Strong Consistency of the Contraction Mapping Method for Frequency Estimation

by T-H. Li and B. Kedem

Strong Consistency of the Contraction Mapping Method for Frequency Estimation

Ta-hsin Li and Benjamin Kedem

Systems Research Center

and Mathematics Department

University of Maryland

College Park, Maryland 20742

June 1991 [1]

Abstract

Consider the stationary process $y_t = \beta \cos(\omega_0 t + \phi) + \epsilon_t$, and a parametric filter $\mathcal{L}_\alpha$, and let $\rho(\alpha)$ be the first-order autocorrelation of the filtered process $\{\mathcal{L}_\alpha(y)_t\}$. Under a certain assumption on the filtered noise spectrum, $\rho(\alpha)$ is contractive at $\cos \omega_0$. It is shown that the sample estimate of $\rho(\alpha)$, denoted by $\hat{\rho}_n(\alpha)$ and obtained from a finite sample of length n, has with probability one a fixed point $\hat{\alpha}_n$ in a neighborhood of $\cos \omega_0$, and that the sequence of fixed points $\{\hat{\alpha}_n\}$ converges with probability one to $\cos \omega_0$. The proof is based on a general result regarding the uniform consistency of the sample autocorrelation. The developed theory is illustrated by two numerical examples pertaining to two different parametric time-invariant filters.

1 Introduction

The classical problem of frequency estimation is of interest in a wide range of engineering and scientific applications. The problem is well-formulated in the signal processing and statistics literature, and has been studied by meany researchers. Recently a general iterative filtering approach, called the contraction mapping (CM) method, was suggested by He and Kedem [6], Yakowitz [18], and Kedem [10] for estimating the frequency of a single sinusoid in additive noise from a finite sample $\{y_0, y_1, \ldots, y_{n-1}\}$ obtained from a process of the form

$$y_t = \beta \cos(\omega_0 t + \phi) + \epsilon_t. \tag{1.1}$$

Here $\beta > 0$ and $\omega_0 \in (0, \pi)$ are constants, ϕ is uniformly distributed on $[0, 2\pi)$, i.e., $\phi \sim U[0, 2\pi)$, and $\{\epsilon_t\}$ is a zero-mean stationary process, independent of ϕ, with spectral distribution function $F(\omega)$ which is continuous at ω_0. The gist of the CM method is as follows. Using a parametric filter $\mathcal{L}_\alpha$ that satisfies the so-called fundamental property

$$\mathrm{Corr}(\mathcal{L}_\alpha(\epsilon)_{t+1}, \mathcal{L}_\alpha(\epsilon)_t) = \alpha, \tag{1.2}$$

a sequence of estimators is constructed by an iterative procedure of the form

$$\alpha_j = \hat{\rho}_n(\alpha_{j-1}) \tag{1.3}$$

where $\hat{\rho}_n(\alpha)$ is an estimator of the first-order autocorrelation of $\{\mathcal{L}_\alpha(y)_t\}$. This procedure determines a fixed-point of the mapping $\hat{\rho}_n(\alpha)$. As will be shown, the sequence of the fixed-points converges to $\cos \omega_0$ as $n \to \infty$.

In this paper we provide an asymptotic analysis of the CM method, focusing on strong (almost sure) consistency. We shall discuss, for a given *finite sample*, the existence of a fixed-point of $\hat{\rho}_n(\alpha)$, which will be referred to as the CM estimator, the convergence of some iterative procedures for finding the fixed-point, and finally, the consistency of the fixed-point as the sample size tends to infinity. We shall show, under appropriate conditions, that the existence, convergence, and consistency of the CM estimator can be established *almost surely* for sufficiently large sample size, provided the parametric filter satisfies the fundamental property (1.2), in addition to some fairly mild conditions.

The fundamental property (1.2) required by the CM method is exhibited by many parametric filters, while many more can be reparametrized so as to satisfy the property (see, for

1

example, [12]). An important special case is the AR(2) filter discussed by Quinn and Fernandes [16] for a special case, and considered in more generality in this paper (see Section 7.2). For their special AR(2) filter, one can check that the fundamental property holds in a limiting sense if the noise has a sufficiently smooth spectral density. Using a variant of the iterative procedure (1.3), Quinn and Fernandes showed that with their AR(2) filter a sequence of estimators can be produced that converges to the unknown frequency almost surely, and that the precision of the estimator is the same as that achieved by the nonlinear least squares (or the maximum likelihood estimator if $\{\epsilon_t\}$ is Gaussian white noise). The main contribution of the present paper is the proof that the CM method is strongly consistent for a much wider class of parametric filters in addition to the special case determined by the AR(2) filter. This, coupled with the Quinn-Fernandes result, shows that the CM method can produce asymptotically efficient estimates.

One of the advantages of the CM method is its computational simplicity. It bypasses time-consuming nonlinear optimization routines needed for the nonlinear least squares method. Some related iterative filtering methods for frequency estimation have been suggested in the literature, of which we mention the work of Kay [9], Kumaresan, Scharf, and Shaw [13], and Dragošević and S. S. Stanković [4].

2 The CM Method

Let the random sequence $\{y_t\}$ be given by (1.1). Consider a parametric linear time-invariant causal filter $\mathcal{L}_\alpha$, indexed by $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, with real-valued impulse response sequence $\{h_j(\alpha)\}_{j=0}^\infty$, where $\underline{\alpha}$ and $\overline{\alpha}$ are constants such that $-1 < \underline{\alpha} < \cos\omega_0 < \overline{\alpha} < 1$. Let $H(\omega; \alpha)$ be the transfer function of $\mathcal{L}_\alpha$ defined by

$$H(\omega; \alpha) := \sum_{j=0}^\infty h_j(\alpha)e^{-ij\omega}.$$

It is easy to see that $H(\omega; \alpha) = \overline{H(-\omega; \alpha)}$, where the overline denotes the complex conjugate operation.

Applying the filter $\mathcal{L}_\alpha$ to $\{y_t\}$ and $\{\epsilon_t\}$ yields the filtered data $\{y_t(\alpha)\}$ and the filtered noise

2

$\{\epsilon_t(\alpha)\}$ defined, respectively, by

$$y_t(\alpha) := \sum_{j=0}^{\infty} h_j(\alpha) \, y_{t-j} \quad \text{and} \quad \epsilon_t(\alpha) := \sum_{j=0}^{\infty} h_j(\alpha) \, \epsilon_{t-j}. \tag{2.1}$$

Let $\rho(\alpha)$ be the first-order autocorrelation of $\{y_t(\alpha)\}$, then the spectral representation of the autocorrelation function gives

$$\rho(\alpha) = \frac{\sigma^2 |H(\omega_0; \alpha)|^2 \cos \omega_0 + \displaystyle\int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \cos \omega \, dF(\omega)}{\sigma^2 |H(\omega_0; \alpha)|^2 + \displaystyle\int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \, dF(\omega)} \tag{2.2}$$

where $\sigma^2 := \beta^2/2$ is the variance of the signal. For convenience, it is always assumed that $\int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \, dF(\omega) = 0$ implies $|H(\omega_0; \alpha)| = 0$, which means that the noise cannot be completely removed without filtering out the signal.

In the sequel, we always assume that for any $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, the filter $\mathcal{L}_\alpha$ satisfies the so-called "fundamental property"

$$\alpha = \rho_\epsilon(\alpha) := \frac{E\{\epsilon_{t+1}(\alpha) \, \epsilon_t(\alpha)\}}{E\{\epsilon_t^2(\alpha)\}}, \tag{2.3}$$

that is,

$$\alpha = \frac{\displaystyle\int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \cos \omega \, dF(\omega)}{\displaystyle\int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \, dF(\omega)}$$

where $\rho_\epsilon(\alpha)$ stands for the first-order autocorrelation of $\{\epsilon_t(\alpha)\}$. Under this assumption, (2.2) reduces to

$$\rho(\alpha) = \alpha^* + C(\alpha) \, (\alpha - \alpha^*) \tag{2.4}$$

where $\alpha^* := \cos \omega_0$, and

$$C(\alpha) := \frac{1}{1 + \gamma(\alpha)}$$

with $\gamma(\alpha)$ being the signal-to-noise ratio of the filtered data $\{y_t(\alpha)\}$ defined by

$$\gamma(\alpha) := \frac{\sigma^2 |H(\omega_0; \alpha)|^2}{\displaystyle\int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \, dF(\omega)}.$$

Clearly, for all $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, $0 < C(\alpha) \leq 1$ with $C(\alpha) = 1$ if and only if $|H(\omega_0; \alpha)| = 0$, or if and only if the filter $\mathcal{L}_\alpha$ does not capture the frequency.

3

The original idea of utilizing (2.4) to estimate α^* was proposed by He and Kedem [6] for a specific filter, known as the α-filter, and extended by Yakowitz [18] to any parametric filter satisfying (2.3). To obtain an estimator of α^*, they employed the iterative procedure

$$\hat{\alpha}_n^{(m)} := \hat{\rho}_n(\hat{\alpha}_n^{(m-1)}) \qquad m = 1, 2, \ldots \tag{2.5}$$

where $\hat{\rho}_n(\alpha)$ is an estimator of $\rho(\alpha)$ on a finite sample of size n. In the numerical analysis literature, this procedure is known as the fixed-point iteration (FPI) (see, for example, [17]) which is used to find a fixed-point of $\hat{\rho}_n(\alpha)$, if exists. The original motivation of using this iterative procedure to estimate α^* was based on the heuristic argument that in the limiting case as the sample size n tends to infinity, the limit of $\hat{\alpha}_n^{(m)}$, denoted by $\hat{\alpha}^{(m)}$, satisfies the equation

$$\hat{\alpha}^{(m)} - \alpha^* = C(\hat{\alpha}^{(m-1)})(\hat{\alpha}^{(m-1)} - \alpha^*)$$

As can be seen, the error in $\hat{\alpha}^{(m)}$ is *contracted* by an amount of $C(\hat{\alpha}^{(m-1)})$ as compared with that in $\hat{\alpha}^{(m-1)}$, and hence the name of contraction mapping (CM) method. It is readily seen that as m tends to infinity $\hat{\alpha}^{(m)}$ converges to α^* monotonically, provided $C(\alpha)$ is uniformly bounded above by some constant which is strictly less than one. This argument, however, does not lead to the conclusion that for a *fixed* sample size n, the sequence $\hat{\alpha}_n^{(m)}$ would converge as m (not n) tends to infinity, not to mention convergence to α^*. This is the problem we are going to deal with in the sequel.

3 Another Look at the CM Method

A close examination of (2.4) reveals that if we define $G(\alpha) := 1 - C(\alpha)$, i.e.,

$$G(\alpha) = \frac{\gamma(\alpha)}{1 + \gamma(\alpha)},$$

then (2.4) can be rewritten as

$$\alpha - \rho(\alpha) = G(\alpha)(\alpha - \alpha^*). \tag{3.1}$$

From this equation, it becomes quite clear that if $G(\alpha) > 0$ (or, equivalently, $\gamma(\alpha) > 0$, i.e., the filter $\mathcal{L}_\alpha$ captures the signal) for all α in a neighborhood of α^*, then α^* would be *the*

4

unique fixed-point of $\rho(\alpha)$ in that neighborhood. This observation leads to the somewhat more general idea of estimating ω_0 by *finding the fixed-point* of $\hat{\rho}_n(\alpha)$ in a neighborhood of α^*. The fixed-point of $\hat{\rho}_n(\alpha)$ will be refrred to as the CM estimator. As will be seen later, if taken to be the first-order sample autocorrelation, $\hat{\rho}_n(\alpha)$ (and in fact variants thereof) does form a contraction mapping, not in the entire interval $[\underline{\alpha}, \overline{\alpha}]$, but in a neighborhood of α^*, provided $C(\alpha^*) < 1$. Therefore, we retain the use of the term "contraction mapping (CM) method" for *any* procedure that finds a fixed-point of $\hat{\rho}_n(\alpha)$, or, equivalently, a zero of the function

$$f_n(\alpha) := \alpha - \hat{\rho}_n(\alpha).$$

Appenrently, (2.5) is only a special procedure of this kind. In fact, one can used any algorithms available in the numerical analysis literature to find the zero of $f_n(\alpha)$. For instance, $\{\hat{\alpha}_n^{(m)}\}$ can be produced by the secant method [17]

$$\hat{\alpha}_n^{(m)} := \hat{\alpha}_n^{(m-1)} - \frac{\hat{\alpha}_n^{(m-1)} - \hat{\alpha}_n^{(m-2)}}{f_n(\hat{\alpha}_n^{(m-1)}) - f_n(\hat{\alpha}_n^{(m-2)})} f_n(\hat{\alpha}_n^{(m-1)}) \qquad m = 1, 2 \ldots \qquad (3.2)$$

which, as will be seen later, converges faster than (2.5) under appropriate conditions.

4 Existence, Convergence, and Consistency

Based on a finite sample $\{y_0(\alpha), y_1(\alpha), \ldots, y_{n-1}(\alpha)\}$, it is alway possible to construct an estimator $\hat{\rho}_n(\alpha)$ as a function of α (for instance, by taking $\hat{\rho}_n(\alpha)$ to be the sample autocorrelation of the filtered data). The following questions are of interest:

1) Whether $\hat{\rho}_n(\alpha)$ has a fixed-point in a neighborhood of α^*;

2) If it does, under what conditions these iterative algorithms converge to the fixed-point;

3) Whether the fixed-point is consistent for estimating α^* as n tends to infinity.

In this section, we would like to answer these questions one by one.

It is worth noting that the almost sure convergence of the CM method has been recently proved in [12] for bandpass filters whose bandwidth shrinks at a certain rate during the iteration. In [12], the zero-crossing rate (ZCR) was used in connection with the Gaussian assumption, resulting in a scheme of the form (1.3) with $\hat{\rho}_n(\alpha)$ replaced by cosine of the *asymptotic* (i.e., $n = \infty$) ZRC of the filtered data.

5

4.1 Existence of the CM Estimator

First of all, let us investigate conditions under which $\hat{\rho}_n(\alpha)$ has a fixed-point in a neighborhood of a specific α_0. For this purpose, we have

Lemma 4.1 *Let α_0 be any fixed number in $[\underline{\alpha}, \overline{\alpha}]$. Suppose that there exist constants K and δ with $0 < K < 1$ and $\delta > 0$ such that $S_\delta(\alpha_0) := \{\alpha : |\alpha - \alpha_0| < \delta\} \subset [\underline{\alpha}, \overline{\alpha}]$ and that with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n) $\hat{\rho}_n(\alpha)$ satisfies*

(a) $|\hat{\rho}_n(\alpha') - \hat{\rho}_n(\alpha'')| \leq K|\alpha' - \alpha''|, \quad \forall \; \alpha', \alpha'' \in \bar{S}_\delta(\alpha_0) := \{\alpha : |\alpha - \alpha_0| \leq \delta\};$

(b) $|\alpha_0 - \hat{\rho}_n(\alpha_0)| \leq (1 - K)\delta.$

Then $\hat{\rho}_n(\alpha)$ has a unique fixed-point in $\bar{S}_\delta(\alpha_0)$ with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n).

<u>PROOF</u>. The assertions follow directly from Theorem 5.2.3 in [17]. $\Diamond$

<u>REMARK 3.1</u> Conditions in Lemma 4.1 can be relaxed considerablely by allowing K and δ to be random variables. In other words, the conclusion in Lemma 4.1 remains valid if conditions (a) and (b) hold with probability tending to one (or with probability one for large n) for some random variables K and δ satisfying $0 < K < 1$ and $\delta > 0$ with probability one. If this is the case, $\bar{S}_\delta(\alpha_0)$ is apparently a random interval. $\Diamond$

<u>REMARK 3.2</u> Under conditions (a) and (b), $\hat{\rho}_n(\alpha)$ becomes a *contraction mapping* on $\bar{S}_\delta(\alpha_0)$. In fact, the contractivity is readily seen from (a). Moreover, combining (a) and (b) gives

$$
\begin{aligned}
|\hat{\rho}_n(\alpha) - \alpha_0| &\leq \; |\hat{\rho}_n(\alpha) - \hat{\rho}_n(\alpha_0)| + |\hat{\rho}_n(\alpha_0) - \alpha_0| \\
&\leq \; K\delta + (1 - K)\delta = \delta
\end{aligned}
$$

for all $\alpha \in \bar{S}_\delta(\alpha_0)$, that is, $\hat{\rho}_n(\alpha)$ maps $\bar{S}_\delta(\alpha_0)$ onto itself. $\Diamond$

According to Lemma 4.1, the existence of a unique fixed-point of $\hat{\rho}_n(\alpha)$ in a neighborhood of α_0 is guaranteed by conditions (a) and (b). An ideal candidate for α_0 in our problem is obviously α^*. Let $e_n(\alpha)$ be the error of the estimator $\hat{\rho}_n(\alpha)$ for estimating $\rho(\alpha)$, i.e., $e_n(\alpha) := \hat{\rho}_n(\alpha) - \rho(\alpha)$. With this notation, $\hat{\rho}_n(\alpha)$ can be written as

$$
\hat{\rho}_n(\alpha) = \rho(\alpha) + e_n(\alpha).
$$

For the existence of a unique fixed-point of $\hat{\rho}_n(\alpha)$ in a neighborhood of α^*, i.e., the existence of the CM estimator, we have the following theorem.

Theorem 4.1 *Suppose that on $\bar{S}_\Delta(\alpha^*) \subset [\underline{\alpha}, \overline{\alpha}]$, $\rho(\alpha)$ is continuously differentiable and $\hat{\rho}_n(\alpha)$ is differentiable with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n). Assume further that $C(\alpha)$ is continuous at α^* with $C(\alpha^*) < 1$. If $\hat{\rho}_n(\alpha)$ is uniformly consistent on $\bar{S}_\Delta(\alpha^*)$ up to the first derivative, i.e., if*

$$\lim_{n \to \infty} \sup_{\alpha \in \bar{S}_\Delta(\alpha^*)} |e_n(\alpha)| = 0 \tag{4.1}$$

$$\lim_{n \to \infty} \sup_{\alpha \in \bar{S}_\Delta(\alpha^*)} |e_n'(\alpha)| = 0 \tag{4.2}$$

in probability (or with probability one), then there exists $0 < \delta \leq \Delta$ such that $\hat{\rho}_n(\alpha)$ has a unique fixed-point in $S_\delta(\alpha^*)$ with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n).

PROOF. Consider the theoretical function $\rho(\alpha)$. The continuity of $C(\alpha)$ at α^* implies that from (2.4), the derivative of $\rho(\alpha)$ at α^* can be written as

$$\begin{aligned} \rho'(\alpha^*) &= \lim_{\alpha \to \alpha^*} \frac{\rho(\alpha) - \rho(\alpha^*)}{\alpha - \alpha^*} \\ &= \lim_{\alpha \to \alpha^*} \frac{\rho(\alpha) - \alpha^*}{\alpha - \alpha^*} \\ &= \lim_{\alpha \to \alpha^*} C(\alpha) = C(\alpha^*). \end{aligned}$$

Since $0 < C(\alpha^*) < 1$, i.e, $\gamma(\alpha^*) > 0$ so that $\mathcal{L}_\alpha$ captures the frequency with $\alpha = \alpha^*$, and since $\rho'(\alpha)$ is continuous, then there exists $0 < \delta \leq \Delta$ such that

$$M := \sup_{\alpha \in \bar{S}_\delta(\alpha^*)} \rho'(\alpha) < 1 \tag{4.3}$$

$$m := \inf_{\alpha \in \bar{S}_\delta(\alpha^*)} \rho'(\alpha) > 0. \tag{4.4}$$

For any $0 < \delta' < \delta$, assumption (4.2), together with (4.3) and (4.4), implies that with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n) there exists K such that

$$0 < m + e_n'(\alpha) \leq \hat{\rho}_n'(\alpha) \leq M + e_n'(\alpha) \leq K < 1 \tag{4.5}$$

7

for all $\alpha \in \bar{S}_{\delta'}(\alpha^*)$. By the mean-value theorem, condition (a) in Lemma 4.1 is guaranteed with δ' in place of δ. Moreover, from (4.1),

$$|\alpha^* - \hat{\rho}_n(\alpha^*)| = |\rho(\alpha^*) - \hat{\rho}_n(\alpha^*)| = |e_n(\alpha^*)| \le (1 - K)\delta' \tag{4.6}$$

with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n), which gives condition (b) in Lemma 4.1. Therefore, by Lemma 4.1, $\hat{\rho}_n(\alpha)$ has a unique fixed-point on $\bar{S}_{\delta'}(\alpha^*)$. Since δ' is arbitrary, this implies that $\hat{\rho}_n(\alpha)$ has a unique fixed-point in the interior of $S_\delta(\alpha^*)$. $\diamond$

<u>Remark 4.3</u> Theorem 4.1 still holds if the continuity of $\rho'(\alpha)$ is replaced by

$$|\rho'(\alpha)| \le M < 1 \qquad \forall\, \alpha \in \bar{S}_\Delta(\alpha^*). \tag{4.7}$$

However, this weaker condition does not guarantee the positivity of $\hat{\rho}'_n(\alpha)$. $\diamond$

<u>Remark 4.4</u> As in Lemma 4.1, inequalities (4.5) and (4.6) imply that $\hat{\rho}_n(\alpha)$ constitutes a contraction mapping in $S_\delta(\alpha^*)$ with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n). It is made possible basically by the requirement that the $\hat{\rho}_n(\alpha)$ be uniformly consistent up to the first derivative and that the filter $\mathcal{L}_\alpha$ passes the frequency for all α in the vicinity of α^*. $\diamond$

4.2 Convergence of Two Iterative Algorithms

Let $\hat{\alpha}_n$ be the CM estimator, i.e., the fixed-point of $\hat{\rho}_n(\alpha)$ in $S_\delta(\alpha^*)$ given by Theorem 4.1. The following theorem guarantees that under suitable conditions, the FPI procedure (2.5) can start anywhere in a neighborhood of $\hat{\alpha}_n$ so as to converge to $\hat{\alpha}_n$.

Theorem 4.2 *Under the conditions in Theorem 4.1, there exist constants $\delta_0 > 0$ and $0 < K < 1$ such that with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n) the sequence $\{\hat{\alpha}_n^{(m)}\}$ defined by (2.5) stays in $S_{\delta_0}(\hat{\alpha}_n)$ and converges at least linearly to $\hat{\alpha}_n$ as m tends to infinity, provided $\hat{\alpha}_n^{(0)} \in S_{\delta_0}(\hat{\alpha}_n)$. Moreover, the convergence is monotone and*

$$|\hat{\alpha}_n^{(m)} - \hat{\alpha}_n| \le K|\hat{\alpha}_n^{(m-1)} - \hat{\alpha}_n| \tag{4.8}$$

for any $m \ge 1$.

PROOF. Since $\hat{\alpha}_n \in S_\delta(\alpha^*)$, there exists $\delta_0 > 0$ such that $S_{\delta_0}(\hat{\alpha}_n) \subset S_\delta(\alpha^*)$. From the proof of Theorem 4.1, a constant K can be found so that $0 < K < 1$ and (4.5) holds for all $\alpha \in S_{\delta_0}(\hat{\alpha}_n)$. According to the mean-value theorem, this implies that condition (a) of Lemma 4.1 holds for all α' and α'' in $S_{\delta_0}(\hat{\alpha}_n)$. In particular,

$$|\hat{\rho}_n(\alpha) - \hat{\alpha}_n| \le K|\alpha - \hat{\alpha}_n| \qquad \forall \ \alpha \in S_{\delta_0}(\hat{\alpha}_n).$$

Therefore, $\hat{\alpha}_n^{(m)} \in S_{\delta_0}(\hat{\alpha}_n)$ for any $m \ge 1$, provided $\hat{\alpha}_n^{(0)} \in S_{\delta_0}(\hat{\alpha}_n)$. The inequality (4.8) and thus the convergence of $\{\hat{\alpha}_n^{(m)}\}$ are consequences of condition (a) in Lemma 4.1 with $S_{\delta_0}(\hat{\alpha}_n)$ in place of $\bar{S}_\delta(\alpha_0)$. The monotonicity is due to the fact that $\hat{\rho}'_n(\alpha)$ is positive in $S_{\delta_0}(\hat{\alpha}_n)$. $\diamondsuit$

REMARK 4.5 Theorem 4.2 remains valid if K and δ_0 are allowed to be random. And if the continuity of $\rho'(\alpha)$ is replaced by (4.7), the convergence of $\{\hat{\alpha}_n^{(m)}\}$ still holds but the monotonicity is not guaranteed any more. $\diamondsuit$

It is evident from (4.3), (4.5), (4.6), and (4.8) that the rate of convergence of the iterative procedure (2.5) is governed by the constant $C(\alpha^*)$ (known as the contraction coefficient) and the estimation accuracy of $\hat{\rho}_n(\alpha)$ up to the first derivative. Usually, the estimation accuracy depends heavily on the sample size n which one can not control. What one can do to accelerate the convergence is to make $C(\alpha^*)$ as small as possible, by using appropriate filters during the iteration. Since decreasing $C(\alpha^*)$ is equivalent to increasing $\gamma(\alpha^*)$, the signal-to-noise ratio after filtering with $\mathcal{L}_{\alpha^*}$, the convergence would be accelerated if the signal could be enhanced in an appropriate way. Since α^* is unknown, the only possibility of enhancing the signal during the iteration relies on $\hat{\alpha}_n^{(m)}$. Some strategies were discussed in [10, 12] upon shrinking the bandwidth of filters.

Now let us consider the secant method defined by (3.2). Under proper conditions, this method has superlinear convergence. To study its convergence, the second derivatives of $\rho(\alpha)$ and $e_n(\alpha)$ are required.

Theorem 4.3 *Under the conditions in Theorem 4.1, if in addition there exists $0 < \delta_0 < \delta$ such that $\rho(\alpha)$ and $e_n(\alpha)$ have second derivatives that are uniformly bounded by D on $\bar{S}_{\delta_0}(\hat{\alpha}_n)$ with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n), then, starting with $\hat{\alpha}_n^{(-1)}$, $\hat{\alpha}_n^{(0)} \in S_{\delta_1}(\hat{\alpha}_n)$ where $\delta_1 := \min\{(1 - K)/D, \delta_0\}$ and K is given by (4.5), the sequence $\{\hat{\alpha}_n^{(m)}\}$ generated by the secant method (3.2) converges at least superlinearly to $\hat{\alpha}_n$*

with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n), and for some $c > 0$ and $0 < \lambda < 1$,

$$|\hat{\alpha}_n^{(m)} - \hat{\alpha}_n| \le c\,\lambda^{p(m)} \tag{4.9}$$

where

$$p(m) := \left(\frac{1 + \sqrt{5}}{2}\right)^m.$$

PROOF. Note that (4.5) and the boundedness of second derivatives imply

$$\left|\frac{f_n''(\alpha)}{2f_n'(\alpha)}\right| = \left|\frac{\rho''(\alpha) + e_n''(\alpha)}{2(1 - \hat{\rho}_n'(\alpha))}\right| \le \frac{D}{1 - K}$$

for all $\alpha \in S_{\delta_1}(\hat{\alpha}_n)$. The remaining proof follows the argument in [17] (pp.292–293). ◇

REMARK 4.6 The secant method may require better initial estimates than the FPI procedure to assure its convergence. Moreover, the secant method is numerically not as stable as the FPI, although it converges faster with proper initial estimates. ◇

4.3 Consistency of the CM Estimator

Suppose that in a neighborhood $S_\delta(\alpha^*)$ of α^*, the function $\hat{\rho}_n(\alpha)$ has a fixed-point $\hat{\alpha}_n$ with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n). We are interested in the consistency of $\hat{\alpha}_n$, referred to as the CM estimators, as n tends to infinity. For this purpose, we notice that from (3.1) we obtain

$$G(\hat{\alpha}_n)\,(\hat{\alpha}_n - \alpha^*) = e_n(\hat{\alpha}_n) \tag{4.10}$$

with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n). Clearly, the behavior of $\hat{\alpha}_n$ depends entirely on that of $G(\alpha)$ (known as the gain coefficient) and of $e_n(\alpha)$ in $S_\delta(\alpha^*)$. For the consistency of $\hat{\alpha}_n$, we have the following results.

Theorem 4.4 *Let $\hat{\alpha}_n$ be the fixed-point of $\hat{\rho}_n(\alpha)$ in $S_\delta(\alpha^*)$. Suppose that $e_n(\alpha)$ satisfies (4.1) in probability (or with probability one) and that $G(\alpha) \ge g$ for some $g > 0$ and all $\alpha \in \bar{S}_\delta(\alpha^*)$. Then $\hat{\alpha}_n$ converges to α^* in probability (or with probability one) as $n \to \infty$. In both cases, the convergence is also in mean-square.*

PROOF. The convergence in probability (or with probability one) follows immediately from (4.1) and (4.10). The mean-square convergence is due to the boundedness of $\hat{\alpha}_n$. ◇

5 Consistency of Sample Autocorrelation

As seen in the proceeding section, the consistency assumption of (4.1)-(4.2) plays an important role in the proof of existence, convergence, and consistency of the CM estimators. To verify this assumption, we specialize in this section to the usual sample autocorrelation $\hat{\rho}_n(\alpha)$, as will be defined later, and investigate its consistency when the filter $\mathcal{L}_\alpha$ satisfies the following conditions:

(H1) $\{h_j\} \neq \{0\}$ and $h_j(\alpha) = 0$ for $j < 0$.

(H2) There exist constants $a_j > 0$ such that

$$\sum_{j=0}^{\infty} j a_j < \infty \quad \text{and} \quad |h_j(\alpha)| \leq a_j$$

for all $j = 0, 1, \ldots$ and all $\alpha \in A$, where A is a closed subset of $[\underline{\alpha}, \overline{\alpha}]$.

(H3) $h_j'(\alpha)$ exists and there are constants $b_j > 0$ such that

$$\sum_{j=0}^{\infty} j b_j < \infty \quad \text{and} \quad |h_j'(\alpha)| \leq b_j$$

for all $j = 0, 1, \ldots$ and all $\alpha \in A$.

These conditions can be easily fulfilled by many commonly-used filters. Some examples will be given in the next section.

Given a *finite* sample $\{y_0, y_1, \ldots, y_{n-1}\}$ of size n, let $\{\hat{y}_t(\alpha)\}$ be the filtered data defined by

$$\hat{y}_t(\alpha) := \sum_{j=0}^{t} h_j(\alpha)\, y_{t-j} \qquad t = 0, 1, \ldots, n-1. \tag{5.1}$$

On the basis of $\{\hat{y}_t(\alpha)\}$, a widely-used estimator of the first-order autocorrelation $\rho(\alpha)$ is the least squares (LS) estimator that minimizes $\sum_{t=1}^{n-1}[\hat{y}_t(\alpha) - \rho\, \hat{y}_{t-1}(\alpha)]^2$, i.e.,

$$\hat{\rho}_n(\alpha) := \frac{\hat{r}_1(\alpha)}{\hat{r}_0(\alpha)} \tag{5.2}$$

where $\hat{r}_0(\alpha)$ and $\hat{r}_1(\alpha)$ are sample variance and first-order autocovariance of $\{\hat{y}_t(\alpha)\}$ defined by

$$\hat{r}_0(\alpha) \quad := \quad \frac{1}{n} \sum_{t=1}^{n-1} \hat{y}_{t-1}^2(\alpha) \tag{5.3}$$

$$\hat{r}_1(\alpha) \quad := \quad \frac{1}{n} \sum_{t=1}^{n-1} \hat{y}_t(\alpha)\, \hat{y}_{t-1}(\alpha). \tag{5.4}$$

11

We would like to show that if (H1)–(H3) are satisfied in a neighborhood of α^*, then (4.1) and (4.2) hold with probability one, provided $\{\epsilon_t\}$ is a linear process, i.e.,

$$\epsilon_t = \sum_{j=-\infty}^{\infty} \psi_j z_{t-j} \tag{5.5}$$

where $\{z_t\}$ are IID$(0,1)$ and $\sum |\psi_j| < \infty$. In this case, the noise spectral distribution function F is given by

$$F(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\omega} \left| \sum \psi_j e^{-ij\lambda} \right|^2 d\lambda$$

where $i := \sqrt{-1}$. Before stating this result, the following lemmas are needed.

Lemma 5.1 *Let $\{\epsilon_t\}$ be a linear process defined by (5.5). Suppose that $\{g_j(\alpha)\}$ and $\{h_j(\alpha)\}$ satisfy assumptions (H1) and (H2). Then as $n \to \infty$,*

$$n^{-1} \sum_{t=0}^{n-\tau-1} \left(\sum_{j=0}^{t+\tau} h_j(\alpha) y_{t-j+\tau} \right) \left(\sum_{j=0}^{t} g_j(\alpha) y_{t-j} \right) \tag{5.6}$$

$$\stackrel{a.s.}{\to} \sigma^2 \, \Re \left\{ H(\omega_0; \alpha) \overline{G(\omega_0; \alpha)} e^{i\tau\omega_0} \right\} + \int_{-\pi}^{\pi} H(\omega; \alpha) \overline{G(\omega; \alpha)} e^{i\tau\omega} \, dF(\omega)$$

uniformly in $\alpha \in A$, where $\tau \geq 0$,

$$G(\omega; \alpha) := \sum_{j=0}^{\infty} g_j(\alpha) e^{-ij\omega},$$

and $\Re\{\cdot\}$ stands for the real part of a complex number.

Proof. See Appendix A. $\Diamond$

Remark 6.1 Lemma 5.1 can be generalized to the case of multiple sinusoids in noise. In this case, the observation $\{y_t\}$ is given by

$$y_t = \sum_{k=0}^{q-1} \beta_k \cos(\omega_k t + \phi_k) + \epsilon_t \tag{5.7}$$

where $q \geq 1$, $\beta_k > 0$ and $0 < \omega_0 < \cdots < \omega_{q-1} < \pi$ are constants, $\{\phi_k\}$ are iid $U[0, 2\pi)$ and independent of $\{\epsilon_t\}$. Under the same conditions as Lemma 5.1, it can be shown that as n tends to infinity,

$$n^{-1} \sum_{t=0}^{n-\tau-1} \left(\sum_{j=0}^{t+\tau} h_j(\alpha) y_{t-j+\tau} \right) \left(\sum_{j=0}^{t} g_j(\alpha) y_{t-j} \right) \tag{5.8}$$

$$\stackrel{a.s.}{\to} \sum_{k=0}^{q-1} \sigma_k^2 \, \Re \left\{ H(\omega_k; \alpha) \overline{G(\omega_k; \alpha)} e^{i\tau\omega_k} \right\} + \int_{-\pi}^{\pi} H(\omega; \alpha) \overline{G(\omega; \alpha)} e^{i\tau\omega} \, dF(\omega)$$

12

uniformly in $\alpha \in A$, where $\sigma_k^2 := \beta_k^2/2$ is the variance of the kth sinusoid. In fact, using the same method, the counterparts of $n^{-1} \sum_{t=0}^{n-\tau-1} I_2(t)$ and $n^{-1} \sum_{t=0}^{n-\tau-1} I_3(t)$ in the proof of Lemma 5.1 can be shown to have the same limits given by (A.6) and (A.7), and in the counterpart of $n^{-1} \sum_{t=0}^{n-\tau-1} I_1(t)$ the cross-product terms with different frequencies converge to zero as $n \to \infty$ since

$$\sum_{t=0}^{n-1} \cos \omega_k t \cos \omega_{k'} t = O(1), \quad \sum_{t=0}^{n-1} \sin \omega_k t \sin \omega_{k'} t = O(1)$$

for $k \neq k'$ and

$$\sum_{t=0}^{n-1} \sin \omega_k t \cos \omega_{k'} t = O(1)$$

for any k and k'. $\diamond$

REMARK 6.2 In the proof of Lemma 5.1, the assumption that $\phi \sim \mathrm{U}[0, 2\pi)$ is not necessary. It is required only if we want $\{y_t\}$ to be stationary. This remark also applies to the case of multiple sinusoids. $\diamond$

Denote by $r_0(\alpha)$ and $r_1(\alpha)$ the variance and the covariance of the filtered data $\{y_t(\alpha)\}$ defined by (2.1), respectively. Then, we have

$$
\begin{aligned}
r_0(\alpha) &= \sigma^2 |H(\omega_0; \alpha)|^2 + \int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \, dF(\omega) \\
&= \sigma^2 \left| \sum h_j(\alpha) e^{-ij\omega} \right|^2 + \sum \sum h_j(\alpha) h_k(\alpha) r_{k-j}^\epsilon \\
r_1(\alpha) &= \sigma^2 |H(\omega_0; \alpha)|^2 \cos \omega_0 + \int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \cos \omega \, dF(\omega) \\
&= \sigma^2 \left| \sum h_j(\alpha) e^{-ij\omega} \right|^2 \cos \omega_0 + \sum \sum h_j(\alpha) h_k(\alpha) r_{k-j+1}^\epsilon.
\end{aligned}
$$

It is readily seen that under assumptions (H2) and (H3), $r_0(\alpha)$ and $r_1(\alpha)$ are differentiable with respect to α and their derivatives, denoted by $r_0'(\alpha)$ and $r_1'(\alpha)$, respectively, can be easily shown to be

$$
\begin{aligned}
r_0'(\alpha) &= 2\sigma^2 \, \Re \left\{ H'(\omega_0; \alpha) \overline{H(\omega_0; \alpha)} \right\} + 2 \int_{-\pi}^{\pi} H'(\omega; \alpha) \overline{H(\omega; \alpha)} \, dF(\omega) \\
r_1'(\alpha) &= 2\sigma^2 \, \Re \left\{ H'(\omega_0; \alpha) \overline{H(\omega_0; \alpha)} \right\} \cos \omega_0 + 2 \int_{-\pi}^{\pi} H'(\omega; \alpha) \overline{H(\omega; \alpha)} \cos \omega \, dF(\omega)
\end{aligned}
$$

where

$$H'(\omega; \alpha) := \sum h_j'(\alpha) e^{-ij\omega}.$$

13

Based on Lemma 5.1, we are now able to obtain consistency of the sample variance $\hat{r}_0(\alpha)$, the sample covariance $\hat{r}_1(\alpha)$, and their derivatives, as estimators of $r_0(\alpha)$, $r_1(\alpha)$, and their derivatives. For this purpose, we have

Lemma 5.2 *Let* $\{\epsilon_t\}$ *be a linear process defined by* (5.5). *Suppose that* $\{h_j(\alpha)\}$ *satisfies* (H1)–(H3). *Then as n tends to infinity,*

$$\hat{r}_0(\alpha) \overset{a.s.}{\to} r_0(\alpha), \quad \hat{r}_1(\alpha) \overset{a.s.}{\to} r_1(\alpha)$$

$$\hat{r}_0'(\alpha) \overset{a.s.}{\to} r_0'(\alpha), \quad and \quad \hat{r}_1'(\alpha) \overset{a.s.}{\to} r_1'(\alpha)$$

uniformly in $\alpha \in A$.

PROOF. See Appendix B. $\Diamond$

Using these lemmas, we are able to obtain the uniform strong consistency of $\hat{\rho}_n(\alpha)$ and $\hat{\rho}_n'(\alpha)$ as follows.

Theorem 5.1 *Let* $\{\epsilon_t\}$ *be a linear process defined by* (5.5) *and suppose that assumptions* (H1)–(H3) *are satisfied. Then as* $n \to \infty$,

$$\hat{\rho}_n(\alpha) \overset{a.s.}{\to} \rho(\alpha) \quad and \quad \hat{\rho}_n'(\alpha) \overset{a.s.}{\to} \rho'(\alpha)$$

uniformly in $\alpha \in A$, *where* $\hat{\rho}_n(\alpha)$ *is defined by* (5.2).

PROOF. From the definition (5.2) of $\hat{\rho}_n(\alpha)$ and the fact that

$$\hat{\rho}_n'(\alpha) = \frac{\hat{r}_0(\alpha)\hat{r}_1'(\alpha) - \hat{r}_1(\alpha)\hat{r}_0'(\alpha)}{\hat{r}_0^2(\alpha)},$$

the assertion in this theorem follows immediately upon using Lemma 5.2. $\Diamond$

REMARK 6.3 There are many other commonly-used estimators of $\rho(\alpha)$. For example, $\hat{\rho}_n(\alpha)$ can be defined as the minimizer of

$$\sum_{t=1}^{n-1}[\hat{y}_t(\alpha) - \rho\hat{y}_{t-1}(\alpha)]^2 + \sum_{t=1}^{n-1}[\hat{y}_{t-2}(\alpha) - \rho\hat{y}_{t-1}(\alpha)]^2,$$

yielding

$$\hat{\rho}_n(\alpha) = \frac{\displaystyle\sum_{t=1}^{n-1}\hat{y}_{t-1}(\alpha)[\hat{y}_t(\alpha) + \hat{y}_{t-2}(\alpha)]}{2\displaystyle\sum_{t=1}^{n-1}\hat{y}_{t-1}^2(\alpha)}. \tag{5.9}$$

It can be shown by a similar argument as in the proof of Lemma 5.2, all these estimators are uniformly equivalent as n tends to infinity, and, therefore, Theorem 5.1 remains valid for these estimators. $\Diamond$

As a consequence of this theorem, the results obtained in Section'4 can be restated as follows:

Corollary 5.1 *Let $\{\epsilon_t\}$ be a linear process defined by (5.5). Suppose that $C(\alpha^*) < 1$ and (H1)–(H3) are satisfied with $A := \bar{S}_\Delta(\alpha^*) \subseteq [\underline{\alpha}, \overline{\alpha}]$ for some $\Delta > 0$. Assume further that $\{h'_j(\alpha)\}$ are continuous on $\bar{S}_\Delta(\alpha^*)$. Then the following results hold for sufficiently large n with probability one:*

(a) *$\hat{\rho}_n(\alpha)$ has a unique fixed-point $\hat{\alpha}_n$ in some $S_\delta(\alpha^*) \subset A$.*

(b) *There exists $S_{\delta_0}(\hat{\alpha}_n) \subset S_\delta(\alpha^*)$ such that the sequence $\{\alpha_n^{(m)}\}$ given by (2.5) converges monotonically and at least linearly to $\hat{\alpha}_n$ as $m \to \infty$, provided $\alpha_n^{(0)} \in S_{\delta_0}(\hat{\alpha}_n)$. In this case, (4.8) also holds for any $m \geq 1$.*

(c) *If in addition $G(\alpha) \geq g > 0$ for all $\alpha \in \bar{S}_\delta(\alpha^*)$, then $\hat{\alpha}_n$ converges to α^* with probability one as $n \to \infty$.*

PROOF. According to Theorem 4.2 and Theorem 4.4, it suffices to check the conditions in Theorem 4.1. To do this, we first notice that the differentiability of $\rho(\alpha)$ and the continuity of $C(\alpha)$ are guaranteed by (H2), (H3), and the continuity of $h'_j(\alpha)$. Moreover, (4.1) and (4.2) are consequences of Theorem 5.1. $\Diamond$

For the convergence of the secant method (3.2), an additional condition on the second derivative of $h_j(\alpha)$ is needed:

(H4) $h''_j(\alpha)$ exists and there are constants $c_j > 0$ such that

$$\sum_{j=0}^{\infty} j c_j < \infty \quad \text{and} \quad |h''_j(\alpha)| \leq c_j$$

for all $j = 0, 1, \dots$ and all $\alpha \in A$.

So, the secant method is more stringent than the FPI. By a similar argument, it is readily seen that under (H4), $\rho(\alpha)$ is twice differentiable and $\hat{\rho}''_n(\alpha)$ is a uniformly consistent estimator of $\rho''(\alpha)$. Therefore, we also have

15

Corollary 5.2 *Suppose that the conditions in Corollary 5.1 hold. If in addition assumption (H4) is satisfied and $\rho''(\alpha)$ is bounded on $A := \bar{S}_\Delta(\alpha^*)$, then there exists $\delta_1 > 0$ such that the sequence $\{\hat{\alpha}_n^{(m)}\}$ given by (3.2) converges to $\hat{\alpha}_n$ as $m \to \infty$ at least superlinearly with probability one for sufficiently large n, provided $\hat{\alpha}_n^{(-1)}$, $\hat{\alpha}_n^{(0)} \in S_{\delta_1}(\hat{\alpha}_n)$.*

PROOF. The assertions follow immediately from Theorem 4.3 upon noting that (H4) and the boundedness of $\rho''(\alpha)$ imply the boundedness of $e_n''(\alpha)$ with probability one for sufficiently large n. $\Diamond$

6 Multiple Sinusoids in Noise

In previous sections, most of the results were restricted to the case of a single sinusoid in noise defined by (1.1). Now let us consider the general case of multiple sinusoids in noise given by (5.7). We would like to discuss conditions for which the CM method provides consistent estimates of the unknown frequencies.

We first notice that for the observation $\{y_t\}$ defined by (5.7), the first-order autocorrelation $\rho(\alpha)$ of filtered data can be expressed as

$$\rho(\alpha) = \frac{\displaystyle\sum_{k=0}^{q-1} \sigma_k^2 |H(\omega_k; \alpha)|^2 \cos\omega_k + \int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \cos\omega \, dF(\omega)}{\displaystyle\sum_{k=0}^{q-1} \sigma_k^2 |H(\omega_k; \alpha)|^2 + \int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \, dF(\omega)}.$$

Define $\alpha_k^* := \cos\omega_k$. Then, under assumption (2.3), this expression reduces to a counterpart of (2.4)

$$\rho(\alpha) = \alpha_k^* + \sum_{j=0}^{q-1} G_j(\alpha)(\alpha_j^* - \alpha_k^*) + C_q(\alpha)(\alpha - \alpha_k^*) \tag{6.1}$$

where

$$C_q(\alpha) := \frac{1}{1 + \displaystyle\sum_{k=0}^{q-1} \gamma_k(\alpha)}, \quad G_j(\alpha) := \frac{\gamma_j(\alpha)}{1 + \displaystyle\sum_{k=0}^{q-1} \gamma_k(\alpha)}$$

$$\gamma_k(\alpha) := \frac{\sigma_k^2 |H(\omega_k; \alpha)|^2}{\displaystyle\int_{-\pi}^{\pi} |H(\omega; \alpha)|^2 \, dF(\omega)}.$$

16

Clearly, $\gamma_k(\alpha)$ is the signal-to-noise ratio of the kth sinusoid after filtering with $\mathcal{L}_\alpha$, $G_j(\alpha) \geq 0$ is the gain coefficient of the jth sinusoid, and $C_q(\alpha)$ is the contraction coefficient satisfying $0 < C_q(\alpha) = 1 - \sum_{j=0}^{q-1} G_j(\alpha) \leq 1$. In particular, from (6.1), we have

$$\rho(\alpha_k^*) = \alpha_k^* + \sum_{j=0}^{q-1} G_j(\alpha_k^*)(\alpha_j^* - \alpha_k^*).$$

Suppose that $C_q(\alpha)$ is continuous and $G_j(\alpha)$ is differentiable at α_k^*. Therefore, $\rho(\alpha)$ is also differentiable at α_k^* and its derivative can be written as

$$
\begin{aligned}
\rho'(\alpha_k^*) & := \lim_{\alpha \to \alpha_k^*} \frac{\rho(\alpha) - \rho(\alpha_k^*)}{\alpha - \alpha_k^*} \\
& = \lim_{\alpha \to \alpha_k^*} \left\{ \sum_{j \neq k} \frac{G_j(\alpha) - G_j(\alpha_k^*)}{\alpha - \alpha_k^*}(\alpha_j^* - \alpha_k^*) + C_q(\alpha) \right\} \\
& = \sum_{j \neq k} G_j'(\alpha_k^*)(\alpha_j^* - \alpha_k^*) + C_q(\alpha_k^*).
\end{aligned}
$$

Clearly, a sufficient condition for $\rho(\alpha)$ to be contractive in a neighborhood of α_k^* is that $\rho'(\alpha)$ is continuous in the vicinity of α_k^* and

$$-1 < \sum_{j \neq k} G_j'(\alpha_k^*)(\alpha_j^* - \alpha_k^*) + C_q(\alpha_k^*) < 1. \tag{6.2}$$

The following theorem guarantees the existence of a unique fixed-point of $\hat{\rho}_n(\alpha)$ in a neighborhood of α_k^* and the convergence of iterative procedures (2.5) and (3.2).

Theorem 6.1 *Under the conditions in Theorem 4.1 about $\rho(\alpha)$ and $e_n(\alpha)$, if (6.2) is satisfied, then there exist $0 < \delta_1 \leq \delta_0 \leq \delta \leq \Delta$ such that the following results hold with probability tending to one as $n \to \infty$ (or with probability one for sufficiently large n).*

(a) *$\hat{\rho}_n(\alpha)$ has a unique fixed-point $\hat{\alpha}_n$ in $S_\delta(\alpha_k^*)$.*

(b) *The sequence $\{\hat{\alpha}_n^{(m)}\}$ defined by (2.5) converges to $\hat{\alpha}_n$ at least linearly as $m \to \infty$, provided $\hat{\alpha}_n^{(0)} \in S_{\delta_0}(\hat{\alpha}_n)$.*

(c) *If the assumptions in Theorem 4.3 about $\rho(\alpha)$ and $e_n(\alpha)$ are also satisfied, the sequence $\{\hat{\alpha}_n^{(m)}\}$ defined by (3.2) converges to $\hat{\alpha}_n$ at least superlinearly as $m \to \infty$, provided $\hat{\alpha}_n^{(-1)}$, $\hat{\alpha}_n^{(0)} \in S_{\delta_1}(\hat{\alpha}_n)$.*

17

PROOF. The assertions can be proved in the same way of showing Theorems 4.1, 4.2, and 4.3. $\Diamond$

Let $\hat{\alpha}_n$ be a fixed-point of $\hat{\rho}_n(\alpha)$ in $S_\delta(\alpha_k^*)$. From (6.1), we obtain

$$\sum_{j=0}^{q-1} G_j(\hat{\alpha}_n)(\hat{\alpha}_n - \alpha_j^*) = e_n(\hat{\alpha}_n). \qquad (6.3)$$

Unlike the case of a single sinusoid, this equation does not lead to the conclusion that $\hat{\alpha}_n \to \alpha_k^*$, even under the assumption that $G_k(\alpha) \geq g > 0$ for all $\alpha \in S_\delta(\alpha_k^*)$. In order to achieve consistency, a much stronger condition is required that prevents $\hat{\alpha}_n$ from converging to false frequency. Obviously, the condition $G_j(\alpha) = 0$ for all $\alpha \notin S_\delta(\alpha_k^*)$ does the job. This condition simply says that those sinusoids with frequencies different from α_k^* must be *completely* filtered out by $\mathcal{L}_\alpha$ with α in the vicinity of α_k^*. In conclusion, $\hat{\rho}_n(\alpha)$ may still have a unique fixed-point in the vicinity of α_k^* and iterative procedures such as (2.5) and (3.2) may still converge to that fixed-point under condition (6.2) and uniform consistency of $\hat{\rho}_n(\alpha)$. However, the fixed-point is not necessarily a consistent estimator of α_k^*, unless other frequencies can be completely cleaned up by $\mathcal{L}_\alpha$ with α in the vicinity of α_k^* (see also [6, 10, 12]). This clearly requires that the frequencies be well-separated.

7 Examples

This section gives two examples of parametric filters that can be applied in frequency estimation using the CM method.

7.1 The α-Filter

The exponentially-weighted moving average filter, or, the "α-filter", was originally studied by He and Kedem [6] for frequency estimation. It can be defined recursively by

$$y_t(\alpha) = \alpha y_{t-1}(\alpha) + y_t$$

where $-1 < \alpha < 1$. It is easy to see that $h_j(\alpha) = 0$ for $j < 0$, $h_j(\alpha) = \alpha^j$ for $j = 0, 1, \ldots,$ and

$$|H(\omega; \alpha)|^2 = \frac{1}{1 - 2\alpha \cos \omega + \alpha^2}.$$

Table 1: Comparison of FPI and Secant Procedures with α-Filter

m	FPI [initial value= 0.6π]		Secant [initial value= $(0.3\pi, 0.6\pi)$]	
	MEAN $\pm$ SDV ($\times\pi$)	MSE ($\times\pi^2$)	MEAN $\pm$ SDV ($\times\pi$)	MSE ($\times\pi^2$)
1	0.495854 ± 0.016533	6.02714×10^{-3}	0.495854 ± 0.016533	6.02714×10^{-3}
2	0.447814 ± 0.017789	1.09008×10^{-3}	0.421869 ± 0.020007	4.03792×10^{-4}
3	0.430492 ± 0.017368	4.12172×10^{-4}	0.420337 ± 0.017723	3.14232×10^{-4}
4	0.424347 ± 0.017484	3.24577×10^{-4}	0.420744 ± 0.018096	3.28011×10^{-4}
5	0.422112 ± 0.017713	3.18207×10^{-4}	0.420741 ± 0.018092	3.27851×10^{-4}
6	0.421277 ± 0.017881	3.21364×10^{-4}	0.420741 ± 0.018092	3.27850×10^{-4}

Some other interesting properties of the α-filter can be found in [11].

Clearly, $0 < C(\alpha^*) < 1$, and (H1)–(H3) are satisfied on any closed subinterval of $(-1, 1)$. When $\{\epsilon_t\}$ is white, it can be easily shown [6] that the fundamental property (2.3) holds for all $\alpha \in (-1, 1)$. Consequently, this filter can be applied to estimate ω_0 using iterative procedures (2.5) or (3.2). In particular, when $\{\epsilon_t\} \sim \text{IID}(0, \sigma_\epsilon^2)$, Corollary 5.1 and Corollary 5.2 guarantee the convergence of these procedures, and also the strong consistency of their limits for estimating α^*. To illustrate the performance of the α-filter, the FPI procedure (2.5) and the secant method (3.2) are applied in the estimation of a single sinusoid in additive Gaussian white noise defined by (1.1) with $\omega_0 = 0.42\pi$, $\phi = 0.1\pi$, and SNR= 3 dB. The initial frequency estimate is taken to be 0.6π (or $\hat{\alpha}_n^{(0)} = \cos 0.6\pi$) for the FPI and $(0.3\pi, 0.6\pi)$ (or $\hat{\alpha}_n^{(-1)} = \cos 0.3\pi$, $\hat{\alpha}_n^{(0)} = \cos 0.6\pi$) for the secant method. Table 1 presents estimated ensemble averages based on 200 independent realizations of size $n = 100$, where m stands for the number of iterations, and the mth estimate of ω_0 is defined by

$$\hat{\omega}_0^{(m)} := \arccos \hat{\rho}_n(\hat{\alpha}_n^{(m-1)}) = \hat{\alpha}_n^{(m)} \qquad m = 1, 2, \ldots \tag{7.1}$$

As can be seen in this table, the FPI and the secant methods work equally well in this experiment, but the latter converges slightly faster at the beginning.

19

7.2 The AR(2) Filter

Considerable improvements in estimation accuracy can be achieved by the following AR(2) filter defined recursively by

$$y_t(\alpha) + \theta(\alpha)\eta y_{t-1}(\alpha) + \eta^2 y_{t-2}(\alpha) = y_t \tag{7.2}$$

where $0 < \eta \leq 1$ and

$$\theta(\alpha) := -\frac{1+\eta^2}{\eta}\alpha.$$

When $\{\epsilon_t\} \sim \text{IID}(0, \sigma_\epsilon^2)$, it is easy to verify, using a formula given in [7], that

$$\rho_\epsilon(\alpha) = -\frac{\eta}{1+\eta^2}\theta(\alpha) = \alpha$$

for all $\alpha \in [\underline{\alpha}, \overline{\alpha}]$ with

$$\underline{\alpha} := -\frac{2\eta}{1+\eta^2} \quad \text{and} \quad \overline{\alpha} := \frac{2\eta}{1+\eta^2}.$$

That is, the fundamental property (2.3) is satisfied by the AR(2) filter (7.2). Its two poles are readily seen to be

$$\zeta_1(\alpha) := \frac{\eta}{2}\left(-\theta(\alpha) + i\sqrt{4 - \theta^2(\alpha)}\right),$$

$$\zeta_2(\alpha) := \frac{\eta}{2}\left(-\theta(\alpha) - i\sqrt{4 - \theta^2(\alpha)}\right)$$

with $|\zeta_1(\alpha)| \equiv |\zeta_2(\alpha)| \equiv \eta$ for all $\alpha \in [\underline{\alpha}, \overline{\alpha}]$. Clearly, whenever $\eta < 1$, the poles are contracted within the unit circle in the complex domain so that the filter (7.2) becomes BIBO-stable, i.e., $\sum |h_j(\alpha)| < \infty$. Stability of (7.2) is extremely important for the on-line implementation of the CM method in frequency tracking. This problem will be addressed separately in another paper.

Note that the impulse response of the AR(2) filter can be written as

$$h_k(\alpha) = \sum_{j=0}^{k} \zeta_1^j(\alpha)\zeta_2^{k-j}(\alpha).$$

Clearly, (H2) is satisfied if $\eta < 1$, since $|h_k(\alpha)| \leq a_k := (k+1)\eta^k$ for all $k \geq 0$. Moreover, since

$$\zeta_1'(\alpha) = -\frac{1+\eta^2}{2}\left(-1 + i\frac{\theta(\alpha)}{\sqrt{4 - \theta^2(\alpha)}}\right),$$

$$\zeta_2'(\alpha) = -\frac{1+\eta^2}{2}\left(-1 - i\frac{\theta(\alpha)}{\sqrt{4 - \theta^2(\alpha)}}\right),$$

20

Table 2: Estimates by FPI with AR(2) Filter (Initial Value = 0.55π)

m	$\eta = 0.95$		$\eta = 0.98$	
	MEAN $\pm$ SDV ($\times\pi$)	MSE ($\times\pi^2$)	MEAN $\pm$ SDV ($\times\pi$)	MSE ($\times\pi^2$)
1	0.498308 ± 0.012725	6.29414×10^{-3}	0.516345 ± 0.013724	9.47067×10^{-3}
3	0.427856 ± 0.007107	1.12234×10^{-4}	0.454814 ± 0.022079	1.69947×10^{-3}
6	0.419949 ± 0.000874	7.67307×10^{-7}	0.421551 ± 0.002866	1.06195×10^{-5}
8	0.419914 ± 0.000867	7.59110×10^{-7}	0.420060 ± 0.000841	7.10807×10^{-7}
10	0.419913 ± 0.000867	7.59001×10^{-7}	0.419925 ± 0.000795	6.37827×10^{-7}

and $|\theta(\alpha)|^2 \leq 4$ for all α with $|\theta(\alpha)|^2 = 4$ if and only if $\alpha = \underline{\alpha}$ or $\alpha = \overline{\alpha}$, then $|\zeta_1'(\alpha)|$ and $|\zeta_2'(\alpha)|$ are uniformly bounded by a constant c for all $\alpha \in A := [\underline{\alpha} + \delta, \overline{\alpha} - \delta]$ where $\delta > 0$. Note that, with obvious notation,

$$h_k'(\alpha) = \sum_{j=0}^{k} j\zeta_1^{j-1}\zeta_2^{k-j}\zeta_1' + \sum_{j=0}^{k}(k-j)\zeta_1^j\zeta_2^{k-j-1}\zeta_2'$$

and thus

$$
\begin{aligned}
|h_k'(\alpha)| &\leq c\sum_{j=0}^{k} j\eta^{k-1} + c\sum_{j=0}^{k}(k-j)\eta^{k-1} \\
&= b_k := ck(k+1)\eta^{k-1}
\end{aligned}
$$

for all $k \geq 0$ and $\alpha \in A$. Assumption (H3) is clearly satisfied if $\eta < 1$. It can be shown similarly that (H4) is also valid under the same condition. As a consequence, Corollary 5.1 and Corollary 5.2 apply to the AR(2) filter for estimating ω_0. (Note that $|H(\omega; \alpha)|^2 > 0$ for all ω and α, and hence $C(\alpha^*) < 1$.)

To demonstrate its performance, we apply the AR(2) filter with the FPI procedure to the same data as in Section 7.1, and Table 2 presents the results for $\eta = 0.95$ and 0.98.

As compared with the α-filter, the AR(2) filter provides greater precision for frequency estimation in terms of smaller variance and mean-square error (MSE). This is because the AR(2) filter is bandpass and enhances considerably the frequence components in the vicinity of the angles of its ploes, which, when $\alpha = \alpha^*$ and $\eta \approx 1$, are approximately $\pm\omega_0$. The role of η can be appreciated by comparing the results for $\eta = 0.95$ and those for $\eta = 0.98$. Clearly, the

closer the parameter η is to 1, the smaller the variance and the MSE, but the lower is the rate of convergence.

Finally, it is worth pointing out that in the extreme case when $\eta = 1$, the CM method using the FPI procedure with the AR(2) filter (7.2) coincides with the method proposed by Quinn and Fernandes [16]. They have shown that in this case, the estimator $\hat{\omega}_n$ is $n^{3/2}$-consistent, just like the nonlinear least squares estimator. For $\eta < 1$, we can show [15] that $\hat{\omega}_n$ is $n^{1/2}$-consistent. However, it is important to note that for $\eta = 1$, the FPI requires a much more accurate initial estimate than it does for $\eta < 1$. In fact, when $\eta = 1$, as proved in [16], the accuracy of the initial estimate is required to be of order n^{-1}. (A modified method in [16] can reduce this order to $n^{-1/2}$.) Therefore, their method must be used in connection with another method that provides satisfactory initial estimates. On the other hand, the CM method with $\eta < 1$ requires the accuracy of the initial estimate to be $O(1)$. By taking the advantage of the flexibility for the choice of η, the CM method does not require any other method for initialization of the procedure while still achieving better and better estimates by increasing η, and eventually, as $\eta \to 1$, obtaining $n^{3/2}$-consistency, as the method proposed in [16]. Furthermore, with a flexible η, the CM method can be applied in the estimation of a time-varing frequency for which the Quinn-Fernandes method may fail due to an excessively narrow bandwidth that can easily "loose" the frequency. Adaptive estimation of time-varying frequencies using the CM method will be addressed elsewhere.

A   Proof of Lemma 5.1

Define

$$Q(t) := \left(\sum_{j=0}^{t+\tau} h_j(\alpha) y_{t-j+\tau} \right) \left(\sum_{j=0}^{t} g_j(\alpha) y_{t-j} \right).$$

From (1.1), $Q(t)$ can be written as

$$Q(t) = I_1(t) + I_2(t) + I_3(t)$$

where

$$I_1(t) \quad := \quad \beta^2 \sum_{j=0}^{t+\tau} \sum_{k=0}^{t} h_j g_k \cos[\omega_0(t - j + \tau) + \phi] \cos[\omega_0(t - k) + \phi]$$

$$I_2(t) \quad := \quad \beta \sum_{j=0}^{t+\tau} \sum_{k=0}^{t} h_j g_k \{\epsilon_{t-j+\tau} \cos[\omega_0(t - k) + \phi] + \epsilon_{t-k} \cos[\omega_0(t - j + \tau) + \phi]\}$$

$$I_3(t) \quad := \quad \sum_{j=0}^{t+\tau} \sum_{k=0}^{t} h_j g_k \epsilon_{t-j+\tau} \epsilon_{t-k}.$$

Here the argument α is omitted in h_j and g_k for notational simplicity.

Using the trigonometric identity

$$\cos \lambda_1 \cos \lambda_2 = \frac{1}{2}[\cos(\lambda_1 - \lambda_2) + \cos(\lambda_1 + \lambda_2)],$$

$I_1(t)$ can be written as $I_1(t) = T_1(t) + T_2(t)$ where

$$T_1(t) \quad := \quad \sigma^2 \sum_{j=0}^{t+\tau} \sum_{k=0}^{t} h_j g_k \cos[\omega_0(k - j + \tau)]$$

$$T_2(t) \quad := \quad \sigma^2 \sum_{j=0}^{t+\tau} \sum_{k=0}^{t} h_j g_k \cos[\omega_0(2t - j - k + \tau) + 2\phi].$$

As $t \to \infty$, assumption (H2) implies that

$$T_1(t) \quad \to \quad \sigma^2 \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} h_j g_k \cos[\omega_0(k - j + \tau)]$$

$$= \quad \sigma^2 \, \Re \left\{ \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} h_j g_k e^{i(k-j+\tau)\omega_0} \right\}$$

$$= \quad \sigma^2 \, \Re \left\{ H(\omega_0; \alpha) \overline{G(\omega_0; \alpha)} e^{i\tau\omega_0} \right\}$$

uniformly in $\alpha \in A$. Therefore,

$$n^{-1} \sum_{t=0}^{n-\tau-1} T_1(t) \to \sigma^2 \, \Re \left\{ H(\omega_0; \alpha) \overline{G(\omega_0; \alpha)} e^{i\tau\omega_0} \right\} \tag{A.1}$$

uniformly in $\alpha \in A$ as $n \to \infty$. Moreover, the following identity holds for any function $u(t,s)$ by interchanging the order of summations:

$$U \quad := \quad \sum_{t=0}^{n-\tau-1} \sum_{j=0}^{t+\tau} \sum_{k=0}^{t} h_j g_k u(t - j + \tau, t - k) \tag{A.2}$$

$$= \quad \sum_{j=\tau}^{n-1} \sum_{k=0}^{j-\tau-1} \sum_{t=0}^{n-j-1} h_j g_k u(t, t + j - k - \tau)$$

$$+ \quad \sum_{k=0}^{n-\tau-1} \sum_{j=0}^{k+\tau} \sum_{t=0}^{n-k-\tau-1} h_j g_k u(t + k - j + \tau, t).$$

23

In fact, by first interchanging the summations over t and j in the first expression and by substituting $t + j - \tau$ for t afterwards, we obtain

$$U = \left(\sum_{j=0}^{\tau-1} \sum_{t=\tau-j}^{n-j-1} + \sum_{j=\tau}^{n-1} \sum_{t=0}^{n-j-1} \right) \sum_{k=0}^{t+j-\tau} h_j g_k u(t, t+j-k-\tau).$$

Then, interchanging the summations over t and k again gives

$$U = \left(\sum_{j=0}^{\tau-1} \sum_{k=0}^{n-\tau-1} \sum_{t=k-j+\tau}^{n-j-1} + \sum_{j=\tau}^{n-1} \sum_{k=0}^{j-\tau-1} \sum_{t=0}^{n-j-1} + \sum_{j=\tau}^{n-1} \sum_{k=j-\tau}^{n-\tau-1} \sum_{t=k-j+\tau}^{n-j-1} \right) h_j g_k u(t, t+j-k-\tau).$$

By finally interchanging the summations over j and k in the first and the last terms and combining them afterwards, we obtain

$$U = \left(\sum_{k=0}^{n-\tau-1} \sum_{j=0}^{k+\tau} \sum_{t=k-j+\tau}^{n-j-1} + \sum_{j=\tau}^{n-1} \sum_{k=0}^{j-\tau-1} \sum_{t=0}^{n-j-1} \right) h_j g_k u(t, t+j-k-\tau).$$

Identity (A.2) follows immediately by substituting $t + k - j + \tau$ for t in the first term. Now applying (A.2) with $u(t, s) = \sigma^2 \cos[\omega_0(t+s) + 2\phi]$, we obtain

$$
\begin{aligned}
\sum_{t=0}^{n-\tau-1} T_2(t) &= \sigma^2 \sum_{j=\tau}^{n-1} \sum_{k=0}^{j-\tau-1} h_j g_k \sum_{t=0}^{n-j-1} \cos[\omega_0(2t+j-k-\tau) + 2\phi] \\
&+ \sigma^2 \sum_{k=0}^{n-\tau-1} \sum_{j=0}^{k+\tau} h_j g_k \sum_{t=0}^{n-k-\tau-1} \cos[\omega_0(2t+k-j+\tau) + 2\phi] \\
&:= U_1 + U_2.
\end{aligned}
$$

Let a_j^h and a_j^g be the constants in (H2) associated with h_j and g_j, respectively. Then, for each j and k, and for any $\alpha \in A$,

$$n^{-1} \left| h_j g_k \sum_{t=0}^{n-j-1} \cos[\omega_0(2t+j-k-\tau) + 2\phi] \right| \leq a_j^h a_k^g$$

and as $n \to \infty$,

$$n^{-1} \sum_{t=0}^{n-j-1} \cos[\omega_0(2t+j-k-\tau) + 2\phi] \overset{a.s.}{\to} 0.$$

According to assumption (H2) and the dominated convergence theorem, $n^{-1} U_2 \overset{a.s.}{\to} 0$ uniformly in $\alpha \in A$. The same assertion is also true for U_2 by a similar argument. Therefore, $n^{-1} \sum_{t=0}^{n-\tau-1} T_2(t) \overset{a.s.}{\to} 0$ uniformly in $\alpha \in A$. Combining this with (A.1) yields

$$n^{-1} \sum_{t=0}^{n-\tau-1} I_1(t) \overset{a.s.}{\to} \sigma^2 \, \Re \left\{ H(\omega_0; \alpha) \overline{G(\omega_0; \alpha)} e^{i\tau\omega_0} \right\} \tag{A.3}$$

uniformly in $\alpha \in A$ as $n \to \infty$.

Write $I_2(t) = T_3(t) + T_4(t)$ where

$$T_3(t) \ := \ \beta \sum_{j=0}^{t+\tau} \sum_{k=0}^{t} h_j g_k \epsilon_{t-j+\tau} \cos[\omega_0(t-k)+\phi]$$

$$T_4(t) \ := \ \beta \sum_{j=0}^{t+\tau} \sum_{k=0}^{t} h_j g_k \epsilon_{t-k} \cos[\omega_0(t-j+\tau)+\phi].$$

Applying (A.2) with $u(t,s) = \beta \epsilon_t \cos(\omega_0 s + \phi)$ gives

$$\sum_{t=0}^{n-\tau-1} T_3(t) \ = \ \sum_{j=\tau}^{n-1} \sum_{k=0}^{j-\tau-1} h_j g_k \sum_{t=0}^{n-j-1} \epsilon_t \cos[\omega_0(t+j-k-\tau)+\phi] \qquad (A.4)$$

$$+ \ \sum_{k=0}^{n-\tau-1} \sum_{j=0}^{k+\tau} h_j g_k \sum_{t=0}^{n-k-\tau-1} \epsilon_{t+k-j+\tau} \cos(\omega_0 t + \phi)$$

$$:= \ U_3 + U_4.$$

Splitting U_3 into two terms, we get

$$U_3 \ = \ \beta \left(\sum_{j=\tau}^{N-1} + \sum_{j=N}^{n-1} \right) \sum_{k=0}^{j-\tau-1} h_j g_k \sum_{t=0}^{n-j-1} \epsilon_t \cos[\omega_0(t+j-k-\tau)+\phi]$$

$$:= \ U_3^{(N)} + (U_3 - U_3^{(N)}).$$

For any $\alpha \in A$,

$$|U_3^{(N)}| \le \beta \sum_{j=\tau}^{N-1} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g \left(\left| \sum_{t=0}^{n-j-1} \epsilon_t \cos \omega_0 t \right| + \left| \sum_{t=0}^{n-j-1} \epsilon_t \sin \omega_0 t \right| \right)$$

with probability one. For each fixed j, it can be shown [1] that

$$\left| \sum_{t=0}^{n-j-1} \epsilon_t \cos \omega_0 t \right| \le O_{a.s.}\left(\sqrt{n \log n} \right) \quad \text{and} \quad \left| \sum_{t=0}^{n-j-1} \epsilon_t \sin \omega_0 t \right| \le O_{a.s.}\left(\sqrt{n \log n} \right)$$

as $n \to \infty$. Therefore, for each fixed N,

$$n^{-1} U_3^{(N)} \le O_{a.s.}\left(\sqrt{\log n/n} \right) \overset{a.s.}{\to} 0$$

uniformly in $\alpha \in A$. This implies that uniformly in $\alpha \in A$,

$$\lim_{N \to \infty} \lim_{n \to \infty} n^{-1} U_3^{(N)} = 0 \qquad a.s. \qquad (A.5)$$

25

Moreover, it is readily seen that for any $\alpha \in A$,

$$
\begin{aligned}
|U_3 - U_3^{(N)}| &\leq \beta \sum_{j=N}^{n-1} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g \sum_{t=0}^{n-j-1} |\epsilon_t| \\
&\leq \beta \sum_{j=N}^{n-1} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g \sum_{t=0}^{n-1} |\epsilon_t| \\
&\leq \beta G \sum_{t=0}^{n-1} |\epsilon_t| \sum_{j=N}^{\infty} a_j^h
\end{aligned}
$$

with probability one, where $G := \sum_{k=0}^{\infty} a_k^g$. Define $\zeta_N(t) := \beta G |\epsilon_t| \sum_{j=N}^{\infty} a_j^h$, then $\{\zeta_N(t)\}$ is strictly stationary for each fixed N, and

$$
|U_3 - U_3^{(N)}| \leq \sum_{t=0}^{n-1} \zeta_N(t)
$$

with probability one for any $\alpha \in A$. According to the strong ergodic theorem (see, for example, [8]), for each fixed N, we have

$$
\lim_{n \to \infty} n^{-1} \sum_{t=0}^{n-1} \zeta_N(t) = \zeta_N \qquad a.s.
$$

for some random variable ζ_N, and

$$
E(\zeta_N) = E\{\zeta_N(0)\} = \beta G\, E|\epsilon_0| \sum_{j=N}^{\infty} a_j^h.
$$

Since $\zeta_n \geq 0$ with probability one, and

$$
\sum_{N=0}^{\infty} E(\zeta_N) = \beta G\, E|\epsilon_0| \sum_{N=0}^{\infty} \sum_{j=N}^{\infty} a_j^h = \beta G\, E|\epsilon_0| \sum_{j=0}^{\infty} (j+1) a_j^h < \infty,
$$

using Chebyshev's inequality, we obtain, for any $\mu > 0$,

$$
\mathrm{P}\left\{ \bigcup_{N=N'}^{\infty} (\zeta_N > \mu) \right\} \leq \sum_{N=N'}^{\infty} \mathrm{P}\{\zeta_N > \mu\} \leq \frac{1}{\mu} \sum_{N=N'}^{\infty} E(\zeta_N) \to 0
$$

as $N' \to \infty$. Therefore, $\mathrm{P}\{$ i.o. $\zeta_N > \mu\} = 0$ for any $\mu > 0$, which implies that $\zeta_N \overset{a.s.}{\to} 0$ as $N \to \infty$. Consequently,

$$
\lim_{N \to \infty} \limsup_{n \to \infty} n^{-1} |U_1 - U_1^{(N)}| = 0 \qquad a.s.
$$

uniformly in $\alpha \in A$. Combining with (A.5) gives $n^{-1} U_3 \overset{a.s.}{\to} 0$ as $n \to \infty$ uniformly in $\alpha \in A$. By switching k and j in the second term of (A.4), U_4 can be written as

$$
\begin{aligned}
U_4 &= \beta \left(\sum_{j=0}^{N-1} + \sum_{j=N}^{n-\tau-1} \right) \sum_{k=0}^{j+\tau} h_k g_j \sum_{t=0}^{n-j-\tau-1} \epsilon_{t+j-k+\tau} \cos(\omega_0 t + \phi) \\
&:= U_4^{(N)} + (U_4 - U_4^{(N)})
\end{aligned}
$$

26

Like $U_3^{(N)}$, it can be shown in a similar way that

$$\lim_{N \to \infty} \lim_{n \to \infty} n^{-1} U_4^{(N)} = 0 \qquad a.s.$$

uniformly in $\alpha \in A$. Furthermore, for all $\alpha \in A$,

$$
\begin{aligned}
|U_4 - U_4^{(N)}| &\leq \beta \sum_{j=N}^{n-\tau-1} \sum_{k=0}^{j+\tau} a_k^h a_j^g \sum_{t=0}^{n-j-\tau-1} |\epsilon_{t+j-k+\tau}| \\
&= \beta \sum_{j=N}^{n-\tau-1} \sum_{k=0}^{j+\tau} a_k^h a_j^g \sum_{t=j-k+\tau}^{n-k-1} |\epsilon_t| \\
&\leq \beta \sum_{j=N}^{n-\tau-1} \sum_{k=0}^{j+\tau} a_k^h a_j^g \sum_{t=0}^{n-1} |\epsilon_t| \\
&\leq \beta H \sum_{t=0}^{n-1} |\epsilon_t| \sum_{j=N}^{\infty} a_j^g
\end{aligned}
$$

with probability one, where $H := \sum_{k=0}^{\infty} a_k^h$. Like $U_3 - U_3^{(N)}$, we have

$$\lim_{N \to \infty} \limsup_{n \to \infty} n^{-1} |U_4 - U_4^{(N)}| = 0 \qquad a.s.$$

Consequently, $n^{-1} U_4 \overset{a.s.}{\to} 0$ as $n \to \infty$ uniformly in $\alpha \in A$. Combining all these results gives $n^{-1} \sum_{t=0}^{n-1} T_3(t) \overset{a.s.}{\to} 0$ uniformly in $\alpha \in A$. The same result can be established for $T_4(t)$ upon noting the resemblance between $T_3(t)$ and $T_4(t)$. Therefore, we have proved

$$n^{-1} \sum_{t=0}^{n-\tau-1} I_2(t) \overset{a.s.}{\to} 0 \tag{A.6}$$

uniformly in $\alpha \in A$.

Finally, let us consider $I_3(t)$. Using (A.2) with $u(t, s) = \epsilon_t \epsilon_s$, we get

$$
\begin{aligned}
\sum_{t=0}^{n-\tau-1} I_3(t) &= \sum_{j=\tau}^{n-1} \sum_{k=0}^{j-\tau-1} h_j g_k \sum_{t=0}^{n-j-1} \epsilon_t \epsilon_{t+j-k-\tau} \\
&\quad + \sum_{k=0}^{n-\tau-1} \sum_{j=0}^{k+\tau} h_j g_k \sum_{t=0}^{n-k-\tau-1} \epsilon_t \epsilon_{t+k-j+\tau} \\
&:= U_5 + U_6
\end{aligned}
$$

Splitting U_5 into two terms gives

$$
\begin{aligned}
U_5 &= \left(\sum_{j=\tau}^{N-1} + \sum_{j=N}^{n-1} \right) \sum_{k=0}^{j-\tau-1} h_j g_k \sum_{t=0}^{n-j-1} \epsilon_t \epsilon_{t+j-k-\tau} \\
&:= U_5^{(N)} + (U_5 - U_5^{(N)}).
\end{aligned}
$$

27

By strong ergodicity (see, for example, [5] and [3]),

$$\lim_{n\to\infty} n^{-1} \sum_{t=0}^{n-j-1} \epsilon_t \epsilon_{t+j-k-\tau} = r^{\epsilon}_{k-j+\tau} \qquad a.s.$$

for each fixed j and k, where

$$r^{\epsilon}_{\tau} := \int_{-\pi}^{\pi} e^{i\tau\omega} dF(\omega).$$

On the other hand,

$$\left| n^{-1} U_5^{(N)} - \sum_{j=\tau}^{N-1} \sum_{k=0}^{j-\tau-1} h_j g_k r^{\epsilon}_{k-j+\tau} \right| \leq \sum_{j=\tau}^{N-1} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g \left| n^{-1} \sum_{t=0}^{n-j-1} \epsilon_t \epsilon_{t+j-k-\tau} - r^{\epsilon}_{k-j+\tau} \right|.$$

Therefore, with probability one,

$$\lim_{N\to\infty} \lim_{n\to\infty} n^{-1} U_5^{(N)} = \sum_{j=\tau}^{\infty} \sum_{k=0}^{j-\tau-1} h_j g_k r^{\epsilon}_{k-j+\tau}$$

$$= \sum_{k=0}^{\infty} \sum_{j=k+\tau+1}^{\infty} h_j g_k r^{\epsilon}_{k-j+\tau}$$

uniformly in $\alpha \in A$. Moreover, for all $\alpha \in A$,

$$|U_5 - U_5^{(N)}| \leq \sum_{j=N}^{n-1} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g \sum_{t=0}^{n-j-1} |\epsilon_t \epsilon_{t+j-k-\tau}|$$

$$\leq \sum_{j=N}^{n-1} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g \sum_{t=0}^{n-1} |\epsilon_t \epsilon_{t+j-k-\tau}|$$

$$\leq \sum_{t=0}^{n-1} \sum_{j=N}^{\infty} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g |\epsilon_t \epsilon_{t+j-k-\tau}|$$

with probability one. Note that in the last inequality the infinite sum

$$Z_N(t) := \sum_{j=N}^{\infty} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g |\epsilon_t \epsilon_{t+j-k-\tau}|$$

converges with probability one, since $Z_N(t) \geq 0$, and by the monotone convergence theorem,

$$E\{Z_N(t)\} = \sum_{j=N}^{\infty} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g E|\epsilon_t \epsilon_{t+j-k-\tau}|$$

$$\leq \sigma_\epsilon^2 \sum_{j=N}^{\infty} \sum_{k=0}^{j-\tau-1} a_j^h a_k^g \leq \sigma_\epsilon^2 G \sum_{j=N}^{\infty} a_j^h < \infty,$$

where $\sigma_\epsilon^2 := \gamma_\epsilon(0)$ and the second expression is a result of Cauchy-Schwarz inequality. Note also that $\{Z_N(t)\}$ is strictly stationary for each fixed N, and by the strong ergodic theorem,

$$n^{-1} \sum_{t=0}^{n-1} Z_N(t) \overset{a.s.}{\to} Z_N$$

28

as $n \to \infty$ for some random variable Z_N with $E(Z_N) = E\{Z_N(0)\} \le \sigma_\epsilon^2 G \sum_{j=N}^\infty a_j^h$. Since $\sum_{N=0}^\infty E(Z_N) \le \sigma_\epsilon^2 G \sum_{j=0}^\infty (j+1) a_j^h < \infty$, we have $Z_N \overset{a.s.}{\to} 0$ as $N \to \infty$. Therefore, uniformly in $\alpha \in A$,

$$\lim_{N \to \infty} \limsup_{n \to \infty} n^{-1} |U_5 - U_5^{(N)}| = 0 \qquad a.s.$$

Combining these results we get

$$n^{-1} U_5 \overset{a.s.}{\to} \sum_{k=0}^\infty \sum_{j=k+\tau+1}^\infty h_j g_k r_{k-j+\tau}^\epsilon$$

uniformly in $\alpha \in A$ as $n \to \infty$. In a similar way, we can also show that

$$n^{-1} U_6 \overset{a.s.}{\to} \sum_{k=0}^\infty \sum_{j=0}^{k+\tau} h_j g_k r_{k-j+\tau}^\epsilon$$

uniformly in $\alpha \in A$ as $n \to \infty$. Consequently, we have proved that uniformly in $\alpha \in A$,

$$\begin{aligned} n^{-1} \sum_{t=0}^{n-\tau-1} I_3(t) &\overset{a.s.}{\to} \sum_{k=0}^\infty \sum_{j=0}^\infty h_j g_k r_{k-j+\tau}^\epsilon \qquad\qquad\qquad (A.7) \\ &= \int_{-\pi}^\pi H(\omega; \alpha) \overline{G(\omega; \alpha)} e^{i\tau\omega} dF(\omega) \end{aligned}$$

as $n \to \infty$. Note that the last quantity is real because of the symmetry of $G(\omega; \alpha)$, $H(\omega; \alpha)$, and $F(\omega)$ as functions of ω. Now (5.6) is proved upon collecting (A.3), (A.6), and (A.7).

B   Proof of Lemma 5.2

First of all, since $\hat{y}_t(\alpha) := 0$ for $t < 0$, $\hat{r}_0(\alpha)$ and $\hat{r}_1(\alpha)$ can be rewritten as

$$\begin{aligned} \hat{r}_0(\alpha) &= n^{-1} \sum_{t=0}^{n-2} \hat{y}_t^2(\alpha) \\ \hat{r}_1(\alpha) &= n^{-1} \sum_{t=0}^{n-2} \hat{y}_{t+1}(\alpha) \hat{y}_t(\alpha). \end{aligned}$$

Define

$$\tilde{r}_0(\alpha) := n^{-1} \sum_{t=0}^{n-1} \hat{y}_t^2(\alpha). \qquad\qquad (B.1)$$

Then it can be shown that under the conditions in Lemma 5.1,

$$\hat{r}_0(\alpha) - \tilde{r}_0(\alpha) \overset{a.s.}{\to} 0$$

uniformly in $\alpha \in A$ as $n \to \infty$. In fact, it is easy to see that

$$
\begin{aligned}
|\hat{r}_0(\alpha) - \tilde{r}_0(\alpha)| &= n^{-1}\hat{y}_{n-1}^2(\alpha) \\
&\leq n^{-1}H^2 + 2Hn^{-1}\sum_{j=0}^{n-1} a_j^h |\epsilon_{n-j-1}| \\
&\quad + n^{-1}\sum_{j=0}^{n-1}\sum_{k=0}^{n-1} a_j^h a_k^h |\epsilon_{n-j-1}\epsilon_{n-k-1}|.
\end{aligned}
$$

Clearly, the first term tends to zero as $n \to \infty$. Since the variance of the second term $\xi_n := 2Hn^{-1}\sum_{j=0}^{n-1} a_j^h |\epsilon_{n-j-1}|$ is bounded by $4H^4\sigma_\epsilon^2 n^{-2}$ and hence $\sum \mathrm{Var}\{\xi_n\} < \infty$, we obtain $\xi_n \overset{a.s.}{\to} 0$. To show that the last term, denoted by θ_n, also vanishes with probability one, let us rewrite θ_n as

$$
\begin{aligned}
\theta_n &= n^{-1}\sum_{j=0}^{n-1}\left(\sum_{k=0}^{j-1} + \sum_{k=0}^{j}\right) a_j^h a_k^h |\epsilon_{n-j-1}\epsilon_{n-k-1}| \\
&\leq n^{-1}\sum_{j=0}^{N-1}\left(\sum_{k=0}^{j-1} + \sum_{k=0}^{j}\right) a_j^h a_k^h |\epsilon_{n-j-1}\epsilon_{n-k-1}| \\
&\quad + n^{-1}\sum_{j=N}^{n-1}\left(\sum_{k=0}^{j-1} + \sum_{k=0}^{j}\right) a_j^h a_k^h \sum_{t=0}^{n-1} |\epsilon_t\epsilon_{t+j-k}| \\
&:= \theta_n^{(N)} + (\theta_n - \theta_n^{(N)}).
\end{aligned}
$$

By strong ergodicity,

$$
\begin{aligned}
n^{-1}\epsilon_{n-j-1}\epsilon_{n-k-1} &= n^{-1}\sum_{t=0}^{n-1}\epsilon_{t-j}\epsilon_{t-k} - n^{-1}\sum_{t=0}^{n-2}\epsilon_{t-j}\epsilon_{t-k} \\
&\overset{a.s.}{\to} r_{k-j}^\epsilon - r_{k-j}^\epsilon = 0
\end{aligned}
$$

for each fixed j and k. Therefore, $\theta_n^{(N)}$ vanishes as $n \to \infty$ for each fixed N, and thus

$$
\lim_{N \to \infty}\lim_{n \to \infty} \theta_n^{(N)} = 0 \qquad a.s.
$$

Moreover, applying the same technique used to derive the limit of $n^{-1}(U_5 - U_5^{(N)})$ in Lemma 5.1 gives

$$
\lim_{N \to \infty}\lim_{n \to \infty} (\theta_n - \theta_n^{(N)}) = 0 \qquad a.s.
$$

Consequently, $\theta_n \overset{a.s.}{\to} 0$ as $n \to \infty$. Combining all these results proves the assertion that $\hat{r}_0(\alpha)$ and $\tilde{r}_0(\alpha)$ are uniformly equivalent as n tends to infinity. Using a similar argument, it can be shown that the uniform equivalence also holds for $\hat{r}_0'(\alpha)$ and $\tilde{r}_0'(\alpha) := 2n^{-1}\sum_{t=0}^{n-1} \hat{y}_t'(\alpha)\hat{y}_t(\alpha)$.

Because of the uniform equivalence, it suffices to show the consistency of $\tilde{r}_0(\alpha)$, $\tilde{r}_0'(\alpha)$, $\hat{r}_1(\alpha)$, and $\hat{r}_1'(\alpha)$. Note that from (5.1), (5.4), and (B.1), the following identities hold:

$$\tilde{r}_0(\alpha) = n^{-1} \sum_{t=0}^{n-1} \left(\sum_{j=0}^{t} h_j(\alpha) y_{t-j} \right) \left(\sum_{j=0}^{t} h_j(\alpha) y_{t-j} \right)$$

$$\hat{r}_1(\alpha) = n^{-1} \sum_{t=0}^{n-2} \left(\sum_{j=0}^{t+1} h_j(\alpha) y_{t-j+1} \right) \left(\sum_{j=0}^{t} h_j(\alpha) y_{t-j} \right)$$

$$\tilde{r}_0'(\alpha) = 2\, n^{-1} \sum_{t=0}^{n-1} \left(\sum_{j=0}^{t} h_j'(\alpha) y_{t-j} \right) \left(\sum_{j=0}^{t} h_j(\alpha) y_{t-j} \right)$$

$$\hat{r}_1'(\alpha) = n^{-1} \sum_{t=0}^{n-2} \left(\sum_{j=0}^{t+1} h_j'(\alpha) y_{t-j+1} \right) \left(\sum_{j=0}^{t} h_j(\alpha) y_{t-j} \right)$$
$$ + \; n^{-1} \sum_{t=0}^{n-2} \left(\sum_{j=0}^{t+1} h_j(\alpha) y_{t-j+1} \right) \left(\sum_{j=0}^{t} h_j'(\alpha) y_{t-j} \right) .$$

Applying Lemma 5.1 to these quantities proves the uniform consistency.

References

[1] H. An, Z. Chen, and E. J. Hannan, "The maximum of the periodogram," *J. Multivariate Anal.*, vol. 13, pp.383–400, 1983.

[2] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1081–1089, 1986.

[3] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1987.

[4] M. V. Dragošević and S. S. Stanković, "A generalized least squares method for frequency estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 6, pp. 805–819, 1989.

[5] E. J. Hannan, *Multiple Time Series*, Wiley, New York, 1970.

[6] S. He and B. Kedem, "Higher-order crossings of an almost periodic random sequences in noise," *IEEE Trans. Information Theory*, vol. 35, no. 3, pp. 360–370, 1990.

[7] S. He and B. Kedem, "The zero-crossing rate of autoregressive processes and its link to unit roots," *J. Time Series Analysis*, vol. 11, no. 3, pp. 201–213, 1990.

[8] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, Academic, New York, 1975.

[9] S. M. Kay, "Accurate frequency estimation at low signal-to-noise ratio," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 3, pp. 540–547, 1984.

[10] B. Kedem, "Contraction mappings in mixed spectrum estimation," Presented at Inst. for Mathematics and Its Applications, Univ. of Minnesoda, Minneapolis, Minnesoda, 1990.

[11] B. Kedem and T. Li, "Monotone gain, first-order autocorrelation and zero-crossing rate," *Ann. Statist.*, to appear, 1991.

[12] B. Kedem and S. Yakowitz, "On the contraction mapping method for frequency estimation," submitted, 1991.

[13] R. Kumaresan, L. L. Scharf, and A. K. Shaw, "An algorithm for pole-zero modeling and spectral analysis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 637-640, 1986.

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

[15] T. Li and B. Kedem, "Asymptotic normality of the contraction mapping estimators for frequency estimation," in preparation, 1991.

[16] B. G. Quinn and J. M. Fernandes, "A fast efficient technique for the estimation of frequency," Dept. of Statist., Univ. of Newcastle, Australia, Tech. Rep. EE8851, 1988.

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

[18] S. Yakowitz, "Some contribution to a frequency location method due to He and Kedem," *IEEE Trans. Information Theory*, vol. 37, no. 4, pp. 1177–1181, 1991.