Abstract—We derive an explicit expression for the covariance of the log-periodogram power spectral density estimator for a zero mean Gaussian process. We do not make the assumption that the spectral components of the process are uncorrelated. Applications to spectral estimation and to cepstral modeling in automatic speech recognition are discussed.

Index Terms—Log-periodogram, speech recognition.

I. INTRODUCTION

We derive an explicit expression for the covariance of the log-periodogram power spectral density estimator. The estimator is obtained from a time series $y_0, y_1, \ldots, y_{k-1}$, which is assumed to have been drawn from a zero mean Gaussian process with arbitrary autocorrelation function. Let

$$Y_\theta = \sum_{n=0}^{k-1} y_n e^{-i\theta n}$$

(1)

denote the discrete Fourier transform (DFT) component of the time series at frequency $\theta$ rad, where $0 \leq \theta \leq \pi$. We also refer to $Y_\theta$ as a spectral component of the time series. The periodogram of the time series is given by $(1/k)|Y_\theta|^2$. We are interested in an explicit expression for the covariance

$$
\text{cov} \left( \log \frac{1}{k} |Y_{\theta_1}|^2, \log \frac{1}{k} |Y_{\theta_2}|^2 \right)
$$

(2)

when $Y_{\theta_1}$ and $Y_{\theta_2}$, $\theta_1 \neq \theta_2$, are not assumed uncorrelated. The natural logarithm is used throughout the letter. The covariance of any two spectral components of a stationary process with autocorrelation function $r_y(m)$ is given by

$$
\text{cov}(Y_{\theta_1}, Y_{\theta_2}) = e^{-i((\theta_1-\theta_2)/2)k} \times \sum_{m=-k}^{k} r_y(m) \frac{\sin \frac{\theta_1 - \theta_2}{2} (k - |m|)}{\sin \frac{\theta_1 - \theta_2}{2}} \cos \frac{\theta_1 + \theta_2}{2} m.
$$

(3)

For finite $k$, the spectral components $\{Y_\theta\}$ are generally correlated. Two noted exceptions are obtained when the process is either white or almost surely periodic with period $k$, and the DFT is sampled at frequencies that are multiples of $2\pi/k$.

The log-periodogram covariance expression we derive here refines a standard form of the covariance obtained under the assumption that the spectral components $\{Y_\theta\}$ are statistically independent Gaussian random variables (see, e.g., [4, Eq. 11]). This assumption follows from asymptotic properties of spectral components of a stationary process with fast decaying autocorrelation function (see, e.g., Brillinger [2, Th. 4.4.1]). In some applications, however, this assumption may not hold either because of lack of stationarity or due to a relatively small $k$. This situation may be encountered in automatic speech recognition applications, where the signal is not strictly stationary, and hence, the frame length $k$ is limited. In this letter, we assume that the process is Gaussian, and hence, its spectral components are jointly Gaussian. We relax the common assumption that the jointly Gaussian spectral components are uncorrelated. In [3, Eq. 2.6], the spectral components were assumed statistically independent, but their densities differed from normal.

The covariance expression derived here generalizes another result from [4]. In [4, Eq. 13], an explicit expression was developed for the covariance between the log-periodogram of a clean signal and the log-periodogram of a corresponding noisy signal at the same frequency. The noise was assumed additive and statistically independent of the signal. Both the signal and noise were assumed zero-mean Gaussian. In that case, the clean and noisy spectral components at any given frequency are correlated, and the log-periodogram covariance can be obtained from the expression we develop for (2).

A key function in our development is the hypergeometric function [6, Ch. 9]. In [4], analytic continuation of this function was used in deriving the covariance between the clean and noisy log-periodogram estimates. Here, we provide a new proof for the more general result, which does not require analytic continuation of the hypergeometric function when $0 < \theta_1, \theta_2 < \pi$.

Explicit expressions for the second-order statistics of the log-periodogram of a given signal are important in the theory of spectral estimation [9], [10]. In addition, they provide some insight into the theory of cepstral coefficients that are extensively used in statistical modeling of speech signals for automatic speech recognition applications (see, e.g., [4], [8] and the references therein). Cepstral coefficients are obtained from the inverse Fourier transform of a log-power spectral density estimate of the signal. Merhav and Lee [8] have shown that cepstral coefficients of a zero-mean stationary Gaussian process, obtained from the windowed autocorrelation power spectral density estimator [9, Sec. 6.2.3], are asymptotically uncorrelated when $k \to \infty$. For finite $k$, the covariance of cepstral components obtained from the log-periodogram power spectral density estimator may be estimated from the log-periodogram covariance expression developed here.
II. MAIN RESULTS

Consider a zero-mean Gaussian process with spectral components \{Y_\theta\} as defined in (1). We are primarily interested in the log-periodogram covariance (2) for \(0 < \theta < \pi\). The log-periodogram covariance associated with the frequencies \(\theta \in \{0, \pi\}\) is given in Section III. The log-periodogram covariance is expressed in terms of the correlation coefficient between \(Y_{\theta_1}\) and \(Y_{\theta_2}\) defined as

\[
\rho^2_{\theta_1 \theta_2} = \frac{\left\{ \frac{1}{k} E\{Y_{\theta_1}Y_{\theta_2}^*\}\right\}^2}{\frac{1}{k} E\{Y_{\theta_1}^2\} \frac{1}{k} E\{Y_{\theta_2}^2\}^2}
\]

where \(Y_{\theta_2}^*\) denotes the complex conjugate of \(Y_{\theta_2}\).

For \(0 < \theta_1, \theta_2 < \pi\), it is well known that the expected value of the log-periodogram is given by (see, e.g., [3, Eq. 2.5], [4, Eq. 10])

\[
E\left\{ \log \frac{1}{k} |Y_{\theta_1}|^2 \right\} = \log E\left\{ \frac{1}{k} |Y_{\theta_1}|^2 \right\} - \gamma
\]

where \(\gamma \approx 0.57721\) is the Euler constant. Furthermore, \(|\rho_{\theta_1 \theta_2}| \leq 1\). Our main result is the explicit expression for the covariance of the log-periodogram given by

\[
\text{cov}\left( \log \frac{1}{k} |Y_{\theta_1}|^2, \log \frac{1}{k} |Y_{\theta_2}|^2 \right) = \sum_{n=-\infty}^{\infty} \frac{1}{n^2} \rho^2_{\theta_1 \theta_2} \delta_n.
\]

When the spectral components \(Y_{\theta_1}\) and \(Y_{\theta_2}\) are assumed uncorrelated, \(\rho_{\theta_1 \theta_2} = 1\) for \(1 = \theta_2\) and \(\rho_{\theta_1 \theta_2} = 0\) otherwise. In that case, (6) is reduced to a well-known form in which it is equal to \(\pi^2/6\) when \(\theta_1 = \theta_2\) and to zero otherwise (see, e.g., [3, Eq. 2.6.1], [4, Eq. 11.1]).

The covariance expression in (6) may be applicable to spectral components that were not necessarily derived from the same process. For example, let \(Y_{\theta}\) and \(W_{\theta}\) denote, respectively, zero-mean Gaussian spectral components from a signal and an additive statistically independent noise process. The squared correlation coefficient between \(Y_{\theta}\) and \(Z_{\theta} = Y_{\theta} + W_{\theta}\) is given by

\[
G_{\theta} = \left( \frac{(1/k)E\{\|Y_{\theta}\|^2\}}{((1/k)E\{\|Z_{\theta}\|^2\})} \right),
\]

which is recognized as the Wiener gain factor for estimating \(Y_{\theta}\) from \(Z_{\theta}\). For this case, we show that the log-periodogram covariance is obtained from (6) by substituting \(\rho^2_{\theta_1 \theta_2}\) with \(G_{\theta}\). This result was originally derived in [4, Eq. (13)] and used for designing a linear estimator for the log-periodogram of the clean signal from the log-periodogram of the noisy signal. The derivation in [4] relied on analytic continuation of the hypergeometric function. The proof of the more general result (6) we provide here does not require such analytic continuation.

III. DERIVATION OF THE LOG-PERIODOGRAM COVARIANCE

Let \(X_1 = (1/k^{1/2})Y_{\theta_1}\) and \(X_2 = (1/k^{1/2})Y_{\theta_2}\) denote two normalized spectral components, which, under our assumptions, are zero-mean jointly Gaussian random variables. Let

\[
R_{12} = \begin{pmatrix} \lambda_1 & \tau_{12} \\ \tau_{12}^* & \lambda_2 \end{pmatrix}
\]

denote the covariance matrix of the vector \((X_1, X_2)'\), where \((\cdot)'\) denotes vector transpose.

Consider first the case where \(0 < \theta_1, \theta_2 < \pi\), and \(\{X_1, X_2\}\) are complex random variables. The \(4 \times 4\) covariance matrix of the real and imaginary components of \(X_1\) and \(X_2\) has the usual structure for spectral components as detailed in [11, Eq. 11–20]. This implies, for example, that the real and imaginary components of \(X_1\) are noncorrelated, and each has the same variance. For this notation, \(\rho_{12}^2 = \langle |X_1|^2/\langle |X_2|^2 \rangle \rangle\). The desired covariance expression (6) is obtained from the second-order derivative of the moment generating function

\[
\Phi_{12}(\mu, \eta) = E\left\{ e^{\mu Y_{12}} e^{\eta Y_{12}^*} \right\} = E\left\{ |X_1|^2 |X_2|^2 \right\}
\]

with respect to \(\mu\) and \(\eta\) at \(\mu = \eta = 0\) and from (5).

The moment generating function is evaluated as

\[
\Phi_{12}(\mu, \eta) = E\left\{ |X_2|^2 e^{\mu Y_{12}} e^{\eta Y_{12}^*} \right\} = E\left\{ |X_1|^2 |X_2|^2 \right\}
\]

where \(\sigma_2^2 = \lambda_1 (1 - \rho_{12}^2)\) denotes the conditional variance of \(X_1\) given \(X_2\). Using (10), we evaluate the inner expected value in (9) in polar coordinates. Define \(\gamma_{12} = (\langle |X_1|^2 \rangle / (\langle |X_2|^2 \rangle))\). Using [5, Eqs. 8.411.1, 6.631.1], we obtain

\[
E\left\{ |X_1|^2 |X_2|^2 \right\} = \sigma_2^2 \Gamma(\mu + 1) e^{-\gamma_{12} |X_2|^2}
\]

\[
\times (1 - \rho_{12}^2)^{\mu+\eta+1} 2 \Gamma(\mu + 1; \eta + 1) F_1(\mu + 1, \eta + 1; \rho_{12}^2)
\]

where

\[
F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{1}{n!} z^n
\]

denotes the degenerate hypergeometric function with convergence region of \(|z| < 1\). Note that the argument of the hypergeometric function in (13) is in its convergence region, and application of the expected value term by term is justified. Applying the linear transformation [6, Eq. 9.5.3]

\[
2 \Gamma(\mu + 1) \Gamma(\eta + 1) 2 F_1(c - a - b; c; z)
\]

(15) to (13), we obtain

\[
\Phi_{12}(\mu, \eta) = \lambda_1^2 \Gamma(\mu + 1) \lambda_2^2 \Gamma(\eta + 1) 2 F_1(-\mu - \eta - 1; \rho_{12}^2).\]

The second-order derivative of (16) at \(\mu = \eta = 0\) is obtained using \(\partial \Gamma(\mu+1)/\partial \mu = \gamma\) at \(\mu = 0\) and \(\partial (\cdot)_n/\partial \mu = -(n-1)!\) at \(\mu = 0\) for \(n > 0\) [4, Eq. A.23].
For the log-periodogram covariance between a clean and a noisy spectral component at a given frequency \( \theta \), it is easy to check that the required structure of the covariance matrix of the real and imaginary components of the clean and noisy spectral components is satisfied. Thus, the log-periodogram covariance can be obtained from (6) using \( \rho_{\theta}^2 = \bar{G}_{\theta} \). In the original proof of this result in [4], the moment generating function was evaluated as in (9) using \( X_1 = (1/k^{1/2})Z_{\theta} \) and \( X_2 = (1/k^{1/2})Y_{\theta} \). As a result, the argument of the hypergeometric function was the negative of the signal-to-noise ratio at the given frequency, which can exceed one in absolute value. In such cases, the series representing the function does not converge, and analytic continuation of the hypergeometric function is necessary [6]. This difficulty can be circumvented if the moment generating function is evaluated as above using \( X_1 = (1/k^{1/2})Y_{\theta} \) and \( X_2 = (1/k^{1/2})Z_{\theta} \) in (9).

The covariance of the log-periodogram when \( \theta_1, \theta_2 \in (0, \pi) \), and when \( \theta_1 \in (0, \pi) \) and \( 0 < \theta_2 < \pi \), may be obtained using a similar approach to that used above. In these two cases, however, analytic continuation of the hypergeometric function is necessary for valid covariance expressions over the full range of \( \theta_2 \). For \( \theta_1, \theta_2 \in (0, \pi) \), \( \Phi_{12}(\mu, \eta) \) was evaluated using [5, Eqs. 3.562.2, 9.240, 8.331, 8.335.1, 9.212.1, 3.381.4] in that order. The resulting hypergeometric function converges for \( \rho_{\theta_2}^2 < 0.5 \) only. Applying the linear transformation [6, Eq. 9.5.1]

\[
2F_1(a, b; c; z) = (1 - z)^{-a}2F_1 \left( a, c - b, c; \frac{z}{z-1} \right)
\]

(17)
gives

\[
\Phi_{12}(\mu, \eta) = \frac{1}{\pi} (2\lambda_2)^\mu \Gamma(\mu + 0.5) \\
\times (2\lambda_2)^\eta \Gamma(\eta + 0.5)2F_1 \left( -\mu - \eta; 0.5; \rho_{\theta_2}^2 \right)
\]

(18)
where convergence is for \( |\rho_{\theta_2}^2| < 1 \). For \( \theta_1 \in (0, \pi) \) and \( 0 < \theta_2 < \pi \), we have applied the same sequence of steps as above and also used [5, 1.320.5]. Here \( \rho_{\theta_2}^2 < 0.5 \), but the resulting hypergeometric function converges only for \( \rho_{\theta_2}^2 < 0.25 \). Applying the linear transformation (17) gives

\[
\Phi_{12}(\mu, \eta) = \frac{1}{\sqrt{\pi}} (2\lambda_2)^{\mu + \eta} \Gamma(\mu + \eta + 1) \\
\times 2F_1 \left( -\mu, -\eta; 1; \rho_{\theta_2}^2 \right)
\]

(19)
with the desired convergence region. The covariance of the log-periodogram corresponding to these two cases can be obtained from the second-order derivatives at \( \mu = \eta = 0 \) of (18) and (19), respectively, and by using (5) and [4, Eqs. 10, A.10]. For \( \theta_1, \theta_2 \in (0, \pi) \), we have

\[
\text{cov} \left( \log \frac{1}{k} \hat{Y}_{\theta_1}, \log \frac{1}{k} \hat{Y}_{\theta_2} \right) = \sum_{n=1}^{\infty} \frac{n!}{(0.5^n)_n} \frac{1}{n^2} \rho_{\theta_1 \theta_2}^2
\]

(20)
and for \( \theta_1 \in (0, \pi) \) and \( 0 < \theta_2 < \pi \), we have

\[
\text{cov} \left( \log \frac{1}{k} \hat{Y}_{\theta_1}^2, \log \frac{1}{k} \hat{Y}_{\theta_2}^2 \right) = \sum_{n=1}^{\infty} \frac{1}{n^2} \left( 2\rho_{\theta_1 \theta_2}^2 \right)^n.
\]

(21)

Note that the transformations (15) and (17) are related. In fact, the former can be obtained from the latter [6]. The transformation (15), however, does not affect the convergence region of the hypergeometric function and, thus, does not provide analytic continuation of that function.

### IV. CEPSTRAL COEFFICIENTS

Consider cepstral coefficients \( \{c_n\} \) obtained from inverse DFT of a windowed log-periodogram as follows:

\[
c_n = \frac{1}{k} \sum_{k=0}^{k-1} w_k \log \left( \frac{1}{k} \hat{Y}_{\theta_k}^2 \right) e^{i\theta_k n}
\]

(22)
where \( n = 0, 1, \ldots, k - 1, \theta_k = (2\pi/k)l, \) and \( \{w_k\} \) denote the window function. The window may be used to exclude undesired components of the periodogram, such as at \( l = 0, k/2 \), or to achieve consistency of the cepstral estimates.

For a window that nulls out the components at \( l = 0, k/2 \), it is straightforward to verify, using (6), that the covariance of the cepstral coefficients is given by

\[
\text{cov}(c_n, c_m) = \frac{4}{k^2} \sum_{l=1}^{k/2-1} \sum_{l'=1}^{k/2-1} w_l w_{l'} \rho_{\theta_l \theta_{l'}}^{2l} \cos(\theta_l n) \cos(\theta_{l'} m).
\]

(23)
The variance of \( c_n, \text{var}(c_n) \), is easily obtained from (23) using \( n = m \). For uncorrelated spectral components, \( \text{var}(c_n) \) approaches zero when \( k \rightarrow \infty \). For correlated spectral components, consistency may be achieved by a proper choice of the window \( w_l \) in (22) similarly to [9, Sec. 6.2.4].

### V. APPLICATIONS

In this section, we briefly mention two applications of our results. In the first application, we provide a theoretical estimate of the covariance of the log of the squared magnitude of the spectral error in autoregressive parameter estimation. In the second application, we have embedded the covariance matrix (23) in a speech recognition system. We have used that matrix to predict the covariance of each state of the hidden Markov process (HMP) that models the acoustic signal from a given word.

For the first application, let \( \{u_t\} \) denote a zero-mean Gaussian autoregressive process with gain \( \sigma \) and coefficients \( \{a_1, \ldots, a_m\}' \). Let \( P \) denote the true probability measure of the process. Let \( a = (a_1, \ldots, a_m)' \). Suppose that \( \sigma \) and \( a \) are estimated in the maximum likelihood sense from \( n \) samples of \( \{u_t\} \). In particular, the estimates are obtained from the sample covariance estimate of the \( m + 1 \times m + 1 \) covariance \( R_{m+1} \) of the process \( \{u_t\} \). Let \( \hat{a}(n) \) denote the estimate of \( a \). It was shown by Mann and Wald [7] that

\[
\sqrt{n}(\hat{a}(n) - a) \rightarrow N_0 \left( 0, \left( \frac{R_{m+1}}{\sigma^2} \right)^{-1} \right) P\text{-weakly as } n \rightarrow \infty.
\]

(24)

This central limit theorem states that the asymptotic distribution of the error in estimating the autoregressive coefficients is zero-mean Gaussian with asymptotic covariance \( (R_{m+1}/\sigma^2)^{-1} \).

If we adopt these asymptotic results as valid for sufficiently large \( n \), then (6) can be used to assess the covariance of \( \log \left( \frac{1}{k} \hat{A}_0 - \hat{A}_0 \right)^2 \), where \( \{\hat{A}_0\} \) and \( \{A_0\} \) denote, respectively, the \( k \)-length DFTs of \( a \) and \( a \) augmented with \( a_0 = 1 \) and where \( k = m + 1 \). In particular, the variance of the
logarithm of the squared magnitude of the spectral error for any $0 < \theta < \pi$ is constant and is given by $\pi^2/6$.

For the second application of automatic speech recognition, we have first studied the extent of the correlation among spectral components of speech signals. We have used the training section of male speakers from the TI-digit database. This database contains two utterances per digit from each member of its 55 speakers. For each of the ten English digits and the word “oh,” we have estimated $\rho_{\theta_1\theta_2}$ from frames of length $k = 200$ samples at 8 kHz sampling rate from all utterances of all speakers. Fig. 1 depicts $\rho_{\theta_1\theta_2}$ for $0 < \theta_1, \theta_2 < \pi$ as obtained from sample covariance estimates for the digit “two.” In this figure, we have suppressed the diagonal terms $\rho_{\theta\theta}$, since all have a unity value. Other digits exhibited varying correlation ranging from zero to approximately 0.8, with the heaviest spectral correlation seen for the digit “six.”

The estimate of $\rho_{\theta_1\theta_2}$ for each digit was subsequently used in (23) to predict the covariance matrix of cepstral vectors from that digit. We have used a rectangular window that suppressed the log-periodogram components at $\theta = 0, \pi$. The predicted covariance was then used in an HMP-based speech recognition system with Gaussian densities. The predicted covariance was attributed to all states of the HMP. The Baum algorithm was used to estimate the mean vector of each Gaussian density only, as well as the transition matrix of the Markov chain. We have maintained only the diagonal terms of the predicted covariance, since the off-diagonal terms of this matrix were negligibly small.

The speech recognition setup we used here is similar to that described in [4]. In particular, we have recognized the ten English digits and the letter “oh” from the male testing section of the TI-digit database. For each digit, two utterances were available from each of 56 speakers. In all cases, we have used ten states, two mixture components per state, and frame length $k = 200$ samples at 8 kHz sampling rate. We have compared our system, which uses the predicted diagonal cepstral covariance matrices, with the standard approach in which diagonal covariance matrices are estimated from the training data for all states of the HMP using the Baum algorithm. We have also compared our system with the system in [4], where a fixed, data-independent, diagonal covariance matrix was attributed to all states of the HMP. In [4], we could use a fixed data-independent matrix since the spectral components were assumed statistically independent. We have practically achieved the same performance in all three systems. The benchmark recognition accuracy for the system using the Baum algorithm for estimating the covariance matrices was 99.19%. The system using the predicted covariance matrices developed here achieved 99.03%. The system in [4], which uses a fixed data-independent covariance matrix for all states, achieved 98.95%. The small differences in performance obtained in these three systems were verified to be statistically insignificant using the bootstrap sampling probability of improvement estimate proposed by Bisani and Ney [1]. These preliminary results are encouraging, since they show that the individually estimated covariance matrices for all states of an HMP for a particular digit could be substituted for by the predicted covariance. This is particularly important when only limited data are available for training of the HMP for each word.

VI. COMMENTS

We have generalized a standard result for the covariance of the log-periodogram of a zero-mean Gaussian process by relaxing the assumption that the spectral components at various frequencies are statistically independent. An interesting related problem is that of estimating the covariance of the log-power spectral density estimate of an autoregressive process given by

$$\log \left( \hat{\sigma}^2 / \hat{A}_0^2 \right).$$

This, however, turns out to be a much harder problem since $\hat{A}_0$ has nonzero mean that asymptotically is given by $A_0$. An asymptotic result for this problem was provided by Merhav and Lee [8].

REFERENCES