The flatness of power spectral zeros and their significance in quadratic estimation

Yongxin Chen and Tryphon T. Georgiou

Abstract—In optimal prediction as well as in optimal smoothing the variance of the optimal estimator is impacted predominantly by the frequency segments where the power spectrum is small or negligible. Indeed, the Szegő’s celebrated theorem characterizes deterministic process as those whose power spectral density fails to be log-integrable by virtue of sufficiently flat spectral zeros. Likewise Kolmogorov’s formula gives an analogous condition for optimal smoothing. We discuss how the flatness of spectral zeros suggests a nested stratification of families of spectral where estimation of a stochastic process over a window of a given size is possible with negligible variance based on observations outside the interval. We then focus on the more general problem of estimating missing data in observation records which are not necessarily contiguous. A key result in the paper (Theorem 3) provides a sufficient condition for being able to estimate missing data with arbitrarily small error variance, in terms of the flatness of the spectral zeros.

I. INTRODUCTION

In optimal prediction the least variance of the estimation error is given by Szegő’s celebrated formula [2, page 176]
\[ \exp \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\theta) d\theta \right), \]
in terms of the spectral density \( f \) of a discrete-time stochastic process. The variance vanishes and, in this case, the process is termed deterministic, when \( f \) fails to be log-integrable due to a “sufficiently flat” zero. Likewise, Kolmogorov’s formula for the variance of the optimal smoothing error [11, page 68]
\[ \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} f^{-1}(\theta) d\theta \right)^{-1} \]
shows that the smoothing error vanishes when a less stringent condition on a spectral zero is in place, namely, in this case, \( f^{-1} \) must fail to be integrable; evidently, in this case, more data is available than in the Szegő problem. It is thus of interest to study how spectral zeros, in conjunction with the extent of the available observation record, affect the estimation error.

In Section II we review notation and basic results from prediction and estimation theory for second-order discrete-time random processes. In Section III we study cases where the observation record misses values over a window which is indexed by a contiguous subset of the integers \( \mathbb{Z} \). Three important cases are being discussed, referred to as the Szegő, Kolmogorov, and Yaglom processes. They correspond to the cases where data is unavailable over the indexing sets \( \{0, 1, \ldots\}, \{0\}, \) and \( \{0, 1, \ldots, n\} \), respectively. In Section IV we consider the case where the unavailable data is indexed by not necessarily continuous regions on integers. We show in Theorem 3 that it is the size of the unavailable data, rather than the particular distribution of the indexing set, that determines the “flatness” of spectral zeros necessary to estimate this part of the record with perfect accuracy. We conclude with an academic example that highlights the insights gained by the theory.

II. PRELIMINARIES

We review some standard notation and the framework for optimal estimation [1], [2]. Throughout \( \mathbb{Z}, \mathbb{R} \) and \( \mathbb{C} \) denote integers, real numbers and complex numbers, respectively, \( \mathbb{C}^n \) denotes column vectors with entries in \( \mathbb{C} \), and \( L^1(d\theta) \) denotes the space of integrable functions with respect to Lebesgue measure on the unit circle.

We consider a scalar, zero-mean stationary random process \( \{X_k : k \in \mathbb{Z}\} \). We denote by \( r_k = E\{X_kX_0^*\}, \) \( k \in \mathbb{Z} \) the correlation coefficients and by \( d\mu \) its spectral measure. As usual, \( ^* \) denotes complex conjugate (transpose), \( E \) denotes the expectation operator, while
\[ r_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\theta} d\mu(\theta), \quad k \in \mathbb{Z}. \]
The spectral measure \( d\mu = d\mu_a + d\mu_s \), where \( d\mu_a \) is the absolutely continuous part of the power spectrum and \( d\mu_s \) is the singular part (which includes spectral lines). The spectral density function is the derivative of \( d\mu_a \), i.e.,
\[ f(\theta) = \frac{d\mu_a(\theta)}{d\theta}. \]
Throughout we assume that \( f \in L^1(d\theta) \). We denote by \( \mathcal{X} := span\{X_k : k \in \mathbb{Z}\} \) the closure of all finite linear combinations of \( \{X_k\} \) with respect to the inner product
\[ \langle \sum_k a_kX_k, \sum_\ell b_\ell X_\ell \rangle := E\{(\sum_k a_kX_k)(\sum_\ell b_\ell X_\ell)^*\} = \sum_{k,\ell} a_kr_{k-\ell}b_{\ell}. \]
This is Hilbert-space isometric to \( L^2(d\mu/2\pi) \), the space of square integrable functions on the unit circle with respect to \( d\mu/2\pi \). The inner product is
\[ \langle a, b \rangle_{d\mu/2\pi} := \int_{-\pi}^{\pi} a(\theta)b(\theta)^* \frac{d\mu(\theta)}{2\pi}, \]
where \(a(\theta) = \sum_k a_k e^{i k \theta}\) and \(b(\theta) = \sum_l b_l e^{i l \theta}\). And the isometric isomorphism is
\[
\mathcal{X} \to L^2(d\mu/2\pi) : \sum_k a_k X_k \mapsto \sum_k a_k e^{i k \theta}.
\]

III. Estimation over an interval

We consider estimating \(X_k\) over an indexing set \(I \subset \mathbb{Z}\) using observations over the complement of \(I\), namely,
\[
I^c := \mathbb{Z} \backslash I.
\]
In particular, we are interested in the question of when the elements of
\[
\mathcal{X}_I := \text{span}\{X_k : k \in I\}
\]
can be approximated with arbitrarily small variance by elements of \(\mathcal{X}_{I^c}\). It is clear that this question amounts to requiring that
\[
\mathcal{X}_{I^c} \subseteq \mathcal{X}.
\]

There are three special problems that have been studied in the literature corresponding to \(I\) being
\[
\{0, 1, 2, \ldots\}, \{0\}, \text{ or } \{0, 1, \ldots, n\}.
\]
These will be referred to as the Szegö, Kolmogorov, and Yaglom problems, respectively [3], [4], [5]. In general, for any given linear combination of “missing values” \(\sum_{k \in I} a_k X_k\), we are interested in the least-variance linear estimator \(\sum_{k \in I^c} b_k X_k\) minimizing
\[
\mathcal{E}\{\left|\sum_{k \in I} a_k X_k - \sum_{k \in I^c} b_k X_k\right|^2\}
\]
and analyze the case where the infimal is zero.

Accordingly, we define
\[
\mathcal{L}_I = \text{span}\{e^{i k \theta} : k \in I\} \subset L^2(d\mu/2\pi)
\]
and consider the case where, for any \(a(\theta) = \sum_{k \in I} a_k e^{i k \theta}\),
\[
\text{dist}(a, \mathcal{L}_{I^c}) := \min_{b \in \mathcal{L}_{I^c}} \|a - b\|_{d\mu/2\pi} = 0.
\]
Without loss of generality, we may assume that the singular part of the spectrum is 0, that is, \(d\mu(\theta) = f(\theta)d\theta\). This assumption does not affect the estimation error since the singular spectrum \(d\mu_s\) corresponds to a deterministic component [6, page 46].

A. Szegö’s prediction

For \(I = \{0, 1, 2, \ldots\}\), Szegö’s celebrated theorem (see [6]) states that
\[
\text{dist}(1, \mathcal{L}_{I^c}) = \sqrt{\frac{1}{2\pi} \int_\pi^\pi \log f(\theta)d\theta}
\]
when \(\log f \in L^1(d\theta)\), otherwise \(\text{dist}(1, \mathcal{L}_{I^c}) = 0\). Furthermore, when
\[
\log f \notin L^1(d\theta)
\]
for any \(a \in \mathcal{L}_I\),
\[
\text{dist}(a, \mathcal{L}_{I^c}) = 0.
\]
Clearly, (3b) implies (3a). The converse follows from the fact that (3a) implies that \(\text{dist}(e^{i k \theta}, \mathcal{L}_I) = 0\) for any \(k \geq 0\).

Condition (3a) represents the vanishing of \(f(\theta)\) about some point on \([-\pi, \pi]\), in a suitable sense, so as to render \(\log f\) not integrable. In fact, \(f\) needs to be “sufficiently flat” about such a “zero.” Rational spectral densities do not satisfy this condition, e.g., \(2 + 2 \cos \theta\) is log-integrable. An example where (3a) holds is the seemingly low-frequency power spectral density \(f(\theta) = e^{-\theta^2/(\pi^2 - \theta^2)}\).

B. Komogorov’s estimation

For \(I = 0\), Kolmogorov’s formula for the “smoothing” error [7], [8] states that
\[
\text{dist}(1, \mathcal{L}_{I^c}) = \sqrt{\frac{1}{2\pi} \int_\pi^\pi (f(\theta)^{-1}d\theta)^{-1}}.
\]
when \(1/f \in L^1(d\theta)\), otherwise \(\text{dist}(1, \mathcal{L}_{I^c}) = 0\). Thus, the condition
\[
1/f \notin L^1(d\theta)
\]
is equivalent to
\[
\text{dist}(1, \mathcal{L}_{I^c}) = 0.
\]
Condition (4a) represents the vanishing of \(f(\theta)\) about some point on \([-\pi, \pi]\), in a suitable sense, so as to render \(1/f\) not integrable. The “flatness” of the zero in this case is less severe than that in the Szegö problem. “Rational zeros” do satisfy this condition, e.g., \(f(\theta) = 2 + 2 \cos \theta\) is such that \(1/f\) is not integrable.

C. Yaglom’s estimation

For \(I = \{0, 1, \ldots, n\}\), it can be shown (see Appendix I) that
\[
\text{dist}(a, \mathcal{L}_{I^c}) = \sqrt{a^* \Gamma^{-1} a},
\]
where
\[
a(\theta) = a_0 + a_1 e^{i \theta} + \cdots + a_n e^{i n \theta} \in \mathcal{L}_I,\]
\[
a = [a_0, a_1, \ldots, a_n]^T \in \mathbb{C}^{n+1},\]
\[
\Gamma = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_n \\ \gamma_{-1} & \gamma_0 & \cdots & \gamma_{n-1} \\ \vdots & \ddots & \ddots & \vdots \\ \gamma_{-n} & \gamma_{-n+1} & \cdots & \gamma_0 \end{bmatrix},\]
\[
\gamma_k = \frac{1}{2\pi} \int_\pi^\pi e^{-i k \theta} f(\theta)d\theta,
\]
assuming \(1/f \in L^1(d\theta)\). Here, \((\cdot)^T\) denotes the transpose.

Turning to approximating perfectly any element in \(\mathcal{L}_I\) by \(\mathcal{L}_{I^c}\), it can be shown that if
\[
\int_\pi^\pi |g_1(\theta)|^2 f(\theta)d\theta = \infty, \text{ for all } g_1(\theta) = g_0 + g_1 e^{i \theta} + \cdots + g_n e^{i n \theta} \neq 0,
\]
then
\[
\text{dist}(a, \mathcal{L}_{I^c}) = 0 \text{ for all } a \in \mathcal{L}_I.
\]
In fact, this is also a necessary condition [8], [9], [10].
The framework where an arbitrary set \( I \subset \mathbb{Z} \), of finite length, indexes missing values was considered by Yaglom [5]. The expressions derived in his paper quantified the variance in estimating each of the corresponding random variables, separately, based on values on the complement \( I^c \).

Condition (6a) represents the vanishing of \( f(\theta) \) about some point on \([-\pi, \pi]\), in the sense that for any trigonometric polynomial \(|g_t(\theta)|^2 \neq 0\) (i.e., for any set of corresponding “rational” zeros) the ratio \(|g_t(\theta)|^2 / f(\theta)\) fails to be integrable. An example where this is the case for \( n = 2 \) is \( f(\theta) = (2 + 2 \cos \theta)^2 \).

D. Deterministic processes across window of arbitrary size

We have just seen conditions on the power spectral density that ensure the corresponding stochastic process to be \( I\)-deterministic for some indexing set \( I \), i.e., \( X_{I^c} = X \). That is, its values over the window \( I \) can be estimated arbitrarily accurately using values over the complement indexing set. These characterizations suggest a stratification of power spectra according to the size of the largest window of “missing data” that can be estimated perfectly using values over the complement indexing set. Thus, we denote by \( D_n \) the set of spectral densities \( f \) such that satisfy (6a) for the specific \( n \geq 0 \). It is easy to see that \( D_{n+1} \subset D_n \) for \( n \geq 0 \).

When the window of missing data can be arbitrarily large, but perfect estimation is still possible based on data that are available on both sides, we denote the corresponding set of spectral densities by \( D_\infty \). Clearly,

\[
D_\infty = \bigcap_{n=0}^{\infty} D_n.
\]

Finally, we denote the Szegö class of deterministic spectral densities by \( D := \{ f : \log f \notin L^1(d\theta) \} \).

Since \( \{0, 1, \ldots\} \supset \{0, 1, \ldots, n\} \) for any \( n \geq 0 \),

\[
D \subset D_\infty.
\]

It is an interesting observation that

\[
D \neq D_\infty.
\]

To see this consider \( \exp(-1/\sqrt{\theta}) \). This is log-integrable and hence it does not belong to \( D \). On the other hand

\[
\int_{-\pi}^{\pi} \frac{|g_t(\theta)|^2}{\exp(-1/\sqrt{|\theta|})} d\theta \to \infty
\]

for any \( n \), as can be seen from the fact that \(|g_t|^2\) can only provide at most \( 2n \) roots at \( \theta = 0 \) while the integral

\[
\int_{-\pi}^{\pi} g_{2n} e^{1/\sqrt{\theta}} d\theta = \infty.
\]

Thus, \( \exp(-1/\sqrt{|\theta|}) \) belongs to \( D_\infty \) but not to \( D \). It is seen that the “flatness” of vanishing of \( \exp(-1/\sqrt{|\theta|}) \) at \( \theta = 0 \) is the distinguishing feature.

Finally, a typical example of an \( f \in D_n \) is the product of a log-integrable function with a non-negative trigonometric polynomial of degree \( n + 1 \) having all its roots on the boundary of the unit disc. Thus, these families are ordered as follows, \( D \subset D_\infty \subset \ldots \subset D_1 \subset D_0 \), and none of these relations holds with equality.

IV. Estimation over non-contiguous indices

We now consider estimating \( X_k \) over an indexing set \( I \), as before, but without the restriction that this consists of contiguous values, i.e., being an interval. We address the same question as to when the stochastic process is \( I\)-deterministic, and we provide conditions for this to be true in terms of the spectral density. The formulation of Yaglom’s estimation problem and characterization of solutions carry over verbatim. More precisely, for

\[
I = \{0, n_1, \ldots, n_m\}
\]

with \( 0 < n_1 < \ldots < n_m \), the following conditions are equivalent,

\[
\int_{-\pi}^{\pi} \frac{|g_t(\theta)|^2}{f(\theta)} d\theta = \infty, \quad \text{for all} \quad (7a)
\]

\[
g_t(\theta) = g_0 + g_1 e^{i(n_1-1)\theta} + \cdots + g_m e^{i(n_m-1)\theta} \neq 0,
\]

and

\[
\text{dist}(a, L_{I^c}) = 0 \quad \text{for all} \quad a \in L_{I^c}. \quad (7b)
\]

Condition (7a) is quite elegant but not amenable to a finite test, in general. One approach is to consider the minima of a sequence of quadratic forms \( \int_{-\pi}^{\pi} |g_t(\theta)|^2 / (f(\theta) + \epsilon) d\theta \) for values of \( \epsilon > 0 \) tending to zero, and assess whether the limit of the sequence of minima diverges; this is a challenging task both numerically and theoretically. A contribution in this paper is to provide a sufficient condition for a random process to be \( I\)-deterministic which is expressed in terms of only the cardinality \(|I|\) of \( I \). We begin with a key lemma whose proof is rather elementary but lengthy.

Lemma 1: Let \( n_1, n_2, \ldots, n_m \) be \( m \) positive integers satisfying \( n_1 < n_2 < \cdots < n_m \). Then for any set of coefficients \( h_1, h_2, \ldots, h_m \in \mathbb{C} \), the polynomial

\[
h(z) = 1 + h_1 z^{n_1} + h_2 z^{n_2} + \cdots + h_m z^{n_m}
\]

can never have a root of order \( > m \).

Proof: Without loss of generality we assume \( h_k \neq 0 \) for all \( 1 \leq k \leq m \) for, otherwise, \( h(z) \) degenerates to a polynomial with fewer terms.

If \( h(z) \) has a root \( z_0 \neq 0 \) of order \( \geq m + 1 \), then \( h(z + z_0) \) has a root at the origin of order \( \geq m + 1 \). Therefore, the coefficients of \( z^k \), for \( 0 \leq k \leq m \), in \( h(z + z_0) \) must all be equal to zero. I.e.,

\[
h(z + z_0) = c_0 + c_1 z + c_2 z^2 + \cdots + c_m z^m
\]

with

\[
c_0 = c_1 = \cdots = c_m = 0. \quad (8)
\]

Expanding \( h(z + z_0) \) we obtain that

\[
\begin{bmatrix}
    c_0 \\
    c_1 \\
    \vdots \\
    c_m
\end{bmatrix} =
\begin{bmatrix}
    1 + h_1 z_0^{n_1} + \cdots + h_m z_0^{n_m} \\
    h_1 \chi(z_0, n_1, 1) + \cdots + h_m \chi(z_0, n_m, 1) \\
    \vdots \\
    h_1 \chi(z_0, n_1, m) + \cdots + h_m \chi(z_0, n_m, m)
\end{bmatrix},
\]

where

\[
\chi(z, n, m) = \begin{cases} 
  z^{n-m} C_n^m & \text{for } 0 < m \leq n, \\
  0 & \text{for } m > n > 0
\end{cases}
\]
and where
\[ C^m_n = \begin{cases} \frac{n(n-1)\cdots(n-m+1)}{m!} & \text{for } 0 < m \leq n, \\ 0 & \text{for } m > n > 0 \end{cases} \]
is used to denote the number of \( m \)-combinations out of a set with \( n \) elements. From (8),
\[
\begin{bmatrix}
C^1_{n_1} & C^1_{n_2} & \cdots & C^1_{n_m} \\
C^2_{n_1} & C^2_{n_2} & \cdots & C^2_{n_m} \\
\vdots & \vdots & \ddots & \vdots \\
C^m_{n_1} & C^m_{n_2} & \cdots & C^m_{n_m}
\end{bmatrix}
\begin{bmatrix}
h_1 z_0^{n_1} \\
h_2 z_0^{n_2} \\
\vdots \\
h_m z_0^{n_m}
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix}. \quad (9)
\]
The coefficient matrix on the left hand side, which we now denote by \( \mathbf{C} \) is nonsingular. To see this note that
\[
\mathbf{C} = \begin{bmatrix}
1 & \cdots & 1 \\
\frac{1}{n_1!} & \cdots & \frac{1}{n_m!}
\end{bmatrix} \mathbf{C}_1 \begin{bmatrix}
n_1 & \cdots & n_m
\end{bmatrix}
\]
where
\[
\mathbf{C}_1 = \begin{bmatrix}
\prod_{k=0}^{n_1-1} (n_1-k) & \cdots & 1 \\
\prod_{k=0}^{n_2-1} (n_2-k) & \cdots & 1 \\
\vdots & \ddots & \vdots \\
\prod_{k=0}^{n_m-1} (n_m-k) & \cdots & 1
\end{bmatrix}.
\]
We now claim that
\[
\det(\mathbf{C}_1) = \prod_{1 \leq k < m} (n_k - n_\ell).
\]
To prove this we first note that
i) \( \det(\mathbf{C}_1) \) is polynomial of degree \( m-1 \) in each of the variables \( n_1, \ldots, n_m \).
ii) \( \det(\mathbf{C}_1) \) vanishes if any two of \( n_1, \ldots, n_m \) are identical.

It follows that
\[
\det(\mathbf{C}_1) = K \prod_{1 \leq k < m} (n_k - n_\ell),
\]
for some constant \( K \), since the right hand side is the only polynomial in \( n_1, \ldots, n_m \) that satisfies i-ii). To determine the value of \( K \) simply evaluate the determinant for the case where
\[ n_k = k, \text{ for } 1 \leq k \leq m. \]
In this case, \( \mathbf{C}_1 \) is upper triangular with \( 0! , 1! , \ldots, (m-1)! \) on the diagonal. Hence,
\[
\det(\mathbf{C}_1) = \prod_{k=0}^{m-1} k !
\]
which coincides with
\[
\prod_{1 \leq k < m} (n_k - n_\ell) = \prod_{1 \leq k < m} (k - \ell).
\]
Therefore \( K = 1 \). Thus, we conclude that \( \det \mathbf{C} \neq 0 \).

Since \( \mathbf{C} \) is invertible, the only solution of (9) is the zero solution. Therefore,
\[
c_0 = 1 + h_1 z_0^{n_1} + \cdots + h_m z_0^{n_m} = 1 \neq 0,
\]
which contradicts (8). This completes the proof. \( \square \)

The following is an easy corollary.

**Corollary 2:** Let \( n_1, n_2, \ldots, n_m \) be \( m \) positive integers satisfying \( n_1 < n_2 < \cdots < n_m \). Then, for any set of coefficients \( g_0, g_1, \ldots, g_m \), the trigonometric polynomial
\[
g(\theta) = g_0 + g_1 e^{i n_1 \theta} + \cdots + g_m e^{i n_m \theta} \neq 0
\]
on \( [-\pi, \pi] \) cannot have a root of order \( > m \).

**Proof:** Once again, without loss of generality we assume \( g_k \neq 0 \) for \( 1 \leq k \leq m \). Otherwise \( g(\theta) \) would degenerate to a trigonometric polynomial with fewer terms. Consider the first term \( g_0 \). If \( g_0 = 0 \), then
\[
g(\theta) = g_1 e^{i n_1 \theta} \left( 1 + \frac{g_2}{g_1} e^{i(n_2-n_1)\theta} + \cdots + \frac{g_m}{g_1} e^{i(n_m-n_1)\theta} \right).
\]
Since \( g_1 e^{i n_1 \theta} \neq 0 \) for any \( \theta \), \( g(\theta) \) has the same roots as
\[
1 + \frac{g_2}{g_1} e^{i(n_2-n_1)\theta} + \cdots + \frac{g_m}{g_1} e^{i(n_m-n_1)\theta},
\]
which is of the same form but with fewer terms. Hence, without loss of generality we may assume that \( g_0 \neq 0 \) as well.

Now we consider the polynomial
\[
h(z) = g_0 + g_1 z^{n_1} + g_2 z^{n_2} + \cdots + g_m z^{n_m}.
\]

By Lemma 1, \( h(z) \) cannot have a root of order \( > m \). In particular, it cannot have a root of order \( > m \) on the unit circle. Every root of \( h(z) \) on the unit circle corresponds to a root of \( g(\theta) \) of the same order, since
\[
|e^{i \theta} - 1| \approx |\theta|.
\]
This completes the proof. \( \square \)

Below we consider \( f \) to be, as usual, the spectral density of a random process. At a place on \( [-\pi, \pi] \) where \( f \) vanishes, say \( \theta_0 \), we define the order of this zero to be
\[
\alpha := \inf \{ \beta > 0 : \limsup_{\theta \to \theta_0} \frac{|\theta - \theta_0|^\beta}{f(\theta)} < \infty \}.
\]
In case the \( \limsup \) condition is not feasible, i.e., the set given above is empty, it is standard to assign as infimal value \( +\infty \).

We are now in a position to state the following main result.

**Theorem 3:** Let \( I \subset \mathbb{Z} \) have cardinality \( |I| \leq m \). Let \( f \) be the spectral density of a random process, and let \( \mathcal{L}_I, \mathcal{L}_{I'} \) be as before, cf. (2). If \( f \) has a zero of order \( \alpha > 2m-1 \), then
\[
\text{dist}(a, \mathcal{L}_{I'}) = 0 \text{ for any } a \in \mathcal{L}_I.
\]

**Proof:** By Corollary 2, any
\[
g_I(\theta) = g_0 + g_1 e^{i n_1 \theta} + \cdots + g_m e^{i n_m \theta} \neq 0
\]
cannot have a zero of order \( > m - 1 \). It follows that
\[
\limsup_{\theta \to \theta_0} \frac{|g_I(\theta)|^2}{f(\theta)} |\theta - \theta_0| = \infty.
\]
We combine this with the fact that
\[
\int_0^\pi \frac{1}{\theta} d\theta = \infty
\]
for any $\epsilon > 0$ to obtain
\[ \int_{-\pi}^{\pi} \frac{|g_I(\theta)|^2}{f(\theta)} d\theta = \infty. \]
By (7a) we conclude that this process is $I$-deterministic and this completes the proof.

V. EXAMPLE

We now highlight the possible conclusions that one can draw for a particular stochastic process and possible efficacy of estimators based on the flatness of its spectral zeros.

We consider a moving average process
\[ X_k = W_k + 2W_{k-1} + W_{k-2} \]
with \{\{W_k\}\} a sequence of independent $N(0, 1)$ random variables (white noise). The spectral density of \{\{X_k\}\} is
\[ f(\theta) = (2 + 2 \cos \theta)^2. \]
The one-step-ahead prediction error is clearly $\geq 1$ because $W_k$ is orthogonal to span \{\{X_{k-1}, X_{k-2}, \ldots\}\}, and therefore \{\{X_k\}\} is not $I$-deterministic for $I = \{0, 1, 2, \ldots\}$. (It can be shown that the one-step-ahead prediction error is in fact equal to 1.) On the other hand, since $f(\theta)$ has a zero of order 4 at the origin, \{\{X_k\}\} is $I$-deterministic for all $I$ with cardinality $|I| \leq 2$ by Theorem 3.

To elaborate on the effect of missing values, we choose two specific indexing sets $I = \{0, 2\}$ with $|I| = 2$, and $J = \{0, 2, 20\}$ with $|J| = 3$. We then numerically compute and compare the variance of optimal estimates of $X_0$ based on a symmetrically expanding window of available observations, from $-n$ to $n$, for a range of $n$. These are shown in Figure 1. For the case where the missing data are indexed by $I$ the variance goes to zero since $|I| = 2$. For the case where there are three missing data, and the indexing set is $J$, the variance levels off and remains bounded away from zero. A rigorous proof of this fact can be constructed by expanding on the arguments in Lemma 1 to show that there is trigonometric polynomial $g_0 + g_1 e^{i2\theta} + g_2 e^{i2\theta}$ with a double root at $\theta = 0$ and, then, setting $g_I$ to this polynomial violates (7a).

VI. CONCLUDING REMARKS

The current paper, as well as a substantial portion of classical literature on the subject, deal with an idealistic scenario where observation records could be infinitely long and estimators are defined in a quadratic sense. In practice, one may be concerned with the decay-rate of the coefficients that are weighing in observations on either side of the time axis. Alternatively, other norms may be placed on the relevance of remote observations and, accordingly, stronger restrictions may be placed on the type of estimators that are desirable or even practical. Yet, it is clear that the part of the frequency spectrum where power is negligibly small will have a predominant role on the efficacy of estimators. It will be interesting and important to investigate further potential tradeoffs amongst the aforementioned elements. In particular, it will be important to consider optimal and suboptimal estimators for stochastic processes that are nearly deterministic. The “memory” reflected in potential long-time-constants may be significant; one would expect that robustness could be compromised. Thus, a possible direction for further research is to study the relevance of spectral zeros in the context of robust filtering and min-max estimation [12, Chapter 7].

VII. APPENDIX I

We first show (5). We note that
\[ \min_{b \in \mathcal{L}_I} \|a - b\|_{d\mu/2\pi} = \min_{v \in \mathcal{V}} \|v\|_{d\mu/2\pi}, \]
where $\mathcal{V} := \{a + b \mid b \in \mathcal{L}_I\}$, or equivalently,
\[ \min \left\{ \|v\|_{d\mu/2\pi} : \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\theta} v(\theta)^* d\theta = a_k, \; k = 0, \ldots, n \right\} = \min \left\{ \|v\|_{d\mu/2\pi} : \langle d_k, v \rangle = a_k, \; k = 0, \ldots, n \right\} \]
where $d_k(\theta) = e^{ik\theta} f(\theta)$ and $\langle \cdot, \cdot \rangle$ is with respect to $d\mu/2\pi$. This problem has the unique solution given by
\[ v = \sum_{k=0}^{n} y_k d_k \]
where $y = [y_0, y_1, \ldots, y_n]^*$ is a solution of $\Gamma y = a$ and
\[ \Gamma = \begin{bmatrix} \langle d_0, d_0 \rangle & \langle d_0, d_1 \rangle & \cdots & \langle d_0, d_n \rangle \\ \langle d_1, d_0 \rangle & \langle d_1, d_1 \rangle & \cdots & \langle d_1, d_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle d_n, d_0 \rangle & \langle d_n, d_1 \rangle & \cdots & \langle d_n, d_n \rangle \end{bmatrix}. \]
This is the same as the matrix $\Gamma$ in (5) which is a Toeplitz matrix with entries the correlation sequence of $1/f$. Since $f$ is integrable, $\int \log(1/f) > -\infty$ and therefore, the Toeplitz matrix $\Gamma$ is non-singular as it corresponds to a non-deterministic stochastic process [2, page 148]. It follows that
\[ \|v\|_{d\mu/2\pi} = a^* \Gamma^{-1} a. \]
Next, we show that (6a) implies (6b). For any $\epsilon > 0$, define
\[
\Gamma_\epsilon = \begin{bmatrix}
\gamma_0^\epsilon & 0 & \gamma_1^\epsilon & \cdots & \gamma_n^\epsilon \\
\gamma_{-1}^\epsilon & \gamma_0^\epsilon & \gamma_1^\epsilon & \cdots & \gamma_{n-1}^\epsilon \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
\gamma_{-n}^\epsilon & \gamma_{-n+1}^\epsilon & \cdots & \gamma_0^\epsilon & \gamma_n^\epsilon
\end{bmatrix}
\]
with $\gamma_k^\epsilon = \int_{-\pi}^{\pi} e^{-ik\theta}/(f(\theta) + \epsilon)d\theta/2\pi$. It's easy to see that
\[
\Gamma_{\epsilon_1} > \Gamma_{\epsilon_2} > 0 \text{ when } \epsilon_1 < \epsilon_2.
\] (11)

We claim that
\[
\lim_{\epsilon \searrow 0} \sigma_{\min}(\Gamma_\epsilon) = \infty,
\] (12)

where $\sigma_{\min}(\Gamma_\epsilon)$ denotes the minimum singular value of $\Gamma_\epsilon$. Assume (12) is not true. Then there exist a $B > 0$, a decreasing sequence of positive numbers $\{\epsilon_k\}$, and a sequence $\{g_k\}$ in $C^{n+1}$ with $\|g_k\| = 1$ such that
\[
g_k^* \Gamma_{\epsilon_k} g_k < B
\] (13)

for all $k = 1, 2, \ldots$. Since the unit sphere in $\mathbb{R}^{2n+2}$ is compact, $\{g_k\}$ has a convergent subsequence. So without loss of generality, we assume $g_k$ converges to $g \in C^{n+1}$ with $\|g\| = 1$. Considering (11) and (13) we get
\[
g^* \Gamma_\epsilon g \leq B
\]

for any $\epsilon > 0$. However, on the other hand, if we take $g_1(\theta) = g_0 + g_1 e^{i\theta} + \cdots + g_n e^{in\theta}$ in (6a) where $g_k$ is the $k$th component of row vector $g^*$, following (6a) we obtain
\[
\lim_{\epsilon \searrow 0} \int_{-\pi}^{\pi} \frac{|g_1(\theta)|^2}{f(\theta) + \epsilon} d\theta = \infty.
\]

Therefore,
\[
\lim_{\epsilon \searrow 0} g^* \Gamma_\epsilon g = \lim_{\epsilon \searrow 0} \int_{-\pi}^{\pi} \frac{|g_1(\theta)|^2}{f(\theta) + \epsilon} d\theta = \infty,
\]

which contradicts (13). We conclude that (12) holds true.

Now let us consider the family of functions
\[
v_\epsilon(\theta) = \sum_{k=0}^{n} y_{e,k} e^{ik\theta} f(\theta) + \epsilon
\]
with $y_\epsilon = [y_{e,0}, y_{e,1}, \ldots, y_{e,n}]^* \in \mathbb{C}^{n+1}$ given by
\[
y_\epsilon = \Gamma_\epsilon^{-1} a.
\]

It's not difficult to see that $v_\epsilon$ satisfies the linear constraints in (10). Then,
\[
\|v_\epsilon\|_{d\mu/2\pi}^2 = \frac{1}{2\pi} \sum_{k,\ell} y_{e,k} y_{e,\ell} \int_{-\pi}^{\pi} \frac{e^{i(k-\ell)\theta}}{f(\theta) + \epsilon} d\theta
\]
\[
\leq \frac{1}{2\pi} \sum_{k,\ell} y_{e,k} y_{e,\ell} \int_{-\pi}^{\pi} \frac{e^{i(k-\ell)\theta}}{f(\theta) + \epsilon}
\]
\[
= y_\epsilon^* \Gamma_\epsilon y = a^* \Gamma_\epsilon^{-1} a
\]
\[
\leq a^* a/\sigma_{\min}(\Gamma_\epsilon),
\]

which would go to 0 as $\epsilon \searrow 0$ since $\sigma_{\min}(\Gamma_\epsilon) \to \infty$. The conclusion $\text{dist}(a, L_{\Gamma_\epsilon}) = 0$ follows.

REFERENCES