$

which implies $\|\boldsymbol{\zeta} - \boldsymbol{\zeta}'\| = \|\boldsymbol{\lambda} - \boldsymbol{\lambda}'\|$. By using a similar argument for $\boldsymbol{\alpha} = \boldsymbol{\phi}(\boldsymbol{\lambda})$, one can establish the second expression in part (b). The proof is thus complete. □

REFERENCES

Abatzoglou, T. J. (1985), "A fast maximum likelihood algorithm for frequency estimation of a sinusoid based on Newton's Method," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 1, pp. 77–89.

Kay, S. M. (1984), "Accurate frequency estimation at low signal-to-noise ratio," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 3, pp. 540–547.

Lewis, B. L., Kretschmer, F. F., and Shelton, W. (1986), *Aspects of Radar Signal Processing*, Norwood, MA: Artech House.

Li, T. H., and Kedem, B. (1994), "Iterative filtering for multiple frequency estimation," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1120–1132.

Li, T. H., and Song, K. S. (2001), "Asymptotic analysis of a contraction mapping algorithm for multiple frequency estimation," Technical Report M958, Statistics Department, Florida State University, Tallahassee, FL. [Online]. Available: http://www.research.ibm.com/people/t/thl.

Li, T. H., and Song, K. S. (2002), "Asymptotic analysis of a fast algorithm for efficient multiple frequency estimation," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2709–2720. Errata: vol. 49, no. 2, p. 529, 2003.

Pisarenko, V. F. (1973), "The retrieval of harmonics by linear prediction," *Geophys. Journal of the Royal Astron. Soc.*, vol. 77, pp. 347–366.

Rice, J. A., and Rosenblatt, M. (1988), “On frequency estimation,” *Biometrika*, vol. 75, no. 3, pp. 477–484.

Rife, D. C., and Boorstyn, R. R. (1976), “Multiple tone parameter estimation from discrete time observations,” *Bell System Technical Journal*, vol. 55, pp. 1389–1410.

Sakai, H. (1984), “Statistical analysis of Pisarenko’s method for sinusoidal frequency estimation,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 1, pp. 95–101.

Shaw, A. K. (1995), “Maximum likelihood estimation of multiple frequencies with constraints to guarantee unit circle roots,” *IEEE Trans. Signal Processing*, vol. 43, no. 3, pp. 796–799.

Song, K. S., and Li, T. H. (1997), “Asymptotic theory of a frequency estimator,” Technical Report M917, Statistics Department, Florida State University, Tallahassee, FL. [Online]. Available: http://www.research.ibm.com/people/t/thl.

Song, K. S., and Li, T. H. (2000), “A statistically and computationally efficient method for frequency estimation,” *Stochastic Processes and Their Applications*, vol. 86, pp. 29–47.

Starer, D., and Nehorai, A. (1992), “Newton algorithm for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise,” *IEEE Transactions on Signal Processing*, vol. 40, no. 6, pp. 1528–1534.

Stoer, J. and Bulirsch, R. (2001), *Introduction to Numerical Analysis*, 3rd Edn., New York: Springer-Verlag.

Stoica, P., Moses, R. L., Friedlander, B., and Söderström, T. (1989), “Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, pp. 378–392.

Stoica, P., and Nehorai, A. (1989), “MUSIC, maximum likelihood, and Cramer-Rao bound,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741.

Stoica, P., and Söderström, T. (1991), “Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies,” *IEEE Transactions on Signal Processing*, vol. 39, no. 8, pp. 1836–1847.

Van Hamme, H. (1991), “Maximum likelihood estimation of superimposed complex sinusoids in white Gaussian noise by reduced effort coarse search (RECS),” *IEEE Transactions on Signal Processing*, vol. 39, no. 2, pp. 536–538.

Walker, A. M. (1971), “On the estimation of a harmonic component in a time series with stationary independent residuals,” *Biometrika*, vol. 58, no. 1, pp. 21–36.

White, L. B. (1993), “An iterative method for exact maximum likelihood estimation of the parameters of a harmonic series,” *IEEE Transactions on Automatic Control*, vol. 38, no. 2, pp. 367–370.