On the problem of parameter estimation in exponential sums

F. Filbir*H. N. Mhaskar† and J. Prestin‡

Abstract

Let $I \geq 1$ be an integer, $\omega_0 = 0 < \omega_1 < \cdots < \omega_I \leq \pi$, and for $j = 0, \cdots, I$, $a_j \in \mathbb{C}$, $a_{-j} = \overline{a_j}$, $\omega_{-j} = -\omega_j$, and $a_j \neq 0$ if $j \neq 0$. We consider the following problem: Given finitely many noisy samples of an exponential sum of the form

$$\tilde{x}(k) = \sum_{j=-I}^{I} a_j \exp(-i\omega_j k) + \epsilon(k), \qquad k = -2N, \cdots, 2N,$$

where $\epsilon(k)$ are random variables with mean zero, each in the range $[-\epsilon, \epsilon]$ for some $\epsilon > 0$, determine approximately the frequencies ω_j. We combine the features of several recent works to use the available information to construct the moments $\tilde{y}_N(k)$ of a positive measure on the unit circle. In the absence of noise, the support of this measue is exactly $\{\exp(-i\omega_j) : a_j \neq 0\}$. This support can be recovered as the zeros of the monic orthogonal polynomial of an appropriate degree on the unit circle with respect to this measure. In the presence of noise, this orthogonal polynomial structure allows us to provide error bounds in terms of ϵ and N. It is not our intention to propose a new algorithm. Instead, we prove that a preprocessing of the raw moments $\tilde{x}(k)$ to obtain $\tilde{y}_N(k)$ enable us to obtain rigorous performance guarantees for existing algorithms. We demonstrate also that the proposed preprocessing enhances the performance of existing algorithms.

MSC: 94A12, 42C15, 65T40

KEYWORDS: frequency analysis, spectral analysis problem, parameter estimation, exponential sums, prony method

1 Introduction

Many applications in digital signal processing, speech recognition [12], direction finding in array antennas [18, 10], as well as applications in numerical analysis [11] lead to the following question, sometimes called the problem of hidden periodicities (PHP). Let $I \geq 1$ be an integer, $\omega_0 = 0 < \omega_1 < \cdots < \omega_I \leq \pi$, and for $j = 0, \cdots, I$, $a_j \in \mathbb{C}$, $a_{-j} = \overline{a_j}$, $\omega_{-j} = -\omega_j$, and $a_j \neq 0$ if $j \neq 0$. Given finitely many noisy samples of an exponential sum of the form

$$\tilde{x}(k) = \sum_{j=-I}^{I} a_j \exp(-i\omega_j k) + \epsilon(k), \qquad k = -2N, \cdots, 2N, \tag{1.1}$$

where $\epsilon(k)$ are random variables with mean zero, determine approximately the frequencies ω_j. We define the moments

$$x(k) = \sum_{j=-I}^{I} a_j \exp(-i\omega_j k), \qquad k \in \mathbb{Z}, \tag{1.2}$$

*Institute of Biomathematics and Biometry, Helmholtz Center Munich, 85764 Neuherberg, Germany, email: filbir@helmholtz-muenchen.de. The research of this author was supported by Deutsche Forschungsgemeinschaft, grant number FI-883/3-1.

†Department of Mathematics, California State University, Los Angeles, California, 90032, USA, email: hmhaska@calstatela.edu. The research of this author was supported, in part, by grant DMS-0908037 from the National Science Foundation and grant W911NF-09-1-0465 from the U.S. Army Research Office.

‡Institute of Mathematics, University of Lübeck, Wallstraße 40, 23560, Lübeck, Germany, email: prestin@math.uni-luebeck.de.

and assume the definition of $\tilde{x}(k)$ similarly extended to all integer values of k, even though only the values for $k = -2N, \cdots, 2N$ are assumed known. For convenience of notation, we will often write $z_j = \exp(-i\omega_j)$. It is useful to think of a (complex) measure that associates the mass a_j with each z_j. The quantities $x(k)$ are then the trigonometric moments of this measure, and PHP is a special case of the more general trigonometric moment problem. We observe that for every $k \in \mathbb{Z}$,

$$\overline{x(k)} = \sum_{j=-I}^{I} \overline{a_j} \exp(i\omega_j k) = \sum_{j=-I}^{I} a_{-j} \exp(-i\omega_{-j} k) = \sum_{\ell=-I}^{I} a_\ell \exp(-i\omega_\ell k) = x(k).$$

Thus, each $x(k)$ is real.

There are many deterministic as well as statistical algorithms for the solution of this problem, e.g., [11, 3, 4]. Our present paper is motivated by the paper [20] of Potts and Tasche where, in particular, the ESPIRIT algorithm introduced by Roy and Kailath [22] was used for solving the PHP. We wish to show that there are some theoretical and perhaps numerical advantages to use suitably windowed average of the auto–correlation sequence of $\{\tilde{x}(k)\}$ as the input to the known algorithms rather than the raw moments $\tilde{x}(k)$. While the raw moments can be viewed as moments of a complex measure on the unit circle, these windowed averages are moments of a positive measure supported on the points $\{z_j\}$. This enables us to realize the points $\{z_j\}$ as the zeros of an orthogonal polynomial with respect to this measure. These, in turn, can be realized as the eigenvalues of a matrix called CMV matrix. Therefore, a perturbation analysis gives us estimates on the accuracy of the resulting zeros in the presence of noise, without any a priori hypothesis on the exact distribution of the noise variables. We note that the statistical solutions to the PHP typically require some hypothesis about the distribution of the noise variables $\epsilon(k)$.

To explain our ideas in further detail, we first review some existing literature on this subject. For simplicity, the notation introduced in this section may not be the same as in the rest of this paper.

Perhaps, the oldest is an algorithm called Prony's method. To describe this and other methods, we assume in this section only that $a_0 = 0$. Let

$$P(z) = \prod_{j=-I, j\neq 0}^{I} (z - z_j) = \sum_{\ell=0}^{2I} b_\ell z^{2I-\ell}. \tag{1.3}$$

Clearly,

$$\sum_{j=-I}^{I} a_j z_j^{-m} P(z_j) = \sum_{\ell=0}^{2I} b_\ell x(2I - m - \ell) = 0, \qquad m = 0, 1, \cdots, 2I - 1. \tag{1.4}$$

Since $b_0 = 1$, one can solve the system of equations (1.4) (using only the data $x(k)$, $k = -2I+1, \cdots, 2I$) to obtain the coefficients of the polynomial P in (1.3), whose zeros are exactly z_j's. The method utilizes the minimal number of moments $x(k)$, and gives exact results in the absence of noise and in exact arithmetic. However, it is very unstable when these conditions are not satisfied. Modifications of this method to stabilize it are recently proposed by Beylkin and Monzón [1] and Potts and Tasche [20]. In particular, the basic idea in [20] is to use all the known moments to consider a matrix of the form $[x(m + k)]$. This matrix has necessarily one or more eigenvalues equal to 0. In place of P, a polynomial is constructed with a combination of the corresponding eigenvectors as the vector of its coefficients. The roots of this polynomial, when pruned to the right number, yield the required frequencies. Potts and Tasche provide error bounds on the coefficients a_j, assuming a deterministic error in the moments $x(k)$. Our idea is to use the moments $x(k)$ to construct another sequence of moments that corresponds to a positive measure on the unit circle supported on the set $\{z_j\}$ in the absence of noise.

The construction of these moments is motivated by another approach for the solution of PHP, proposed by Jones, Njåstad, and Saff [8]. If μ is any positive measure supported on the points $\{z_j\}_{j=-I, j\neq 0}^{I}$, then the monic polynomial of degree $2I$, orthogonal on the unit circle with respect to μ to all polynomials of a smaller degree, is given by P, defined in (1.3), regardless of what the measure μ is. Jones, Njåstad, and Saff consider an approximation to the measure μ that associates the mass $|a_j|^2$ with each z_j. This approximating measure being absolutely continuous with respect to the Lebesgue measure on the unit circle, the whole sequence of orthogonal polynomials with respect to this measure can be defined. Jones, Njåstad, and Saff find the zeros of the orthogonal polynomials of higher and higher degrees. The $2I$

zeros with the largest magnitude are then the desired approximations to z_j's. The rate of convergence of this process was found in [17] to be $\mathcal{O}(N^{-1})$. This rate was improved by Pan [16], however, at the cost of losing the orthogonal polynomial structure. The main idea in our paper is to use the available data to construct the moments of a positive measure supported on the points $\{z_j\}_{j=-I,j\neq 0}^{I}$. The zeros of the orthogonal polynomial of degree $2I$ are the points $\{z_j\}_{j=-I,j\neq 0}^{I}$ exactly, if there were no noise. Since these zeros are eigenvalues of the CMV matrix [23, Section 4.2], a perturbation analysis for the eigenvalues of matrices can be used as in [20] to estimate the error in the presence of noise. We will do this without making any assumption regarding the distribution of the random variables $\epsilon(k)$.

The idea behind the construction of such a measure originates from a third approach to the problem, proposed in [14, 15]. We consider a suitable filter g; i.e., a smooth, even function on $\mathbb{R}$, vanishing near the origin and outside of $[-1, 1]$, and examine the power pattern; i.e., a trigonometric polynomial

$$\mathbb{P}(x) = \sum_{k=-2N}^{2N} g\left(\frac{|k|}{2N}\right) x(k) \exp(ikx).$$

The location of the maxima of $|\mathbb{P}(x)|$ are approximately the points $\pm\omega_j$ for which $|a_j|$ is maximum. A comparison of $\mathbb{P}(x)$ at these locations with the power pattern with all moments $x(k)$ replaced by 1 enables us to estimate the a_j's. We then subtract these frequencies to obtain a new set of moments, and continue. This process appears to be very stable under noise, and does not require an a priori knowledge of the number of frequencies present in the moments. In practice, we always got the exact locations for ω_j's in this way. However, since $\mathbb{P}$ is a trigonometric polynomial of order $2N$, the theoretically guaranteed accuracy is only $\mathcal{O}(N^{-1})$, and cannot be improved. Nevertheless, an essential aspect of our work is that the kernel

$$\sum_{k=-2N}^{2N} g\left(\frac{|k|}{2N}\right) \exp(ikx)$$

is highly localized around the origin. We utilize this high localization to construct the moments corresponding to the positive measure supported on the points $\{z_j\}_{j=-I}^{I}$. Thereby, we obtain the advantage of the stability of the method in [8], while obtaining exact answers as in the Prony method. In addition, because we have retained the orthogonal polynomial structure in the process, we are able to use perturbation analysis for the corresponding CMV matrices to estimate the errors in the case when the noisy data $\{\tilde{x}(k)\}$ is considered.

To summarize, we propose to replace the moments $\tilde{x}(k)$ of a complex measure on the unit circle by the moments of a positive measure on the unit circle, which can be computed using only the avaiable data. In contrast to the papers [8, 16], this positive measure is a linear combination of the dirac delta measures supported at the points z_j if the number of moments is sufficiently large, rather than an approximation to such a measure. Therefore, in the absence of noise, the points z_j are recovered exactly, while retaining the stability advantages of known algorithms. While the algorithms in [20] retrieve the points very accurately, that paper does not give any error bounds on the results in the presence of noise. The orthogonal polynomial structure allows us to do so. A numerical example worked out at our request by Potts (Example 3) shows that the results in [20] improved significantly by starting with our modified moments as inputs.

In Section 2, we review certain preliminary material concerning orthogonal polynomials on the unit circle, our localized kernel, and recall a version of Höffding's inequality. In Section 3, we consider the problem studied in [8]; i.e., the estimation of ω_j's from $x(k)$'s without noise. The effect of having noisy data will be studied in Section 4. We present a few numerical examples in Section 5, using essentially the idea in [20], mainly to illustrate the feasibility and advantages of starting with our modified moments as the input to the known algorithms. We are grateful to Daniel Potts and Manfred Tasche for their helpful comments and the computation for one of the examples.

2 Review

In this section, we will recall certain facts needed in our paper. In Section 2.1, we review some aspects of the theory of orthogonal polynomials on the unit circle. In Section 2.2, we recall certain basic results about the localized trigonometric polynomial kernels. In Section 2.3, we recall a version of the Höffding's inequality in the theory of probability which will be useful in estimating the perturbations caused by the noise in the moments $\tilde{x}(k)$.

2.1 Orthogonal polynomials on the unit circle

The material in this section is based mainly on [23]. For integer $n \geq 0$, let Π_n denote the class of all complex algebraic polynomials of degree at most n. We find it convenient to extend this notation also when n is not an integer by defining $\Pi_n = \Pi_{\lfloor n \rfloor}$. Let $n \geq 0$ be an integer, and μ be any positive measure supported on the complex unit circle $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ having at least n points in its support. We may use the Gram–Schmidt orthogonalization procedure to obtain monic polynomials $\Phi_k(\mu; z) = z^k + \sum_{j=1}^{k} q_{k,j}(\mu) z^{k-j} \in \Pi_k$ such that

$$\int_{\mathbb{T}} \overline{z}^{\ell} \Phi_k(\mu; z) d\mu(z) = 0, \qquad \ell = 0, \cdots, k-1,\ k = 1, \cdots, n. \tag{2.1}$$

The system of monic polynomials satisfying (2.1) is unique. In particular, if the support of μ consists of n points $\zeta_1, \cdots, \zeta_n$, then $\Phi_n(\mu; z) = \prod_{j=1}^{n}(z - \zeta_j)$, regardless of the measure μ itself. The polynomial $\Phi_k(\mu)$ is called the monic orthogonal polynomial with respect to the measure μ. Associated with these are the polynomials $\Phi_k^*(\mu; z) := 1 + \sum_{j=1}^{k} \overline{q_{k,j}(\mu)} z^j \in \Pi_k$, $k = 0, \cdots, n$. We define the Verblunsky coefficients α_k by

$$\alpha_k(\mu) := \overline{\Phi_{k+1}(\mu; 0)}, \qquad k = 0, \cdots, n-1. \tag{2.2}$$

We have $|\alpha_k(\mu)| < 1$ for $k = 0, \cdots, n-1$, and the following Szegö recurrence holds:

$$\Phi_{k+1}(\mu; z) = z\Phi_k(\mu; z) - \overline{\alpha_k(\mu)} \Phi_k^*(\mu; z), \qquad k = 0, \cdots, n-1. \tag{2.3}$$

For $k = 0, \cdots, n-1$, all the zeros of $\Phi_k(\mu)$ are in the open unit disc $\{z \in \mathbb{C} : |z| < 1\}$. The zeros of $\Phi_n(\mu)$ are in the closure of this disc; i.e., some or all of these zeros may be on $\mathbb{T}$. The zeros of Φ_k are eigenvalues of the CMV matrix $\mathcal{C}^{(k)}(\mu)$ which we now define. First, let $\rho_k(\mu) := (1 - |\alpha_k(\mu)|^2)^{1/2}$, $k = 0, \cdots, n$. Then for $k = 0, \cdots, n$, the CMV matrix (for order k) $\mathcal{C}^{(k)}(\mu)$ is a $k \times k$ matrix such that $\mathcal{C}^{(k)}_{\ell,m} = 0$ if (ℓ, m) is not of the form $(2j - m, 2j)$ or $(2j - m, 2j - 1)$, for $m = 0, \pm 1, \pm 2$. These possibly nonzero entries are defined by

$$\begin{gathered}
\mathcal{C}_{2j-2,2j-1} = \rho_{2j-2}\alpha_{2j-1},\ \mathcal{C}_{2j-1,2j-1} = -\alpha_{2j-2}\alpha_{2j-1},\ \mathcal{C}_{2j,2j-1} = \alpha_{2j}\rho_{2j-1},\ \mathcal{C}_{2j+1,2j-1} = \rho_{2j-1}\rho_{2j}, \\
\mathcal{C}_{2j-2,2j} = \rho_{2j-2}\rho_{2j-1},\ \mathcal{C}_{2j-1,2j} = -\alpha_{2j-2}\rho_{2j-1},\ \mathcal{C}_{2j,2j} = -\alpha_{2j-1}\alpha_{2j},\ \mathcal{C}_{2j+1,2j} = -\alpha_{2j-1}\rho_{2j},
\end{gathered} \tag{2.4}$$

where we have used the abbreviation $\alpha_\ell = \alpha_\ell(\mu)$, $\rho_\ell = \rho_\ell(\mu)$, $\mathcal{C} = \mathcal{C}^{(k)}(\mu)$. We remark that the CMV matrix is defined in [23] as an infinite matrix, and $\mathcal{C}^{(k)}$ is its principal minor of order k.

We remark that since μ has at least n points in its support, there exists a positive constant $Z = Z(\mu)$ such that the following inequality holds:

$$Z \sum_{k=0}^{n-1} |b_k|^2 \leq \int_{\mathbb{T}} \left| \sum_{k=0}^{n-1} b_k z^k \right|^2 d\mu(z), \tag{2.5}$$

for all $b_0, \cdots, b_{n-1} \in \mathbb{C}$.

Next, let $\mathbf{y} = \{y(k)\}_{k \in \mathbb{Z}}$ be any sequence of complex numbers, $n \geq 1$ be any integer, and assume that for any complex numbers $b_0, \cdots, b_{n-1}$,

$$\sum_{j,k=0}^{n-1} b_j \overline{b_k} y(j-k) > 0. \tag{2.6}$$

A well known theorem, known as the Herglotz theorem [9, Theorem 7.6] implies that there exists a (not necessarily unique) positive measure μ on $\mathbb{T}$ such that $y(\ell) = \int_{\mathbb{T}} z^\ell d\mu(z)$, $\ell = -n+1, \cdots, n-1$. The condition (2.6) implies that $\int_{\mathbb{T}} |P(z)|^2 d\mu(z) > 0$ for all $P \in \Pi_{n-1}$. Consequently, the support of the measure μ must contain at least n points on $\mathbb{T}$. When the measure μ is one of the measures constructed from $\mathbf{y}$ as just explained, we will write $\Phi_k(\mathbf{y})$, $\alpha_k(\mathbf{y})$, etc. in place of $\Phi_k(\mu)$, $\alpha_k(\mu)$, etc.

The orthogonal polynomials $\Phi_k(\mathbf{y})$ and the Verblunsky coefficients can be written explicitly in terms of the numbers $\{y(j)\}$. Let

$$D_k(\mathbf{y}) = \det[\overline{y(j-\ell)}]_{j,\ell=0}^k, \qquad k = 0, \cdots, n. \tag{2.7}$$

Then

$$\begin{aligned}
\Phi_k(\mathbf{y}; z) &= D_k(\mathbf{y})^{-1} \begin{vmatrix} y(0) & y(1) & \cdots & y(k) \\ y(-1) & y(0) & \cdots & y(k-1) \\ \vdots & \vdots & \vdots & \\ y(-k+1) & y(-k+2) & \cdots & y(1) \\ 1 & z & \cdots & z^k \end{vmatrix}, \\
\alpha_k(\mathbf{y}) &= (-1)^k \frac{\det[y(j-\ell-1)]_{j,\ell=0}^k}{D_k(\mathbf{y})}.
\end{aligned} \tag{2.8}$$

The zeros of Φ_k are the eigenvalues of $\mathcal{C}^{(k)}(\mathbf{y})$. Of course, we do not mean to suggest these formulas for numerical computations; that is a separate research problem.

2.2 Localized kernel

For a Lebesgue measurable function $F : \mathbb{R} \to \mathbb{R}$, we define

$$\|F\|_\infty = \operatorname*{ess\,sup}_{x \in \mathbb{R}} |F(x)|, \qquad \|F\|_1 = \int_{\mathbb{R}} |F(x)| dx.$$

Let $S \geq 2$ be an integer, $H : \mathbb{R} \to [0,1]$ be an S times continuously differentiable function, $H(t) = H(-t)$, $t \in \mathbb{R}$, $H(t) = 0$ if $|t| \geq 1$. We assume that

$$H_1 = \|H\|_1 = \int_{-1}^{1} H(t) dt > 0.$$

Since $0 \leq H(t) \leq 1$ for all $t \in [-1,1]$, necessarily, $H_1 < 2$. We write

$$\Psi_N(x) := \Psi_N(H; x) := \sum_{m \in \mathbb{Z}} H\left(\frac{|m|}{N}\right) \exp(imx), \qquad x \in \mathbb{R}. \tag{2.9}$$

Then Ψ_N is a real valued, even, trigonometric polynomial of order at most $N-1$.

The following proposition was proved essentially in [14], but was not stated exactly in the following form.

Proposition 2.1 *Let $N \geq 1$ be an integer. If*

$$N \geq \sqrt{\frac{\|H''\|_\infty}{3H_1}}, \tag{2.10}$$

then

$$NH_1/2 \leq \max_{x \in \mathbb{R}} |\Psi_N(x)| = \Psi_N(0) \leq 2N - 1. \tag{2.11}$$

Further,

$$|\Psi_N(x)| \leq 2\frac{\|H^{(S)}\|_1}{H_1}\left\{1 + 2(1-2^{-S})\sum_{k=1}^{\infty} k^{-S}\right\}\frac{\Psi_N(0)}{(N|x|)^S} =: L\frac{\Psi_N(0)}{(N|x|)^S}, \qquad |x| \leq \pi, \ x \neq 0. \tag{2.12}$$

Proof. Since $H(t) \geq 0$ for all $t \in \mathbb{R}$, we have $\max_{x\in\mathbb{R}} |\Psi_N(x)| = \Psi_N(0)$. Since $0 \leq H(t) \leq 1$ for $t \in \mathbb{R}$, and $H(t) = 0$ for $t \geq 1$, $\Psi_N(0) = \sum_{|k|\leq N-1} H(|k|/N) \leq 2N - 1$. This shows the second inequality in (2.11). The sum representing $(1/N)\Psi_N(0)$ is the trapezoidal rule approximation to $\int_{-1}^{1} H(t)dt = H_1$. Hence, the error formula for this rule [21, Eqn. (4.10-12), p. 122] gives

$$|(1/N)\Psi_N(0) - H_1| \leq \frac{\|H''\|_\infty}{6N^2}.$$

Hence, $\Psi_N(0) \geq NH_1 - \frac{\|H''\|_\infty}{6N}$. If N satisfies (2.10), then this leads to the first inequality in (2.11).

Let

$$\hat{H}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} H(x)\exp(ixt)dx, \qquad t \in \mathbb{R}. \tag{2.13}$$

Since H is compactly supported and at least twice continuously differentiable, the following Poisson summation formula holds ([2, Section 5.1.5]):

$$\Psi_N(x) = 2\pi N \sum_{k\in\mathbb{Z}} \hat{H}(N(x + 2\pi k)), \qquad |x| \leq \pi. \tag{2.14}$$

Integrating by parts several times in (2.13), we obtain that

$$|\hat{H}(t)| \leq \frac{1}{2\pi}\|H^{(S)}\|_1 |t|^{-S}, \qquad t \in \mathbb{R},\ t \neq 0.$$

Hence, (2.14) implies that

$$|\Psi_N(x)| \leq \|H^{(S)}\|_1 N \sum_{k\in\mathbb{Z}} \frac{1}{N^S|x + 2\pi k|^S}, \qquad |x| \leq \pi,\ x \neq 0. \tag{2.15}$$

Since $|x| \leq \pi$,

$$|x + 2\pi k| \geq 2\pi|k| - |x| \geq \pi(2|k| - 1) \geq (2|k| - 1)|x|, \qquad k = \pm 1, \pm 2, \cdots.$$

Since $S \geq 2$, $\displaystyle\sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{1}{(2|k| - 1)^S} = 2(1 - 2^{-S})\sum_{k=1}^{\infty} k^{-S}$. Thus, (2.15) implies

$$|\Psi_N(x)| \leq \|H^{(S)}\|_1 \left\{1 + 2(1 - 2^{-S})\sum_{k=1}^{\infty} k^{-S}\right\} \frac{N}{(N|x|)^S}.$$

Therefore, (2.12) follows from the first inequality in (2.11). □

We remark that

$$2\left\{1 + 2(1 - 2^{-S})\sum_{k=1}^{\infty} k^{-S}\right\} \leq 7, \qquad S \geq 2.$$

2.3 Höffding's inequality

We need the following sharper version of the Hoeffding's inequality (cf. [19, p. 191]). It is proved in [7], but not stated in this way. The following formulation is in [13, Proposition 5.1].

Proposition 2.2 *Let (Q, μ) be a probability space, $n \geq 1$ be an integer, and $\{X_k\}$, $k = 1, \cdots, n$ be independent random variables on Q, each with range contained in a compact interval $[a, b]$ and expectation equal to m. Then for any $\delta \in (0, (b - a)/2]$,*

$$\mathit{Prob}\left(\left|n^{-1}\sum_{k=1}^{n} X_k - m\right| \geq \delta\right) \leq 2\exp\left(-\frac{4n\delta^2}{G(b - a)^2}\right), \tag{2.16}$$

where

$$G := \frac{4}{3\log 3 - 2}. \tag{2.17}$$

In particular, if $C > 0$, and n is large enough so that $\log n/n \le (GC)^{-1}$, then

$$\textsf{Prob}\left(\left|n^{-1}\sum_{k=1}^{n} X_k - m\right| \ge \frac{b-a}{2}\sqrt{\frac{GC\log n}{n}}\right) \le 2n^{-C}. \tag{2.18}$$

3 Frequency analysis with exact data

In this section, we follow the lead in [8], and consider the question of finding the frequencies ω_j, given the moments $x(k)$ (without noise) as in (1.2), $k = -2N, \cdots, 2N$. First, we introduce some further notation.

We will write $I^* := 2I$ if $a_0 = 0$ and $I^* := 2I + 1$ if $a_0 \ne 0$. Thus, I^* is the number of distinct frequencies present in $x(k)$'s. Let $\mathcal{F} = \{\pm 1, \cdots, \pm I\}$ if $a_0 = 0$ and $\{0, \pm 1, \cdots, \pm I\}$ if $a_0 \ne 0$. We define

$$q = \min_{\substack{j,k\in\mathcal{F}\\ j\ne k}} |\omega_j - \omega_k|, \quad M := \max_{k\in\mathcal{F}} \frac{1}{|a_k|}\sum_{j\in\mathcal{F}, j\ne k} |a_j|, \tag{3.1}$$

$$A := \sum_{j\in\mathcal{F}} |a_j|, \qquad A_2 := \textstyle\sum_{j\in\mathcal{F}} |a_j|^2. \tag{3.2}$$

We denote the measure that associates the mass $|a_j|^2$ with each z_j, $j \in \mathcal{F}$ by μ. The corresponding moments $y(\ell)$ are defined by

$$y(\ell) := \int_{\mathbb{T}} z^\ell d\mu(z) = \sum_{j\in\mathcal{F}} |a_j|^2 z_j^\ell.$$

The constant Z is defined as in (2.5) for this μ; i.e.,

$$Z\sum_{k=0}^{I^*-1} |b_k|^2 \le \sum_{j\in\mathcal{F}} |a_j|^2 \left|\sum_{k=0}^{I^*-1} b_k z_j^k\right|^2 = \sum_{k,\ell=0}^{I^*-1} b_k \overline{b_\ell}\, y(k-\ell), \qquad b_0, \cdots, b_{I^*-1} \in \mathbb{C}. \tag{3.3}$$

In principle, the number I^* can be calculated as the rank of the matrix $Y := [y(k-\ell)]_{k,\ell=0}^N$. In this discussion, we assume that I^* is known. For $n = 0, \cdots, I^*$, we will denote $\Phi_n = \Phi_n(\mu)$. We will fix a function H as in Section 2.2, and recall the constant L as in (2.12).

Theorem 3.1 *Let $N \ge 1$ be an integer,*

$$y_N(\ell) := \frac{1}{\Psi_N(0)}\sum_{m\in\mathbb{Z}} H\left(\frac{|m|}{N}\right) x(m)\frac{x(m+\ell) + x(m-\ell)}{2}, \qquad \ell \in \mathbb{Z}, \tag{3.4}$$

and $\mathbf{y}_N := \{y_N(\ell)\}_{\ell\in\mathbb{Z}}$.
(a) *Each $y_N(\ell)$ is real, and $y_N(-\ell) = y_N(\ell)$, $\ell \in \mathbb{Z}$. Let*

$$\lambda_{k,N} := \Re e\left\{a_k \sum_{j\in\mathcal{F}} \overline{a_j}\frac{\Psi_N(\omega_k - \omega_j)}{\Psi_N(0)}\right\}, \qquad k \in \mathcal{F}. \tag{3.5}$$

Then

$$y_N(\ell) = \sum_{k\in\mathcal{F}} \lambda_{k,N} \exp(-i\omega_k \ell) = \sum_{k\in\mathcal{F}} \lambda_{k,N} z_k^\ell, \qquad \ell \in \mathbb{Z}. \tag{3.6}$$

(b) *Let $N \ge I^*$ satisfy (2.10). Then*

$$|\lambda_{k,N} - |a_k|^2| \le \frac{LM|a_k|^2}{(Nq)^S}, \qquad k \in \mathcal{F}, \tag{3.7}$$

and
$$|y_N(\ell) - y(\ell)| \le \frac{LMA_2}{(Nq)^S}. \tag{3.8}$$
In particular, if
$$N \ge q^{-1}(2LM)^{1/S}, \tag{3.9}$$
then $\lambda_{k,N} \ge (1/2)|a_k|^2 > 0$*,* $k \in \mathcal{F}$*, and*
$$(Z/2)\sum_{j=0}^{I^*-1}|b_j|^2 \le \sum_{j,\ell=0}^{I^*-1} b_j\overline{b_\ell}y_N(j-\ell), \qquad b_0,\cdots,b_{I^*-1}\in\mathbb{C}. \tag{3.10}$$
Thus, the monic orthogonal polynomial $\Phi_{I^*,N} := \Phi_{I^*}(\mathbf{y}_N)$ *exists, and* $\{z_j\}_{j\in\mathcal{F}}$ *is the set of its zeros.*

We observe that if $x(k)$'s are known for $k = -2N,\cdots,2N$, then the formula (3.4) allows us to calculate $y_N(\ell)$'s for $\ell = -N,\cdots,N$. In turn, the polynomial $\Phi_{I^*,N}$ can be calculated from the moments $y_N(\ell)$, $|\ell| \le 2I^*$; i.e., using the moments $x(k)$, $|k| \le N + 2I^* - 1$.

Proof of Theorem 3.1. Let
$$w_{k,N} = a_k\sum_{j\in\mathcal{F}}\overline{a_j}\frac{\Psi_N(\omega_k-\omega_j)}{\Psi_N(0)}, \qquad k\in\mathcal{F}, \tag{3.11}$$
so that
$$\lambda_{k,N} = \frac{w_{k,N}+\overline{w_{k,N}}}{2}, \qquad k\in\mathcal{F}.$$

Let $k \in \mathcal{F}$. We observe first that
$$\overline{w_{k,N}} = \overline{a_k}\sum_{j\in\mathcal{F}}a_j\frac{\Psi_N(\omega_k-\omega_j)}{\Psi_N(0)} = a_{-k}\sum_{j\in\mathcal{F}}\overline{a_{-j}}\frac{\Psi_N(-\omega_{-k}+\omega_{-j})}{\Psi_N(0)} = a_{-k}\sum_{\ell\in\mathcal{F}}\overline{a_\ell}\frac{\Psi_N(\omega_\ell-\omega_{-k})}{\Psi_N(0)} = w_{-k,N}. \tag{3.12}$$
Next, using the fact that Ψ_N is an even function and $x(m)$'s are all real, we obtain
$$\begin{aligned}
\Psi_N(0)\sum_{k\in\mathcal{F}}w_{k,N}\exp(-i\omega_k\ell) &= \sum_{k\in\mathcal{F}}a_k\exp(-i\omega_k\ell)\sum_{j\in\mathcal{F}}\overline{a_j}\Psi_N(\omega_k-\omega_j)\\
&= \sum_{k\in\mathcal{F}}a_k\exp(-i\omega_k\ell)\sum_{j\in\mathcal{F}}\overline{a_j}\Psi_N(\omega_j-\omega_k)\\
&= \sum_{k\in\mathcal{F}}a_k\exp(-i\omega_k\ell)\sum_{j\in\mathcal{F}}\overline{a_j}\sum_{m\in\mathbb{Z}}H\left(\frac{|m|}{N}\right)\exp(im\omega_j-im\omega_k)\\
&= \sum_{m\in\mathbb{Z}}H\left(\frac{|m|}{N}\right)\sum_{j\in\mathcal{F}}\overline{a_j}\exp(im\omega_j)\sum_{k\in\mathcal{F}}a_k\exp(-i\omega_k(\ell+m))\\
&= \sum_{m\in\mathbb{Z}}H\left(\frac{|m|}{N}\right)\overline{x(m)}x(m+\ell)\\
&= \sum_{m\in\mathbb{Z}}H\left(\frac{|m|}{N}\right)x(m)x(m+\ell).
\end{aligned}$$
Therefore, it follows from (3.12) that
$$\begin{aligned}
\sum_{k\in\mathcal{F}}\lambda_{k,N}\exp(-i\omega_k\ell) &= \frac{1}{2}\left\{\sum_{k\in\mathcal{F}}w_{k,N}\exp(-i\omega_k\ell)+\sum_{k\in\mathcal{F}}w_{-k,N}\exp(-i\omega_k\ell)\right\}\\
&= \frac{1}{2}\left\{\sum_{k\in\mathcal{F}}w_{k,N}\exp(-i\omega_k\ell)+\sum_{k\in\mathcal{F}}w_{-k,N}\exp(-i\omega_{-k}(-\ell))\right\}
\end{aligned}$$

$$= \frac{1}{2}\left\{\sum_{k\in\mathcal{F}} w_{k,N}\exp(-i\omega_k\ell)+\sum_{j\in\mathcal{F}} w_{j,N}\exp(-i\omega_j(-\ell))\right\}$$
$$= \frac{1}{\Psi_N(0)}\sum_{m\in\mathbb{Z}} H\left(\frac{|m|}{N}\right)x(m)\frac{x(m+\ell)+x(m-\ell)}{2}.$$

This proves (3.6).

Next, since N satisfies (2.10), we may use (2.12) and (3.1) to obtain that

$$\left|\sum_{j\in\mathcal{F},j\neq k}\overline{a_j}\frac{\Psi_N(\omega_k-\omega_j)}{\Psi_N(0)}\right| \le \frac{L}{(Nq)^S}\sum_{j\in\mathcal{F},j\neq k}|a_j| \le \frac{LM|a_k|}{(Nq)^S}, \qquad k\in\mathcal{F}.$$

Therefore,

$$|w_{k,N}-|a_k|^2| \le |a_k|\left|\sum_{j\in\mathcal{F},j\neq k}\overline{a_j}\frac{\Psi_N(\omega_k-\omega_j)}{\Psi_N(0)}\right| \le \frac{LM|a_k|^2}{(Nq)^S}, \qquad k\in\mathcal{F}. \tag{3.13}$$

In view of (3.5), this yields (3.7). The estimate (3.8) follows immediately. If N satisfies (3.9) as well, then the fact that $\lambda_{k,N}\ge(1/2)|a_k|^2$ is also clear from (3.7). Together with (3.3), this implies (3.10). The statements regarding the monic orthogonal polynomials follow from the general theory outlined in Section 2.1. □

4 Perturbations

In this section, we consider the general problem to approximate the frequencies ω_j using the noisy data $\tilde{x}(k)$ as defined in (1.1). Let $\{\epsilon(k)\}_{k\in\mathbb{Z}}$ be a sequence of independent random variables, each with mean 0, and each in the (almost) range $[-\epsilon,\epsilon]$ for some $\epsilon>0$. In spite of the symbol, we do not assume that ϵ is “small”. One of the features of our theory is that even if ϵ “overwhelms” the signal strengths $|a_j|$, one can still obtain with high probability a good estimate on the ω_j's provided N is large enough. If the range of $\epsilon(k)$'s is not compact, we need to estimate the tail probability that they will lie in a certain sufficiently large interval, and this probability will then enter into our estimates. Therefore, this assumption is not a serious obstacle to the application of our theory. Let the variance of $\epsilon(k)$ (and hence, of $\tilde{x}(k)$) be $v(k)$, $k=-2N,\cdots,2N$. We write $\tilde{\epsilon}_m(\ell)=(\epsilon(m+\ell)+\epsilon(m-\ell))/2$, $\delta_{0,\ell}=1$ if $\ell=0$ and 0 otherwise. We do not assume that the $\epsilon(k)$'s are identically distributed. We assume that the random variables $\{\epsilon(k)\epsilon(j)\}$ are also independent. If the variances $\{v(k)\}_{k=-N}^N$ are not known, they may be estimated using standard statistical methods by taking several samples of each $\tilde{x}(k)$. This will also help to reduce the variances, as is well known. We will consider the variances $\{v(k)\}_{k=-N}^N$ to be known. First, we will consider the case when I^* is known. At the end, we will discuss some ideas about estimating this, which is a much more delicate task in the presence of noise.

It is natural to expect that the moments $\tilde{y}_N(\ell)$'s can be defined by simply replacing $x(k)$'s by $\tilde{x}(k)$'s in the definition (3.4). However, this is not an unbiased estimator. The correct estimator is given for $\ell=-N,\cdots,N$ by

$$\tilde{y}_N(\ell)=\frac{1}{\Psi_N(0)}\left\{\sum_{m\in\mathbb{Z}}H\left(\frac{|m|}{N}\right)\tilde{x}(m)\frac{\tilde{x}(m+\ell)+\tilde{x}(m-\ell)}{2}-\delta_{0,\ell}\sum_{m\in\mathbb{Z}}H\left(\frac{|m|}{N}\right)v(m)\right\}, \tag{4.1}$$

where H is a fixed function as defined in Section 2.2. We observe that $\tilde{y}_N(\ell)=\tilde{y}_N(-\ell)$ is real for each $\ell\in\mathbb{Z}$. Also, using the stochastic independence of the $\epsilon(k)$'s, it is easy to verify that the expected value of $\tilde{y}_N(\ell)$ is $y_N(\ell)$, $\ell=-N,\cdots,N$. We define $\tilde{\mathbf{y}}_N=\{\tilde{y}_N(\ell)\}_{\ell\in\mathbb{Z}}$.

Theorem 4.1 *Let $C>0$, $N\ge\max(7,I^*)$ be an integer, satisfying in addition to (2.10), (3.9), also*

$$\frac{LMA_2}{8\epsilon(A+\epsilon)(Nq)^S}\le\sqrt{\frac{G(C+2)\log N}{N}}\le 1, \tag{4.2}$$

where G is defined in (2.17). Then with probability exceeding $1-2N^{-C}$, we have

$$|\tilde{y}_N(\ell) - y(\ell)| \leq 16\frac{\epsilon(A+\epsilon)}{H_1}\sqrt{\frac{G(C+2)\log N}{N}}, \qquad \ell = -N, \cdots, N. \tag{4.3}$$

In addition if N also satisfies

$$8I^*\frac{\epsilon(A+\epsilon)}{H_1}\sqrt{\frac{G(C+2)\log N}{N}} \leq Z/4, \tag{4.4}$$

then with probability exceeding $1-2N^{-C}$,

$$(Z/4)\sum_{j=0}^{I^*-1}|b_j|^2 \leq \sum_{j,k=0}^{I^*-1} b_j\overline{b_k}\tilde{y}_N(j-k), \qquad b_0, \cdots, b_{I^*-1} \in \mathbb{C}. \tag{4.5}$$

Thus, the monic orthogonal polynomial $\tilde{\Phi}_{I^*,N} := \Phi_{I^*}(\mathbf{y}_N)$ exists.

The points $\{z_j\}_{j\in\mathcal{F}}$ will be estimated by the zeros of $\tilde{\Phi}_{I^*,N}$. We will discuss this estimation after the proof of Theorem 4.1.

Proof of Theorem 4.1. As expected, we will use Proposition 2.2 in order to prove this theorem. First, let $N \geq 3$ be an integer, satisfying

$$G(C+2)\log N \leq N.$$

We will prove that with probability exceeding $1-2N^{-C}$, we have

$$|\tilde{y}_N(\ell) - y_N(\ell)| \leq 8\frac{\epsilon(A+\epsilon)}{H_1}\sqrt{\frac{G(C+2)\log N}{N}}, \qquad \ell = -N, \cdots, N. \tag{4.6}$$

First, let $\ell \in \{-N, \cdots, N\}$ be fixed, and we define random variables $X_m(\ell)$, $Y_m(\ell)$, $Z_m(\ell)$ by

$$\begin{aligned}
X_m(\ell) &:= \frac{2N-1}{\Psi_N(0)}H\left(\frac{|m|}{N}\right)x(m)\tilde{\epsilon}_m(\ell), \\
Y_m(\ell) &:= \frac{2N-1}{\Psi_N(0)}H\left(\frac{|m|}{N}\right)\epsilon(m)\frac{x(m+\ell)+x(m-\ell)}{2}, \\
Z_m(\ell) &:= \frac{2N-1}{\Psi_N(0)}H\left(\frac{|m|}{N}\right)\{\epsilon(m)\tilde{\epsilon}_m(\ell) - \delta_{0,\ell}v(m)\}.
\end{aligned}$$

We observe that $H\left(\frac{|m|}{N}\right) = 0$ if $|m| \geq N$. Therefore,

$$\begin{aligned}
\tilde{y}_N(\ell) &= y_N(\ell) + \frac{1}{\Psi_N(0)}\sum_{m\in\mathbb{Z}} H\left(\frac{|m|}{N}\right)x(m)\tilde{\epsilon}_m(\ell) + \frac{1}{\Psi_N(0)}\sum_{m\in\mathbb{Z}} H\left(\frac{|m|}{N}\right)\epsilon(m)\frac{x(m+\ell)+x(m-\ell)}{2} \\
&\quad + \frac{1}{\Psi_N(0)}\sum_{m\in\mathbb{Z}} H\left(\frac{|m|}{N}\right)\{\epsilon(m)\tilde{\epsilon}_m(\ell) - \delta_{0,\ell}v(m)\} \\
&= y_N(\ell) + \frac{1}{2N-1}\sum_{m=-N+1}^{N-1} X_m(\ell) + \frac{1}{2N-1}\sum_{m=-N+1}^{N-1} Y_m(\ell) + \frac{1}{2N-1}\sum_{m=-N+1}^{N-1} Z_m(\ell).
\end{aligned} \tag{4.7}$$

Using the fact that $\mathbb{E}(\epsilon(k)) = 0$ for all k and the stochastic independence of $\epsilon(k)$'s, we deduce that for every $m \in \mathbb{Z}$ (and $\ell \in \mathbb{Z}$),

$$\mathbb{E}(X_m(\ell)) = \mathbb{E}(Y_m(\ell)) = \mathbb{E}(Z_m(\ell)) = 0.$$

Since $\Psi_n(0) \geq NH_1/2$, and $H(t) \leq 1$ on $\mathbb{R}$, the random variables $X_m(\ell)$, $Y_m(\ell)$, $Z_m(\ell)$ satisfy

$$|X_m(\ell)| \leq 4A\epsilon/H_1, \ |Y_m(\ell)| \leq 4A\epsilon/H_1, \ |Z_m(\ell)| \leq 8\epsilon^2/H_1.$$

We note that each of the sequences $\{X_m(\ell)\}$, $\{Y_m(\ell)\}$, $\{Z_m(\ell)\}$ is independent. Therefore, Proposition 2.2 is applicable. We use it with $C+2$ in place of C to obtain that each of the following estimates holds with probability exceeding $1-2N^{-C-2}$:

$$\begin{aligned}
\left|\frac{1}{2N-1}\sum_{m=-N+1}^{N-1} X_m(\ell)\right| &\leq 4\frac{A\epsilon}{H_1}\sqrt{\frac{G(C+2)\log N}{N}},\\
\left|\frac{1}{2N-1}\sum_{m=-N+1}^{N-1} Y_m(\ell)\right| &\leq 4\frac{A\epsilon}{H_1}\sqrt{\frac{G(C+2)\log N}{N}},\\
\left|\frac{1}{2N-1}\sum_{m=-N+1}^{N-1} Z_m(\ell)\right| &\leq 8\frac{\epsilon^2}{H_1}\sqrt{\frac{G(C+2)\log N}{N}}.
\end{aligned}$$

The estimates (4.6) follow for each ℓ immediately from these estimates and (4.7), with probability exceeding $1-6N^{-C-2}$. Since $N\geq 7$, the probability that at least one of these inequalities does not hold for at least one of the ℓ's in the range $|\ell|\leq N$, is at most $6(2N+1)N^{-C-2}\leq 2N^{-C}$. If N satisfies all the conditions (2.10), (3.9), and (4.2), then we may use (3.8) to deduce (4.3) from (4.6).

Next, if N satisfies (4.4) as well, then any $b_0,\cdots,b_{I^*-1}$ with $\sum_{j=0}^{I^*-1}|b_j|^2=1$, we have with probability exceeding $1-2N^{-C}$,

$$\begin{aligned}
&\left|\sum_{j,k=0}^{I^*-1} b_j\overline{b_k}\tilde{y}_N(j-k)-\sum_{j,k=0}^{I^*-1} b_j\overline{b_k}y_N(j-k)\right|\\
&\leq 8\frac{\epsilon(A+\epsilon)}{H_1}\sqrt{\frac{G(C+2)\log N}{N}}\left(\sum_{j=0}^{I^*-1}|b_j|\right)^2\leq 8I^*\frac{\epsilon(A+\epsilon)}{H_1}\sqrt{\frac{G(C+2)\log N}{N}}\leq Z/4.
\end{aligned}$$

In view of (3.10), this implies (4.5). The existence of the polynomial $\tilde{\Phi}_{I^*,N}$ follows from the general theory outlined in Section 2.1. □

We will now discuss the estimation of the points $\{z_j\}$ by the zeros $\{\tilde{z}_j\}$ of $\tilde{\Phi}_{I^*,N}$. Since both of these sets are sets of eigenvalues of suitable CMV matrices, we will use the following proposition, called Bauer–Fike theorem in [5, Theorem 7.2.2, p. 321], to help us in this estimation. The Euclidean norm of a complex vector $\mathbf{z}$ will be denoted by $\|\mathbf{z}\|$. For a matrix $\mathcal{X}$, $\|\mathcal{X}\|$ is the operator norm of $\mathcal{X}$ with Euclidean norms on both its domain and codomain spaces.

Proposition 4.1 *Let $n\geq 1$ be an integer, $\mathcal{A}$, $\mathcal{E}$ be any $n\times n$ matrices with complex entries, $\mathcal{X}$ be an invertible matrix such that*

$$\mathcal{X}^{-1}\mathcal{A}\mathcal{X}=\ \textit{diag}\ (w_1,\cdots,w_n), \tag{4.8}$$

and $\tilde{w}$ be any eigenvalue of $\mathcal{A}+\mathcal{E}$. Then

$$\min_{1\leq j\leq n}|w_j-\tilde{w}|\leq\|\mathcal{X}^{-1}\|\|\mathcal{X}\|\|\mathcal{E}\|. \tag{4.9}$$

In the sequel, $c,c_1,\cdots,$ will denote positive constants depending on H, S, and μ only. Their value may be different at different occurrences, even within a single formula. The notation $B\sim B_1$ means $cB_1\leq B\leq c_1B_1$.

Theorem 4.2 *Let $C>0$. There exists a constant $c^*>0$, depending on μ, H, C, and S alone satisfying the following property with probability exceeding $1-2N^{-C}$. Let $N\geq\max(3,I^*)$ satisfy (2.10), (3.9), (4.4), (4.2), and*

$$c^*\epsilon(A+\epsilon)\sqrt{\frac{\log N}{N}}<q/\pi. \tag{4.10}$$

For each $\ell\in\mathcal{F}$, let

$$B_\ell=\left\{z\in\mathbb{C}:|z-\tilde{z}_\ell|\leq c^*\epsilon(A+\epsilon)\sqrt{\frac{\log N}{N}}\right\}$$

Then each disc B_ℓ contains exactly one of the points $\{z_j\}_{j\in\mathcal{F}}$.

Proof. In this proof, we will write $\mathcal{C}$ (respectively, $\tilde{\mathcal{C}}$) in place of the CMV matrix $\mathcal{C}^{(I^*)}(\mu)$ (respectively, $\mathcal{C}^{(I^*)}(\tilde{\mathbf{y}}_N)$. Similarly, D_k, α_k (respectively, $\tilde{D}_k$, $\tilde{\alpha}_k$) will denote the determinants and Verblunsky coefficients for μ (respectively, the sequence $\tilde{\mathbf{y}}_N$). Then the points z_j (respectively, $\tilde{z}_j$), $j \in \mathcal{F}$ are the eigenvalues of $\mathcal{C}$ (respectively, $\tilde{\mathcal{C}}$). In particular, the eigenvalues of $\mathcal{C}$ are all distinct. Therefore, there exists an invertible matrix $\mathcal{X}$ such that (4.8) is satisfied with $n = I^*$ and $w_k = z_k$ ([21, Theorem 10-8, p. 489]). Since the matrix $\mathcal{C}$ depends only on the measure μ, so does the matrix $\mathcal{X}$. In particular, the quantity $\|\mathcal{X}^{-1}\|\|\mathcal{X}\|$ is a positive constant depending only on μ.

All the following statements in this discussion will hold with probability exceeding $1 - 2N^{-C}$.

We recall from (4.3) that

$$|\tilde{y}_N(\ell) - y(\ell)| \le 16\frac{\epsilon(A+\epsilon)}{H_1}\sqrt{\frac{G(C+2)\log N}{N}} \le c\epsilon(A+\epsilon)\sqrt{\frac{\log N}{N}}, \qquad \ell = -N, \cdots, N. \tag{4.11}$$

Clearly, the same upper bound holds (with different constants) for $|D_n - \tilde{D}_n|$ for $n = 0, \cdots, I^* - 1$, and hence, with yet different constants, also for $|\alpha_n - \tilde{\alpha}_n|$ and $|\rho_n - \tilde{\rho}_n|$ for $n = 0, \cdots, I^* - 1$. Consequently, for each $j, k = 0, \cdots, I^* - 1$,

$$|\mathcal{C}_{j,k} - \tilde{\mathcal{C}}_{j,k}| \le c\epsilon(A+\epsilon)\sqrt{\frac{\log N}{N}}.$$

Since $\|\mathcal{C} - \tilde{\mathcal{C}}\|$ is at most the Frobenius norm of $\mathcal{C} - \tilde{\mathcal{C}}$, we deduce easily that

$$\|\mathcal{C} - \tilde{\mathcal{C}}\| \le c\epsilon(A+\epsilon)\sqrt{\frac{\log N}{N}}. \tag{4.12}$$

In view of Proposition 4.1, we obtain that there exists a constant $c^* > 0$ depending only on H, S, and μ such that for each $\ell \in \mathcal{F}$, the disc B_ℓ contains at least one of the points $\{z_k\}_{k\in\mathcal{F}}$. Now, if $j \ne k$ then

$$|z_k - z_j|^2 = 2 - 2\cos(\omega_j - \omega_k) = 4\sin^2(|\omega_j - \omega_k|/2) \ge (4/\pi^2)|\omega_j - \omega_k|^2 \ge (2q/\pi)^2.$$

So, if N satisfies (4.10), then each of the discs B_ℓ contains exactly one of the points $\{z_k\}_{k\in\mathcal{F}}$. □

Finally, we discuss some ideas on estimating the number I^*. This can be done using the methods discussed in [15]. The following discussion is more in the spirit of the present paper. The bounds can be improved using Proposition 2.2 if some estimates are known on the spectral norm of the matrix $[\epsilon(j-k)]_{j,k=-N}^N$, but we will keep the discussion simple by using only the assumptions we have already made. All the following statements in this discussion will hold with probability exceeding $1 - 2N^{-C}$. Let $Y = [y(j-k)]_{j,k=-N}^N$, $\tilde{Y}_N = [\tilde{y}_N(j-k)]_{j,k=-N}^N$, and $\{\xi_k\}$ (respectively, $\{\tilde{\xi}_k\}$) be the eigenvalues of Y (respectively, $\tilde{Y}_N$), arranged in the decreasing order.

The estimate (4.11) implies a crude estimate

$$\|Y - \tilde{Y}_N\| \le c\epsilon(A+\epsilon)\sqrt{N\log N}. \tag{4.13}$$

According to a corollary to the Wieldant–Hoffman theorem [5, Corollary 8.1.6], this shows that

$$\max_{1\le j\le N} |\xi_k - \tilde{\xi}_k| \le c\epsilon(A+\epsilon)\sqrt{N\log N}. \tag{4.14}$$

However, unless ϵ is very small, this is not a useful bound. To reduce ϵ, one may wish to sample the random variables $\tilde{x}(k)$ K times. The average of these samples, $\overline{\tilde{x}_k}$ satisfies

$$\overline{\tilde{x}_k} = x(k) + \overline{\epsilon(k)},$$

where $\overline{\epsilon(k)}$ is the average of the $\epsilon(k)$'s. Since all the quantities here are real, the overline denotes the average in this context, rather than the complex conjugate. In view of Proposition 2.2,

$$\mathsf{Prob}\left(|\overline{\epsilon(k)}| \ge \epsilon\sqrt{\frac{G(C+2)\log K}{K}}\right) \le 2K^{-C}, \qquad k = -N, \cdots, N.$$

Thus, with probability exceeding $1-2K^{-C}$, we may replace ϵ by $\epsilon\sqrt{\dfrac{G(C+2)\log K}{K}}$. Therefore, with probability exceeding $(1-2K^{-C})(1-2N^{-C})$, we obtain in place of (4.14),

$$\max_{1\le j\le N}|\xi_k-\tilde{\xi}_k|\le c\epsilon\sqrt{N\log N}\left(A+\epsilon\sqrt{\frac{G(C+2)\log K}{K}}\right)\sqrt{\frac{G(C+2)\log K}{K}}. \tag{4.15}$$

We need to choose K large enough so that this bound does not exceed $Z/2$. Since $\xi_{I^*}\ge Z$, we may then find I^* by looking at the number of eigenvalues of $\tilde{Y}_N$ which exceed $Z/2$.

5 Numerical examples

In this section, we present a few numerical examples to illustrate the theory. It is not our intention to present efficient algorithms to compute orthogonal polynomials; merely to observe the effect of considering the moments $\tilde{y}_N(k)$ rather than $\tilde{x}(k)$. In each example, we will report the average error in a number of trials. There are two approaches here. One approach is to take the average of each $\tilde{x}(k)$ before starting, as we indicated during our discussion of estimating I^*. In our examples below, we take the other approach. In each trial, we will use just one sample $\tilde{y}(k)$, $|k|\le N$. We then compute the frequencies with this one observation, and calculate the error from the actual frequencies. The average of these errors is reported. In the first two examples, we assume I^* to be known.

For the function H, we choose the following:

$$H(t):=\begin{cases}1, & \text{if } 0\le t\le 1/2,\\ \exp\left(-\dfrac{\exp(2/(1-2t))}{1-t}\right), & \text{if } 1/2<t<1,\\ 0, & \text{if } t\ge 1.\end{cases}$$

For computing orthogonal polynomials, we simply use Prony's method. We consider the eigenvector $(v_0,\cdots,v_{I^*})^T$ of the matrix

$$M=[\tilde{y}(I^*-j-\ell)]_{j,\ell=0}^{I^*}$$

corresponding to the eigenvalue 0, as in (1.4) (where one more equation is added to get a square matrix with eigenvalue 0). The desired orthogonal polynomial is given by $\sum_{k=0}^{I^*}v_kz^{I^*-k}$. If the zeros of this polynomial are $(r_1\exp(i\theta_1),\cdots,r_{I^*}\exp(i\theta_{I^*}))$, then the computed frequencies $\tilde{\omega}_j$'s are taken to be $\{\theta_1,\cdots,\theta_{I^*}\}$ in an ascending order. Because of certain technicalities concerning the implementation of the commands `svd` and `eig` of Matlab, we actually computed the vector $(v_0,\cdots,v_{I^*})$ as the (right) singular vector corresponding the to smallest singular value of M. We found the zeros of this polynomial using the `roots` command.

EXAMPLE 1. We consider the example given in [8]:

$$x(k)=\sin(k\pi/2)+\sin(k\pi/3)+\sin(k\pi/6)+10\sin(3k\pi/4)+\epsilon(k),\qquad k=-199,\cdots,199, \tag{5.1}$$

where $\epsilon(k)$ is a normal random variable with mean zero and standard deviation $\sqrt{0.02}$ (variance 0.02). An average of 100 trials gave the following frequencies:

$$(\pm 0.52348954499631,\pm 1.04707921259201,\pm 1.57071968582955,\pm 2.35619363762873)$$

as an approximation to the real frequencies:

$$(\pm 0.52359877559830,\pm 1.04719755119660,\pm 1.57079632679490,\pm 2.35619449019234)$$

In addition, we computed the maximum absolute error in each trial. The average of these was $9.201629e-4$. □

EXAMPLE 2. This example is motivated by a similar example in [20]. We considered $I=100$, and the frequencies

$$\omega_j=(j\pi/101)(1+n_j),\qquad j=1,\cdots,I,$$

where each n_j is a uniformly distributed random variable in the range $[-0.5, 0.5]$. With different values of N as indicated in Table 1, we then considered the moments

$$\tilde{x}(k) = \sum_{j=1}^{I} \cos(k\omega_j) + \epsilon(k), \qquad |k| \le N + 200,$$

where $\epsilon(k)$ is a uniformly distributed random variable in the range $[-0.002, 0.002]$. We used the same procedure as in Example 1 to estimate the frequencies, $\{\tilde{\omega}_j\}$. In each trial, we computed the maximum of the errors $|\tilde{\omega}_j - \omega_j|/q$, where q is the minimal separation among the frequencies in that trial. In the table below, err_N denotes the average over 100 trials of these maximum errors. The maximum value of the minimal separation as obtained among these trials is indicated on line 3 of this table. It is noted that this minimal separation is far less than $\pi/100$. □

N	256	512	1024
err_N	0.43033271423691	0.52085197284142	0.35016180491495
Maximum q	0.00887441453240	0.00987858969739	0.00922718056415

Table 1: Average over 100 trials for $\mathrm{err}_N := \max_j |\tilde{\omega}_j - \omega_j|/q$ for the indicated values of N, where q is the minimal separation among the frequencies in each trial. The number of moments used is $4N + 1$. The maximum value for the minimal separation over all the trials is also indicated.

Example 3. We consider

$$\tilde{x}(k) = 34 + 300\cos(k\pi/4) + \cos(k\pi/2) + \epsilon(k), \qquad k = -1024, \cdots, 1024, \tag{5.2}$$

where $\epsilon(k)$ is a random variable uniformly distributed in the range $[-3, 3]$. Thus, in addition to the large differences in the magnitudes at different frequencies, also the noise is three times the strength of the weakest signal at $\pi/2$. The method which we used in the other two examples did not give satisfactory results. At our request, Daniel Potts ran the algorithm ESPIRIT as in [20], but starting with the moments $\tilde{y}(k)$ as defined in (4.1). As an average over 500 trials, he recovered the frequencies as

$$(-3.992361996552509e - 16, \pm 0.785398165178676, \pm 1.570153903610014).$$

In particular, the weakest frequency $\pi/2$ was detected with an accuracy of $6.4242e - 4$ in spite of the noise being 3 times the strength of the corresponding signal. We note that the error reported in [20] using this algorithm and the correct bound on the number of frequencies is $2.3783e - 2$. The noise in that paper is in the range $[0, 3]$. Thus, using the moments $\tilde{y}_N(k)$ in place of the raw moments $\tilde{x}_N(k)$ gives a remarkable improvement, in spite of the noise having 4 times as much variance. □

References

[1] G. Beylkin and L. Monzón, *On approximations of functions by exponential sums,* Appl. Comput. Harmon. Anal. **19** (2005) 17–48.

[2] P. L. Butzer and R. J. Nessel, "Fourier analysis and approximation, Vol. 1", Academic Press, New York, 1971.

[3] K. S. Eckhoff, *Accurate reconstructions of functions of finite regularity from truncated Fourier series expansions,* Math. Comput., **64** (1995), 671–690.

[4] A. Gelb and E. Tadmor, *Detection of edges in spectral data,* Appl. Comput. Harmon. Anal. **7** (1999), no. 1, 101–135.

[5] G. H. Golub and C. F. van Loan, "Matrix computations", The Johns Hopkins University Press, Baltimore, 1996.

[6] E. Hewitt and K. Stromberg, "Real and abstract analysis", Springer, New York, 1975.

[7] F. J. Hickernell, I. H. Sloan, and G. W. Wasilkowski, *On tractability of weighted integration over bounded and unbounded regions in R^s,* Math. Comp. **73** (2004), no. 248, 1903–1911.

[8] W. B. Jones, O. Njåstad, and E. B. Saff, *Szegö polynomials associated with Wiener–Levinson filters,* J. Comput. Appl. Math., **32** (1990) 387-406.

[9] Y. Katznelson, "An Introduction to Harmonic Analysis", John Wiley & Sons, Inc., New York, 1968.

[10] H. Krim and M. Vidberg, *Sensor array signal processing: two decades later,* Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, LIDS P-2282, 1995.

[11] C. Lanczos, "Applied analysis", Dover, New York, 1988.

[12] J. D. Markel and A. N. Gray Jr., "Linear prediction of speech", Springer Verlag, New York, 1976.

[13] H. N. Mhaskar, *On the tractability of multivariate integration and approximation by neural networks,* J. Complexity **20** (2004), no. 4, 561–590.

[14] H. N. Mhaskar and J. Prestin, *On the detection of singularities of a periodic function,* Adv. Comput. Math., **12** (2000), 95–131.

[15] H. N. Mhaskar and J. Prestin, *On local smoothness classes of periodic functions,* J. Fourier Anal. Appl. **11** (2005), no. 3, 353–373.

[16] K. C. Pan, *A refined Wiener-Levinson method in frequency analysis,* SIAM J. Math. Anal. **27** (1996), no. 5, 1448–1453.

[17] K. C. Pan and E. B. Saff, *Asymptotics for zeros of Szegö polynomials associated with trigonometric polynomial signals,* J. Approx. Theory **71** (1992), no. 3, 239–251.

[18] S. U. Pillai, "Array signal processing", Springer Verlag, New York, 1989.

[19] D. Pollard, *Convergence of stochastic processes,* Springer Verlag, New York, 1984.

[20] D. Potts and M. Tasche, *Parameter estimation for exponential sums by approximate Prony method,* Signal Processing, to appear.

[21] A. Ralston and P. Rabinowitz, "A first course in numerical analysis", McGraw Hill, New York, 1978.

[22] R. Roy and T. Kalaith, *ESPRIT – estimation of signal parameters via rotational invariance techniques,* IEEE Trans. Acoust. Speech Signal Process. **37** (1989), no. 7, 984–995.

[23] B. Simon, " Orthogonal polynomials on the unit circle, Part 1, Classical theory", American Mathematical Society Colloquium Publications, 54, Part 1. American Mathematical Society, Providence, RI, 2005.

[24] G. Szegö, "Orthogonal Polynomials", Amer. Math. Soc. Colloq. Publ. **23**, Amer. Math. Soc., Providence, 1975.