MAP Model Order Selection Rule for 2-D Sinusoids in White Noise

Mark Kliger, *Student Member, IEEE,* and Joseph M. Francos, *Senior Member, IEEE*

***Abstract*—We consider the problem of jointly estimating the number as well as the parameters of two-dimensional (2-D) sinusoidal signals, observed in the presence of an additive white Gaussian noise field. Existing solutions to this problem are based on model order selection rules and are derived for the parallel one-dimensional (1-D) problem. These criteria are then adapted to the 2-D problem using heuristic arguments. Employing asymptotic considerations, we derive a maximum *a posteriori* (MAP) model order selection criterion for jointly estimating the parameters of the 2-D sinusoids and their number. The proposed model order selection rule is strongly consistent. As an example, the model order selection criterion is applied as a component in an algorithm for parametric estimation and synthesis of textured images.**

***Index Terms*—Model order selection, maximum *a posteriori* estimation, random fields, 2-D parameter estimation, 2-D sinusoids, texture parametric model.**

I. Introduction

WE consider the problem of jointly estimating the number as well as the parameters of two-dimensional (2-D) sinusoidal signals observed in the presence of an additive white Gaussian noise field. This problem is, in fact, a special case of a much more general problem: From the 2-D Wold-like decomposition [6], we have that any 2-D regular and homogeneous discrete random field can be represented as a sum of two mutually orthogonal components: a purely-indeterministic field and a deterministic one. The deterministic component is further orthogonally decomposed into a harmonic field and a countable number of mutually orthogonal evanescent fields. In this paper, we consider the special case where the deterministic component consists of a finite (unknown) number of harmonic components, while the purely-indeterministic component is assumed to be a white noise field.

A solution to this problem is an essential component in many image processing and multimedia data processing applications. For example, in indexing and retrieval systems of multimedia data that employ the textural information in the imagery components of the data, e.g., [23], the identification of similar textured surfaces, is highly sensitive to errors in estimating the orders of the models of the deterministic components of the textures. More specifically, this indexing approach employs the 2-D Wold decomposition based parametric model of each textured segment in the image as the index to this segment. Therefore, an accurate and robust procedure for estimating the orders as well as the parameters of the models of the deterministic components of the textures is an essential component in any such indexing and retrieval system. Similar requirements are posed by parametric content-based image coding and representation methods.

The same type of problem, i.e., joint estimation of the model order and the parameters for a sum of 2-D sinusoidal signals observed in additive noise, naturally arises in processing 2-D SAR data, and in space time adaptive processing (STAP) of airborne radar data. In these problems, however, the observed random field is complex valued, where for each scatterer one frequency parameter corresponds to the range information, whereas the second frequency parameter is the Doppler.

Many algorithms have been derived to estimate the parameters of sinusoids observed in additive white Gaussian noise. Most of these assume the number of sinusoids is *a priori* known. However this assumption does not always hold in practice. In the past three decades the problem of model order selection for 1-D signals has received considerable attention. Existing model order selection rules can be classified to two classes: algebraic criteria and information-theoretic criteria. The algebraic criteria (see, e.g., [8], [24], [25], [27], and the references therein) employ the eigenvalue or the singular value decomposition to the sample covariance matrix of the data to determine the number of dominant sinusoidal components. Information theoretic model order selection rules are based (directly or indirectly) on three popular criteria: The Akaike information criterion (AIC) [1], the minimum description length (MDL) [19], [20], and the maximum *a posteriori* (MAP) probability criterion [21]. All these criteria have a common form composed of two terms: a data term and a penalty term, where the data term is the log-likelihood function evaluated for the assumed model.

However, most of the papers dedicated to the problem of model order selection are concerned with various models of one-dimensional (1-D) signals, while the problem of modeling multidimensional fields has received considerably less attention. To the best of our knowledge no such criterion has been rigorously developed, and adaptation of existing solutions that were derived for 1-D data models may be misleading. Stoica *et al.*, [22] proposed a cross-validation selection rule and demonstrated its asymptotic equivalence to the Generalized Akaike Information Criterion (GAIC). The suggested criterion is not derived for any specific model. The penalty term is given by $kK(N)$, where k is the number of model parameters, N is the length of the observed data vector, and $K(N)$ is some penalty term, which is a function of N. In [16], this criterion is employed to detect the

Manuscript received June 3, 2003; revised August 3, 2004. This work was supported in part by the Israel Ministry of Science under Grants 1233198 and 1233301. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Derong Li.

The authors are with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva 84105, Israel (e-mail: francos@ee.bgu.ac.il).

Digital Object Identifier 10.1109/TSP.2005.849203

number of sinusoids in 1-D and 2-D signals. The penalty term for 2-D signals is the same as in the 1-D case. The penalty parameter is chosen as $K(N) = c \log\log N$, where $c > 2$. Stoica *et al.* in [22] and Li *et al.* in [16] arrived at this choice of $K(N)$ by using consistency arguments based on [10], [11]. However, in [10] and [11], consistency of an order-selection criterion for autoregressive moving average (ARMA) models is proved, while the model considered in [16] is that of sinusoids in noise. Moreover, for the data model of 1-D sinusoids observed in white noise, Quinn [17] derives conditions for strong consistency of any model order selection criterion. The penalty term of the criterion in [16] does not satisfy Quinn's consistency conditions, even for the 1-D problem.

Djuric [5] proposed a MAP order selection rule for 1-D sinusoids observed in additive white noise. Kavalieris and Hannan [13] prove strong consistency of a criterion that indirectly employs the MDL principle. In this framework the observation noise is modeled as an autoregression of an unknown order. In the special case where the noise process in [13] is assumed to be a white noise process, the resulting criterion is identical to the MAP criterion derived in [5]. In this paper, following the information theoretic approach and the Bayesian methodology, we derive a MAP model order selection criterion for jointly estimating the number and the parameters of 2-D sinusoids observed in the presence of an additive white Gaussian noise field. Moreover, in [15], the strong consistency of the derived criterion is established.

This paper is organized as follows: In Section II, we define our notations, whereas in Section III, we formally define the MAP model order selection problem. The MAP model order selection criterion is derived in Section IV. In Section V, we provide some numerical examples and Monte Carlo simulations to better illustrate the performance of the proposed criterion. In Section VI, we present and analyze texture estimation and synthesis results obtained by applying the proposed model selection rule to a number of natural textures. In Section VII, we provide our conclusions.

II. Notations and Definitions

Let $\{y(n,m)\}, (n,m) \in \Psi(S,T)$, where $\Psi(S,T) = \{(i,j) \mid 0 \le i \le S-1, 0 \le j \le T-1\}$ is the observed 2-D real valued random field such that

$$y(n,m) = h(n,m) + u(n,m). \tag{1}$$

The field $\{u(n,m)\}$ is a 2-D zero mean, white Gaussian field with finite variance σ^2. The field $\{h(n,m)\}$ is the harmonic random field

$$h(n,m) = \sum_{i=1}^{k} C_i \cos(n\omega_i + m\nu_i) + G_i \sin(n\omega_i + m\nu_i) \tag{2}$$

where (ω_i, ν_i) are the spatial frequencies of the ith harmonic. The C_i's and G_i's are the unknown amplitudes of the sinusoidal components in the observed realization. Obviously, it is assumed that $k \ll \min(S,T)$.

Let us define the following matrix notations:

$$\mathbf{y} = [y(0,0), \ldots, y(0,T-1), y(1,0), \ldots \\ y(1,T-1), \ldots, y(S-1,T-1)]^T. \tag{3}$$

The vectors $\mathbf{u}$ and $\mathbf{h}$ are similarly defined. Rewriting (1), we have $\mathbf{y} = \mathbf{h} + \mathbf{u}$. Let $\mathbf{\Gamma}$ denote the covariance matrix of the observed field. Thus, $\mathbf{\Gamma} = \sigma^2 \mathbf{I}_{ST\times ST}$, where $\mathbf{I}_{ST\times ST}$ is an $ST \times ST$ identity matrix, and $|\mathbf{\Gamma}| = \sigma^{2ST}$. In addition, define

$$\mathbf{a} = [C_1, G_1, C_2, G_2, \ldots, C_k, G_k]^T. \tag{4}$$

Let

$$\mathbf{e}_i = \Big[e^{j[0\omega_i + 0\nu_i]}, e^{j[0\omega_i + 1\nu_i]}, \ldots \\ e^{j[0\omega_i + (T-1)\nu_i]}, \ldots, e^{j[(S-1)\omega_i + (T-1)\nu_i]}\Big]^T \tag{5}$$

and let us define the following $ST \times 2k$ matrix:

$$\mathbf{D} = [\mathrm{Re}(\mathbf{e}_1), \mathrm{Im}(\mathbf{e}_1), \mathrm{Re}(\mathbf{e}_2), \mathrm{Im}(\mathbf{e}_2), \ldots, \mathrm{Re}(\mathbf{e}_k), \mathrm{Im}(\mathbf{e}_k)]. \tag{6}$$

Using the foregoing notations. we have that

$$\mathbf{y} = \mathbf{Da} + \mathbf{u}. \tag{7}$$

Let $\{\Psi_i\}$ be a sequence of rectangles such that $\Psi_i = \{(n,m) \in \mathcal{Z}^2 \mid 0 \le n \le S_i - 1, 0 \le m \le T_i - 1\}$.

Definition 1: The sequence of subsets $\{\Psi_i\}$ is said to tend to infinity (we adopt the notation $\Psi_i \to \infty$) as $i \to \infty$ if $\lim_{i\to\infty} \min(S_i, T_i) = \infty$, and $0 < \lim_{i\to\infty} (S_i/T_i) < \infty$. To simplify notations, we will omit in the following the subscript i. Thus, the notation $\Psi(S,T) \to \infty$ implies that both S and T tend to infinity as functions of i and at roughly the same rate.

Let $\boldsymbol{\theta}_k \in \mathbf{\Theta}_k$ denote the parameter vector of the harmonic field, i.e.,

$$\boldsymbol{\theta}_k = [C_1 \quad G_1 \quad \omega_1 \quad \nu_1 \quad \cdots \quad C_k \quad G_k \quad \omega_k \quad \nu_k]^T \tag{8}$$

where for all l, C_l, G_l are real and bounded. Assume further that $\omega_l, \nu_l \in [0, 2\pi)$, where $\min(|\omega_l - \omega_j|) \ge \delta$ or $\min(|\nu_l - \nu_j|) \ge \delta$ for $l \ne j$. Hence, the parameter space $\mathbf{\Theta}_k$ is a subset of the $4k$ dimensional Euclidian space. By the above assumption, we further conclude that $\mathbf{D}$ has rank $2k$ and that the corresponding $2k \times 2k$ Gram matrix $\mathbf{D}^T\mathbf{D}$ is of rank $2k$ as well.

III. MAP Model Selection Criterion

Let $p(k)$ denote the *a priori* probability of the kth model, where k denotes the unknown number of sinusoidal components in the data model given by (1) and (2).

It is assumed that there are Q competing models, where $Q > K$ (K being the actual number of sinusoidal components) and where each model is equiprobable. That is

$$p(k) = \frac{1}{Q}, \quad k \in Z_Q \tag{9}$$

where

$$Z_Q = \{0, 1, 2, \ldots, Q-1\}. \tag{10}$$

The MAP estimate of K is the value of k that maximizes the *a posteriori* probability $p(k \,|\, \mathbf{y})$, where $k \in Z_Q$. More specifically

$$\begin{aligned}\hat{K}_{\mathrm{MAP}} &= \arg\max_{k\in Z_Q}\{p(k \,|\, \mathbf{y})\}\\ &= \arg\max_{k\in Z_Q}\left\{\frac{p(\mathbf{y} \,|\, k)p(k)}{p(\mathbf{y})}\right\}\\ &= \arg\max_{k\in Z_Q}\{p(\mathbf{y} \,|\, k)\}\\ &= \arg\max_{k\in Z_Q}\{\log p(\mathbf{y} \,|\, k)\}\end{aligned} \tag{11}$$

where $p(\mathbf{y} \,|\, k)$ denotes the marginal probability of $\mathbf{y}$, given that there are k sinusoidal components in the data.

Let

$$\mathbf{w} = [\omega_1, \omega_2, \ldots, \omega_k, \nu_1, \nu_2, \ldots, \nu_k]^T. \tag{12}$$

In addition, let $\mathcal{R}^+$ denote the positive real line, and let $\Omega_k = [0, 2\pi)^{2k}$. Thus, we have that $\sigma \in \mathcal{R}^+, \mathbf{a} \in \mathcal{R}^{2k}$, and $\mathbf{w} \in \Omega_k$. Using these notations, the marginal probability density $p(\mathbf{y} \,|\, k)$ is expressed by

$$\begin{aligned}&p(\mathbf{y} \,|\, k)\\ &= \int_{\Omega_k}\int_{\mathcal{R}^+}\int_{\mathcal{R}^{2k}} p(\mathbf{y} \,|\, k, \mathbf{w}, \sigma, \mathbf{a})p(\mathbf{w}, \sigma, \mathbf{a} \,|\, k)\, d\mathbf{a}\, d\sigma\, d\mathbf{w}\end{aligned} \tag{13}$$

where $p(\mathbf{w}, \sigma, \mathbf{a} \,|\, k)$ is the *a priori* probability of $\mathbf{w}, \sigma$, and $\mathbf{a}$, given that there exist k sinusoidal components in the observed data.

IV. Derivation of the Criterion

A. Priors Selection

Inspecting (11) and (13), we conclude that finding $\hat{m}_{\mathrm{MAP}}$, using the observed data only, requires that some assumptions be made regarding the prior distribution of the model parameters $p(\mathbf{w}, \sigma, \mathbf{a} \,|\, k)$. Clearly, our goal is to derive a model selection rule based on a noninformative prior about the parameters. In other words, the selected prior should be chosen such that it represents the lack of *a priori* knowledge of the values of the problem parameters, before the data is observed. (See, e.g., [2] for a detailed discussion of the problem of choosing noninformative priors).

Clearly

$$p(\mathbf{w}, \sigma, \mathbf{a} \,|\, k) = p(\sigma, \mathbf{a} \,|\, \mathbf{w}, k)p(\mathbf{w} \,|\, k). \tag{14}$$

Since the sinusoidal frequencies are assumed independent of each other (i.e., that they are not harmonically related), the lack of *a priori* knowledge of the frequencies is modeled by assuming the frequencies (ω_i, ν_i) to be uniformly distributed in Ω_k. Thus

$$p(\mathbf{w} \,|\, k) = \frac{1}{(2\pi)^{2k}}. \tag{15}$$

Given that $\mathbf{w}$ and k are known, $\mathbf{D}$ is also known, and the observation model (7) becomes a linear regression model, where the observations are subject to a zero mean white Gaussian observation noise with variance σ^2, such that $\mathbf{a}, \sigma$ are independent but unknown. Hence

$$p(\sigma, \mathbf{a} \,|\, \mathbf{w}, k) = p(\mathbf{a} \,|\, \mathbf{w}, k)p(\sigma \,|\, \mathbf{w}, k). \tag{16}$$

For this problem, it is shown in [2] that in the space defined by $\mathbf{a}$ and $\log\sigma$, the shape of the likelihood function (given here by $p(\mathbf{y} \,|\, k, \mathbf{w}, \sigma, \mathbf{a})$) is "data translated," i.e., it is invariant to translations that result from the different values these parameters assume in different realizations of the observed data. Hence, the idea that little is known *a priori* relative to the information contained in the observed data is expressed by choosing a prior distribution such that $p(\mathbf{a} \,|\, \mathbf{w}, k)$ and $p(\log\sigma \,|\, \mathbf{w}, k)$ are *locally* uniform [2] or, equivalently, that

$$p(\mathbf{a} \,|\, \mathbf{w}, k) = \gamma(k) \tag{17}$$

and

$$p(\sigma \,|\, \mathbf{w}, k) = \xi\sigma^{-1} \tag{18}$$

where ξ is some finite positive constant, and $\gamma(k)$ is some finite positive function of k, which is a *constant* for any given k. However, (17) and (18) result in improper prior distribution (16), if assumed valid on $\mathcal{R}^+ \times \mathcal{R}^{2k}$. We therefore emphasize that (17) and (18) represent only the *local behavior* of the prior distribution in the region where the likelihood function is appreciable and not over the entire admissible range. This clearly follows from the assumptions on the boundedness of both the amplitudes of the sinusoids and the noise variance. In other words, the priors (17) and (18) represent the true priors only over the range where the likelihood function is appreciable, whereas, in fact, the priors decay to zero outside this range to ensure they represent proper probability density functions. We further elaborate on this point in the next subsection, where the likelihood function is employed.

B. Evaluation of the a Posteriori Distribution

In this subsection, we derive an approximate expression for the *a posteriori* probability distribution $p(\mathbf{y} \,|\, k)$ given in (13). Since the noise field $\{u(n, m)\}$ is Gaussian, we have, using (7)

$$\begin{aligned}&p(\mathbf{y} \,|\, k, \mathbf{w}, \sigma, \mathbf{a})\\ &\quad = p(\mathbf{u} \,|\, \sigma)\\ &\quad = (2\pi)^{-\frac{ST}{2}}|\boldsymbol{\Gamma}|^{-1/2}\exp\left\{-\frac{1}{2}\mathbf{u}^T\boldsymbol{\Gamma}^{-1}\mathbf{u}\right\}\\ &\quad = (2\pi\sigma^2)^{-\frac{ST}{2}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{D}\mathbf{a})^T(\mathbf{y} - \mathbf{D}\mathbf{a})\right\}.\end{aligned} \tag{19}$$

Let

$$\begin{aligned}\hat{\mathbf{a}} &= (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{y}\\ \hat{\mathbf{y}} &= \mathbf{D}\hat{\mathbf{a}}\end{aligned} \tag{20}$$

and let $\mathbf{P}^{\perp}$ denote the projection matrix defined by

$$\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{D}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T. \tag{21}$$

Using these notations, we have that

$$\begin{aligned}(\mathbf{y}-\mathbf{D}\mathbf{a})^T(\mathbf{y}-\mathbf{D}\mathbf{a}) &= (\mathbf{y}-\hat{\mathbf{y}})^T(\mathbf{y}-\hat{\mathbf{y}}) + (\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{D}^T\mathbf{D}(\mathbf{a}-\hat{\mathbf{a}}) \\ &= \mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y} + (\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{D}^T\mathbf{D}(\mathbf{a}-\hat{\mathbf{a}}). \end{aligned} \tag{22}$$

Applying the priors and evaluating the marginal distribution, we have

$$\begin{aligned} p(\mathbf{y},\mathbf{w},\sigma|k) &= \int_{\mathcal{R}^{2k}} p(\mathbf{y}|k,\mathbf{w},\sigma,\mathbf{a})p(\mathbf{w},\sigma,\mathbf{a}|k)d\mathbf{a} \\ &= p(\sigma|\mathbf{w},k)p(\mathbf{w}|k)\int_{\mathcal{R}^{2k}}(2\pi\sigma^2)^{-\frac{ST}{2}} \\ &\quad \times \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{D}\mathbf{a})^T(\mathbf{y}-\mathbf{D}\mathbf{a})\right\}p(\mathbf{a}|\mathbf{w},k)d\mathbf{a} \\ &= p(\sigma|\mathbf{w},k)p(\mathbf{w}|k)(2\pi\sigma^2)^{-\frac{ST}{2}}\exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}\right\} \\ &\quad \times \int_{\mathcal{R}^{2k}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{D}^T\mathbf{D}(\mathbf{a}-\hat{\mathbf{a}})\right\}p(\mathbf{a}|\mathbf{w},k)d\mathbf{a}. \end{aligned} \tag{23}$$

It is well known that

$$(\sqrt{2\pi}\sigma)^{-2k}|\mathbf{D}^T\mathbf{D}|^{1/2}\int_{\mathcal{R}^{2k}} \times\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{D}^T\mathbf{D}(\mathbf{a}-\hat{\mathbf{a}})\right\}\,d\mathbf{a} = 1. \tag{24}$$

Hence, for every $\epsilon > 0$, there exists $\eta(\epsilon) > 0$ such that for all $\eta > \eta(\epsilon)$

$$\begin{aligned}(\sqrt{2\pi}\sigma)^{-2k}|\mathbf{D}^T\mathbf{D}|^{1/2} &\times \int_{\eta}^{\infty}\cdots\int_{\eta}^{\infty}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{D}^T\mathbf{D}(\mathbf{a}-\hat{\mathbf{a}})\right\}d\mathbf{a} \\ &< \frac{\epsilon}{2}. \end{aligned} \tag{25}$$

Let $\mathcal{A}_k^{\epsilon} = [-\eta(\epsilon),\eta(\epsilon)]^{2k}$. Following the discussion in the previous subsection, it is assumed that on $\mathcal{A}_k^{\epsilon}$, the prior on the amplitude vector is a constant given by (17) for any given k, whereas outside this subset, it decays to zero to ensure that it represents a proper probability density function. Hence, we have that

$$\begin{aligned} p(\mathbf{y},\mathbf{w},\sigma|k) &\geq p(\sigma|\mathbf{w},k)p(\mathbf{w}|k)(2\pi\sigma^2)^{-\frac{ST}{2}}\exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}\right\} \\ &\quad \times\int_{\mathcal{A}_k^{\epsilon}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{D}^T\mathbf{D}(\mathbf{a}-\hat{\mathbf{a}})\right\}p(\mathbf{a}|\mathbf{w},k)d\mathbf{a} \\ &= p(\sigma|\mathbf{w},k)p(\mathbf{w}|k)\gamma(k)(2\pi\sigma^2)^{-\frac{ST}{2}}\exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}\right\} \\ &\quad \times\int_{\mathcal{A}_k^{\epsilon}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{D}^T\mathbf{D}(\mathbf{a}-\hat{\mathbf{a}})\right\}d\mathbf{a} \\ &> p(\sigma|\mathbf{w},k)p(\mathbf{w}|k)\gamma(k)(2\pi\sigma^2)^{-\frac{ST}{2}} \\ &\quad \times\exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}\right\}\frac{(\sqrt{2\pi}\sigma)^{2k}}{|\mathbf{D}^T\mathbf{D}|^{1/2}}(1-\epsilon) \end{aligned} \tag{26}$$

where the last inequality results from bounding the integral using (24) and (25). On the other hand

$$\begin{aligned} p(\mathbf{y},\mathbf{w},\sigma|k) &\leq p(\sigma|\mathbf{w},k)p(\mathbf{w}|k)\gamma(k)(2\pi\sigma^2)^{-\frac{ST}{2}}\exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}\right\} \\ &\quad \times\int_{\mathcal{R}^{2k}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{a}-\hat{\mathbf{a}})^T\mathbf{D}^T\mathbf{D}(\mathbf{a}-\hat{\mathbf{a}})\right\}d\mathbf{a} \\ &= p(\sigma|\mathbf{w},k)p(\mathbf{w}|k)\gamma(k)(2\pi\sigma^2)^{-\frac{ST}{2}} \\ &\quad \times\exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}\right\}\frac{(\sqrt{2\pi}\sigma)^{2k}}{|\mathbf{D}^T\mathbf{D}|^{1/2}}. \end{aligned} \tag{27}$$

Since ϵ is arbitrarily small $p(\mathbf{y},\mathbf{w},\sigma\,|\,k)$ is approximated to an arbitrarily small error by

$$p(\mathbf{y},\mathbf{w},\sigma\,|\,k) \approx p(\sigma\,|\,\mathbf{w},k)p(\mathbf{w}\,|\,k)\gamma(k)(2\pi\sigma^2)^{-\frac{ST}{2}} \times\exp\left\{-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}\right\}\frac{(\sqrt{2\pi}\sigma)^{2k}}{|\mathbf{D}^T\mathbf{D}|^{1/2}}. \tag{28}$$

Next, we evaluate $p(\mathbf{y},\mathbf{w}\,|\,k)$. Substituting (28) and using similar considerations, we have

$$\begin{aligned} p(\mathbf{y},\mathbf{w}\,|\,k) &= \int_{\mathcal{R}^+} p(\mathbf{y},\mathbf{w},\sigma\,|\,k)\,d\sigma \\ &\approx \gamma(k)\xi 2^{-2k-1}\pi^{-\frac{ST+2k}{2}}\Gamma\left(\frac{ST-2k}{2}\right) \\ &\quad \times|\mathbf{D}^T\mathbf{D}|^{-\frac{1}{2}}(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})^{-\frac{ST-2k}{2}} \end{aligned} \tag{29}$$

where $\Gamma(\cdot)$ is the standard Gamma function (see, e.g., [9] for the integration result).

Finally, to obtain an expression for the conditional probability $p(\mathbf{y}\,|\,k)$, we have to evaluate

$$p(\mathbf{y}\,|\,k) = \int_{\Omega_k} p(\mathbf{y},\mathbf{w}\,|\,k)\,d\mathbf{w}. \tag{30}$$

Since a direct analytic solution of this integration problem does not exist, we derive an approximate solution, employing the Laplace integration method (see, e.g., [4] and [26]). The Laplace method considers an integral of the form

$$\int_{\mathbf{a}}^{\mathbf{b}} g(\mathbf{t})e^{xh(\mathbf{t})}\,d\mathbf{t}$$

where $\mathbf{a},\mathbf{b},\mathbf{t},\boldsymbol{\tau}$ are vectors, x is a large positive parameter, and $h(\mathbf{t})$ is real. The approximation is based on the observation that if $h(\mathbf{t})$ has a maximum at $\mathbf{t}=\boldsymbol{\tau}$ and $h(\mathbf{t}) < h(\boldsymbol{\tau})$ when $\mathbf{t}\neq\boldsymbol{\tau}$, whereas $g(\boldsymbol{\tau})\neq 0$; then, for large x, the modulus of the integrand will have a sharp maximum at a point very close to $\boldsymbol{\tau}$, and most of the contribution to the integral will arise from the immediate vicinity of this maximum point. The integral can then be evaluated approximately by expanding both g and h in the neighborhood of $\mathbf{t}=\boldsymbol{\tau}$.

Rewrite (29) in the following form:

$$p(\mathbf{y},\mathbf{w}|k) \approx C(k)|\mathbf{D}^T\mathbf{D}|^{-\frac{1}{2}}\exp\left\{-\frac{ST-2k}{2}\log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})\right\} \tag{31}$$

where $C(k) = \gamma(k)\xi 2^{-2k-1}\pi^{-((ST+2k)/2)}\Gamma((ST-2k)/2)$ is a function of k only. Let $\hat{\mathbf{w}}$ denote the ML estimate of $\mathbf{w}$ for the observed realization, based on the data model (1) and (2) and assuming the model order is k. In addition, let $\hat{\mathbf{D}}$ and $\hat{\mathbf{P}}^{\perp}$ denote the matrices $\mathbf{D}$ and $\mathbf{P}^{\perp}$, respectively, with $\mathbf{w}$ substituted by its ML estimate $\hat{\mathbf{w}}$. It is well known that for the data model (1) and (2), the maximum likelihood estimate of $\mathbf{w}$ is obtained through minimization of the quadratic form $\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}$. Hence, the minimum point of $\log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})$ is obtained by substituting $\mathbf{w}$ with its ML estimate.

Substituting (31) into (30), we observe that for any given realization of $\mathbf{y}$, the integration is carried out over $\mathbf{w}$, while the observations vector is treated as a vector of known constants. Due to its definition, it is clear that $\mathbf{D}$ is full rank and that $|\mathbf{D}^T\mathbf{D}|^{-1/2}$ is a continuous function of the unknown frequency parameters. Since the approximate solution is a function of $\hat{\mathbf{w}}$, we employ the almost sure convergence of $\hat{\mathbf{w}}$ to the correct value of $\mathbf{w}$ (see Lemma 2 in Appendix A) to actually evaluate (30). Since $\hat{\mathbf{w}}$ converges a.s. to the correct value of $\mathbf{w}$, $\hat{\mathbf{D}}$ is a.s. full rank. Hence, $|\hat{\mathbf{D}}^T\hat{\mathbf{D}}|^{-1/2} \neq 0$ a.s. Therefore, for each realization, the conditions required to employ the Laplace asymptotic approximation to evaluate (30) around the ML estimate of $\mathbf{w}$ obtained for this realization are satisfied (as in a standard deterministic problem). Thus, as $\Psi(S,T) \to \infty$

$$\begin{aligned} p(\mathbf{y}\,|\,k) &\approx \int_{\Omega_k} C(k)|\mathbf{D}^T\mathbf{D}|^{-\frac{1}{2}} \exp\left\{-\frac{ST-2k}{2}\log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})\right\} d\mathbf{w} \\ &\approx C(k)|\hat{\mathbf{D}}^T\hat{\mathbf{D}}|^{-\frac{1}{2}} \exp\left\{-\frac{ST-2k}{2}\log(\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y})\right\} \\ &\quad \times (2\pi)^k|\hat{\mathbf{H}}_{\mathrm{ML}}|^{-\frac{1}{2}}\left(\frac{ST-2k}{2}\right)^{-k}(1+o(1)) \\ &= C(k)|\hat{\mathbf{D}}^T\hat{\mathbf{D}}|^{-\frac{1}{2}}(\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y})^{-\frac{ST-2k}{2}}2^{2k}\pi^k|\hat{\mathbf{H}}_{\mathrm{ML}}|^{-\frac{1}{2}} \\ &\quad \times (ST-2k)^{-k}(1+o(1)) \end{aligned} \tag{32}$$

where we have

$$\mathbf{H}_{\mathrm{ML}} = \begin{bmatrix} \frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial \mathbf{w}_1^2} & \frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial \mathbf{w}_1\partial \mathbf{w}_2} & \cdots & \frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial \mathbf{w}_1\partial \mathbf{w}_{2k}} \\ \frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial \mathbf{w}_2\partial \mathbf{w}_1} & \frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial \mathbf{w}_2^2} & \cdots & \frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial \mathbf{w}_2\partial \mathbf{w}_{2k}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial \mathbf{w}_{2k}\partial \mathbf{w}_1} & \frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial \mathbf{w}_{2k}\partial \mathbf{w}_2} & \cdots & \frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial \mathbf{w}_{2k}^2} \end{bmatrix} \tag{33}$$

which denotes the Hessian matrix of $\log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})$. When evaluated at $\mathbf{w} = \hat{\mathbf{w}}$, it will be denoted by $\hat{\mathbf{H}}_{\mathrm{ML}}$. Using the next lemma, the computation is completed.

Lemma 1:

$$|\hat{\mathbf{H}}_{\mathrm{ML}}| = \mathcal{C}(\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y})^{-2k}S^{4k}T^{4k}(1+o(1)) \text{ a.s.} \tag{34}$$

where $\mathcal{C}$ denotes some "generic" constant, whose exact value is of no importance to us in the current context (and may vary from usage to usage).

Proof: See Appendix A. □

Substituting (34) and the explicit expression of $C(k)$ into (32), we have

$$\begin{aligned} p(\mathbf{y}\,|\,k) &\approx \mathcal{C}\gamma(k)\xi 2^{-1}\pi^{-\frac{ST}{2}}\Gamma\left(\frac{ST-2k}{2}\right)|\hat{\mathbf{D}}^T\hat{\mathbf{D}}|^{-\frac{1}{2}} \\ &\quad \times \left(\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y}\right)^{-\frac{ST-4k}{2}} S^{-2k}T^{-2k} \\ &\quad \times (ST-2k)^{-k}(1+o(1)) \text{ a.s.} \end{aligned} \tag{35}$$

It is possible to further simplify (35) by observing that $|\mathbf{D}^T\mathbf{D}| = \mathcal{C}S^{2k}T^{2k}(1+o(1))$. [See (56) in Appendix A for the derivation of a similar conclusion]. Since $\hat{\mathbf{w}} \to \mathbf{w}$ a.s., and since $|\mathbf{D}^T\mathbf{D}|$ is a continuous function of $\mathbf{w}$, we have that $|\hat{\mathbf{D}}^T\hat{\mathbf{D}}| = \mathcal{C}S^{2k}T^{2k}(1+o(1))$ a.s. Furthermore, employing the asymptotic properties of the Gamma function the contribution of the $\Gamma((ST-2k)/2)$ is approximated by $((ST)/2)^{-k}\Gamma((ST)/2)$ (see, e.g., [3]).

Substituting these approximations into (35), the final form of the model selection criterion can be readily established for $\Psi(S,T) \to \infty$:

$$\begin{aligned} \hat{K}_{\mathrm{MAP}} &= \arg\min_{k\in Z_Q}\{-\log p(\mathbf{y}\,|\,k)\} \\ &\approx \arg\min_{k\in Z_Q}\left\{\frac{ST-4k}{2}\log(\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y}) + \frac{1}{2}2k\log ST\right. \\ &\quad + k\log ST - \log\Gamma\left(\frac{ST}{2}\right) + 2k\log ST \\ &\quad \left. + k\log(ST-2k) - \log(\gamma(k)) - k\log 2\right\} \\ &= \arg\min_{k\in Z_Q}\left\{\frac{ST-4k}{2}\log(\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y}) + 4k\log ST\right. \\ &\quad \left. + k\log(ST-2k)\right\} \\ &= \arg\min_{k\in Z_Q}\left\{\frac{ST}{2}\log(\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y}) + 5k\log ST\right\} \end{aligned} \tag{36}$$

where in the third equality, the term $\log\Gamma((ST)/2)$ is omitted because it is not a function of k, while $-\log(\gamma(k)) - k\log 2$ is asymptotically negligible compared with the other terms, since they are increasing functions of ST. Finally, $k\log(ST-2k)$ is approximated by $k\log ST$ as $\Psi(S,T) \to \infty$.

Furthermore, recall that K denotes the correct number of sinusoids in the field. Then, it is proved in [15] that as $\Psi(S,T) \to \infty$

$$\hat{K}_{\mathrm{MAP}} \to K \text{ a.s.} \tag{37}$$

V. Numerical Examples

To illustrate the performance of the proposed model order selection rule, we present some numerical examples. In the examples below, the data field was generated with five equiamplitude sinusoidal components, and we define

$$\mathrm{SNR}_i = 10\log\frac{C_i^2+G_i^2}{2\sigma^2}. \tag{38}$$

TABLE I
PERFORMANCE OF MAP CRITERION FOR VARIOUS VALUES OF SNR_i: NUMBER OF TIMES MODEL ORDER k WAS DECIDED, OUT OF 100 EXPERIMENTS

SNR_i	k=0	k=1	k=2	k=3	k=4	k=5	k=6
-20dB	95	5	0	0	0	0	0
-19dB	86	12	2	0	0	0	0
-18dB	33	33	23	11	0	0	0
-17dB	8	17	29	28	18	0	0
-16dB	0	0	15	18	33	34	0
-15dB	0	0	0	3	19	78	0
-14dB	0	0	0	0	2	98	0
-13dB	0	0	0	0	0	100	0
-12dB	0	0	0	0	0	100	0

The noise is a white Gaussian noise field with variance σ^2, which is chosen to yield the desired signal-to-noise ratio (SNR). In these experiments, the signal-to-noise ratio of each component SNR_i varies in the range of -20 dB to -10 dB, in steps of 1 dB. For each SNR, 100 Monte Carlo experiments are performed. The data field dimensions are 64×64. The frequencies of the sinusoidal components are $\{-2\pi 0.155, 2\pi 0.253\}, \{-2\pi 0.155, 2\pi(0.253 + 1/64)\}$, $\{2\pi 0.112, 2\pi 0.274\}, \{2\pi(0.112 + 1/64), 2\pi 0.274\}$, $\{2\pi(0.112 + 1/128), 2\pi(0.274 + 1/64)\}$. The amplitudes are given by $C_i = G_i = 1, i = 1, \ldots, 5$. The performance results of the proposed MAP selection criterion are summarized in Table I for various values of SNR_i. The simulation results, which are tested for model orders ranging from 0 to 100, demonstrate that even for modest dimensions of the observed field, and relatively low SNRs, i.e., as low as − 14 dB, the error rates of the MAP model order selection criterion are very low. In all the experiments, the MAP criterion never estimated the model order to be higher than 5. (Hence, Table I lists the results only for $k = 0, \ldots, 6$.)

VI. EXPERIMENTAL RESULTS

In this section, we present some experimental results to illustrate the performance of the suggested order selection algorithm on images of natural textures. In [7], a parametric texture model that is based on the 2-D Wold decomposition was presented. It was shown that the Wold decomposition based texture model is successful in estimating the texture parameters and in reproducing the original texture using *only* the estimated parameters. However, for estimating the number of harmonic components in the given texture field an *ad-hoc* procedure was adopted. It is based on a search for the isolated peaks of the magnitude of the transfer function of the observed texture field linear predictor. In this section, we demonstrate the performance of the texture analysis/synthesis procedure when the *ad-hoc* model order selection rule is replaced by the MAP order selection rule proposed in the previous sections.

In general, the assumption that the purely indeterministic component is a white noise field does not hold for natural textures. However, for structured textures where the harmonic component is much stronger than the purely indeterministic component, we can assume *only* for the purpose of estimating the order of the model that the purely indeterministic component is a white noise field and apply the MAP model order selection rule derived in the previous sections to estimate the

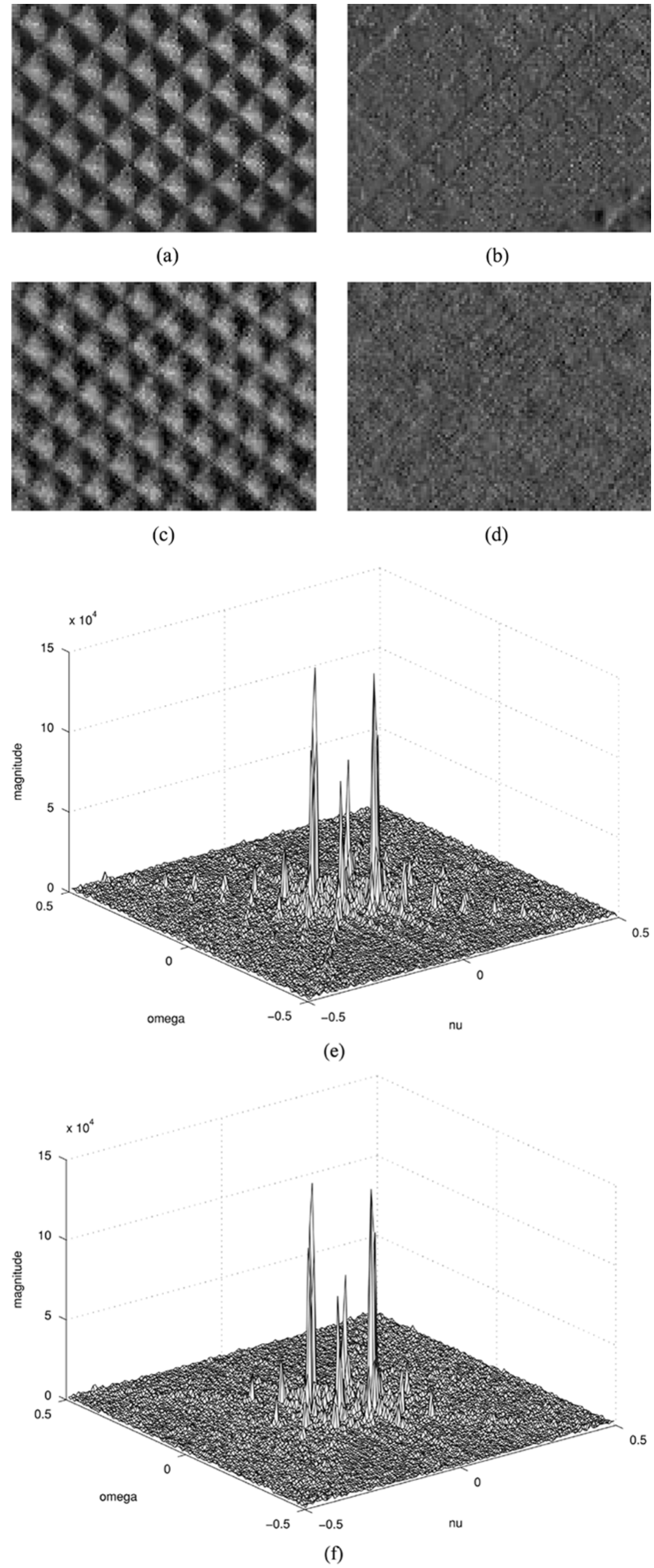

Fig. 1. Natural texture with 15 estimated sinusoids. (a) Original texture. (b) Residual component. (c) Synthesized texture. (d) Synthesized purely indeterministic component. (e) Magnitude of DFT of original texture. (f) Magnitude of DFT of synthesized texture.

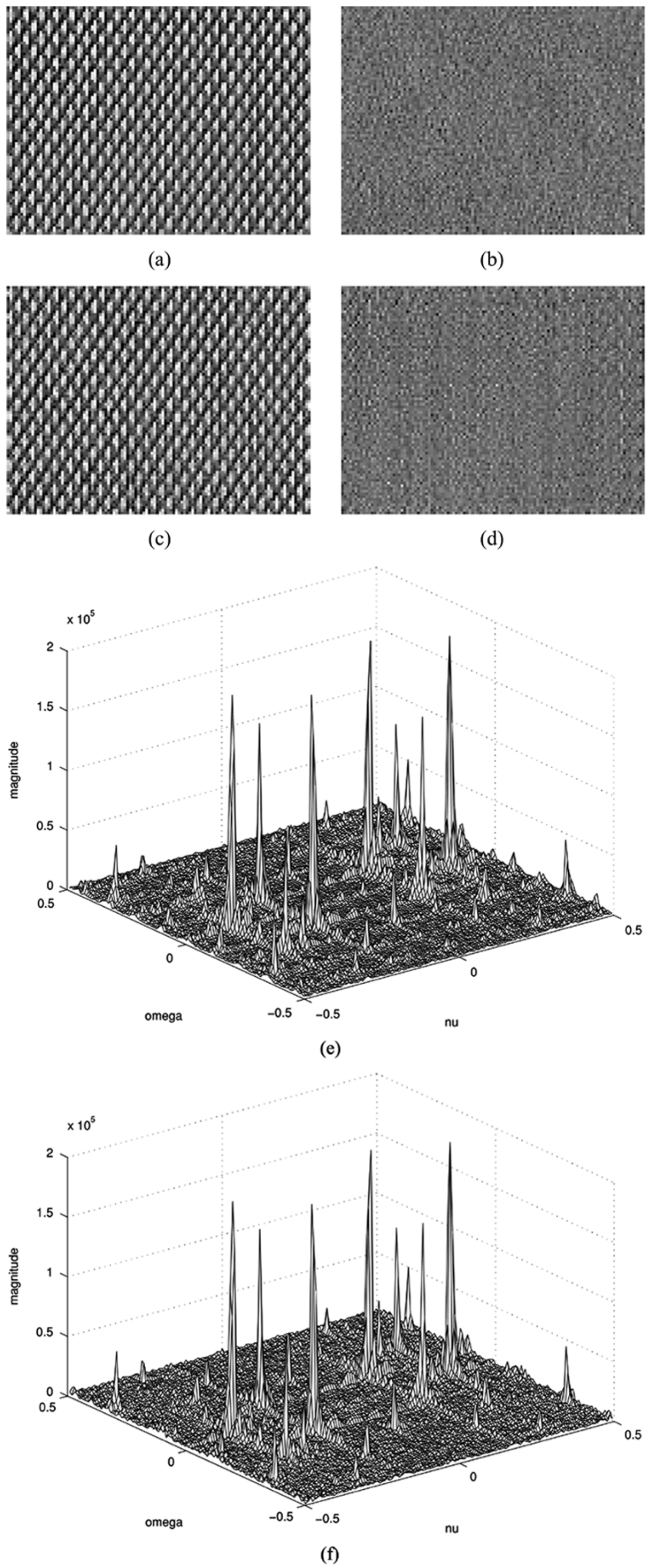

Fig. 2. Natural texture with 22 estimated sinusoids. (a) Original texture. (b) Residual component. (c) Synthesized texture. (d) Synthesized purely indeterministic component. (e) Magnitude of DFT of original texture. (f) Magnitude of DFT of synthesized texture.

number of harmonic components. Nevertheless, achieving high-quality synthesis from the estimated parameters requires modeling and estimation of the colored nature of the purely indeterministic component. Thus, in the following, we assume that the purely indeterministic component is a real-valued Gaussian AR field

$$w(n,m) = -\sum_{(k,\ell)\in S_{N,M}\setminus\{(0,0)\}} b(k,\ell)w(n-k,m-\ell) + u(n,m) \quad (39)$$

where $S_{N,M} = \{(i,j)\,|\, i=0, 0 \le j \le M\} \bigcup \{(i,j)\,|\, 1 \le i \le N, -M \le j \le M\}$. The driving noise of the AR model is a zero-mean white Gaussian field. In all the examples, we assume an $S_{(6,6)}$ NSHP AR model for the purely indeterministic component since it is large enough to provide high-quality synthesis results for the tested textures.

All the textures presented here are natural textures, and hence, the true parameters are unknown. The synthesis algorithm reconstructs the original textures using *only* the estimated parameters. In all the examples presented, the original image is such that it can be bounded by a 84×112 pixel box.

The MAP order selection rule involves the evaluation of a ML estimate of the model parameters. In [7], a conditional maximum-likelihood algorithm for jointly estimating the parameters of the harmonic, evanescent, and purely indeterministic components of the texture was developed. However, this algorithm requires the solution of a nonlinear least-squares (NLLS) problem for the spectral support parameters of the harmonic components. Due to the required multidimensional search, this estimator is computationally demanding. We therefore use, in this paper, a suboptimal (relative to the maximum likelihood estimator) but computationally efficient algorithm (since no multidimensional search in the parameter space is required), for estimating the texture model parameters. The algorithm that we use is an iterative, periodogram-based estimation algorithm. In the first stage, the parameters of the harmonic component are estimated, and their contribution to the observed realization is removed. Ideally, the obtained residual is the purely indeterministic component of the texture. In a second stage, a 2-D AR model of the residual is estimated. Note that in this case, where all the deterministic components have already been removed, the procedure of obtaining a maximum-likelihood estimate of the AR model parameters is reduced to a solution of a linear least squares problem.

Figs. 1 and 2 depict examples of natural textures and the results of the synthesis algorithm that employs the estimated parameters. The proposed MAP order selection algorithm is employed to estimate the number of sinusoids in each texture. The number of estimated sinusoids in the textures of Figs. 1 and 2 are 15 and 22, respectively.

VII. Conclusion

In this paper, we have presented a solution to the problem of jointly estimating the number as well as the parameters of 2-D sinusoidal signals observed in the presence of an additive white Gaussian noise field. Following the Bayesian methodology and employing asymptotic considerations, a strongly consistent MAP model order selection criterion has been developed.

Similar to criteria derived for 1-D problems, the proposed criterion has a log-likelihood term and a penalty term.

The performance of the proposed algorithm for finite-dimensional data is illustrated using Monte Carlo simulations. The simulation results demonstrate that even for modest dimensions of the observed field and relatively low SNRs, the error rates of the MAP model order selection criterion are very low. The MAP order selection rule is applicable to a wide variety of problems in which 2-D harmonic components are observed in the presence of an additive Gaussian noise field. Using the derived model order selection rule, we present an improved solution for the problem of parameter estimation and synthesis of natural textures. The synthesis results obtained by the suggested algorithm are both visually and statistically very similar to the originals and, in some cases, indistinguishable.

APPENDIX A

Lemma 1:

$$|\hat{\mathbf{H}}_{\mathrm{ML}}| = \mathcal{C}(\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y})^{-2k}S^{4k}T^{4k}(1+o(1)) \text{ a.s.} \tag{40}$$

where $\mathcal{C}$ denotes some "generic" constant whose exact value is of no importance to us in the current context.

In the Proof of Lemma 1, we will use the following results.

Lemma 2: Let $\hat{\boldsymbol{\theta}}$ and $\hat{\sigma}^2$ denote the ML estimates of the parameter vector $\boldsymbol{\theta}$ of the harmonic component and the variance σ^2 of the noise field in the data model given by (1), (2). Then, as $\Psi(S,T) \to \infty$

$$\hat{\boldsymbol{\theta}} \to \boldsymbol{\theta} \text{ a.s.} \tag{41}$$

$$\hat{\sigma}^2 \to \sigma^2 \text{ a.s.} \tag{42}$$

Proof: The data model given by (1), (2) is the special case of the more general model given in [18]. Therefore, this lemma is a straightforward result of Theorems 1 and 2 in [18]. ∎

Lemma 3: Let $u(n,m)$ be a real-valued 2-D Gaussian white noise field with zero mean and variance σ^2. Then, for any $\tau > 0$ and any integer $l \geq 0$

$$S^{-(l+\frac{1}{2}+\tau)}T^{-\frac{1}{2}} \sup_{\omega,\nu} \left| \sum_{n=0}^{S-1}\sum_{m=0}^{T-1} n^l u(n,m) e^{j[n\omega+m\nu]} \right| \to 0 \text{ a.s. as } \Psi(S,T)\to\infty. \tag{43}$$

Proof: This lemma is the result of a straightforward manipulation of Lemma 3 [18]. ∎

In addition, in the sequel, we will use the following asymptotic results that hold for $\mu \neq 2\pi n$ for any integer n:

$$\sum_{s=0}^{m-1} e^{js\mu} = O(1)$$
$$\sum_{s=0}^{m-1} s e^{js\mu} = O(m)$$
$$\sum_{s=0}^{m-1} s^2 e^{js\mu} = O(m^2). \tag{44}$$

We are now in a position to prove Lemma 1.

Proof: Taking the first and second partial derivatives of $\log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})$ with respect to ω_i and ω_l and evaluating their values at the ML estimate of $\mathbf{w}$, we have

$$\left.\frac{\partial \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial\omega_i}\right|_{\mathbf{w}=\hat{\mathbf{w}}} = (\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y})^{-1} \left.\frac{\partial(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial\omega_i}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \tag{45}$$

and

$$\left.\frac{\partial^2 \log(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial\omega_l\partial\omega_i}\right|_{\mathbf{w}=\hat{\mathbf{w}}} = -(\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y})^{-2}\left(\left.\frac{\partial(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial\omega_l}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\right) \times \left(\left.\frac{\partial(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial\omega_l}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\right) + (\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y})^{-1}\left.\frac{\partial^2(\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y})}{\partial\omega_l\partial\omega_i}\right|_{\mathbf{w}=\hat{\mathbf{w}}}. \tag{46}$$

Since $\hat{\mathbf{w}}$ is the ML estimate of $\mathbf{w}$, it is also an extremal point of $\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}$. Hence

$$\mathbf{y}^T \left.\frac{\partial\mathbf{P}^{\perp}}{\partial\omega_i}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \mathbf{y} = 0. \tag{47}$$

Substituting (47) into (46), we can rewrite $\hat{\mathbf{H}}_{\mathrm{ML}}$ in the following form:

$$\hat{\mathbf{H}}_{\mathrm{ML}} = (\mathbf{y}^T\hat{\mathbf{P}}^{\perp}\mathbf{y})^{-1}\hat{\mathcal{H}}_{\mathrm{ML}} \tag{48}$$

where

$$\hat{\mathcal{H}}_{\mathrm{ML}} = \begin{bmatrix} \frac{\partial^2\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}}{\partial\mathbf{w}_1^2} & \frac{\partial^2\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}}{\partial\mathbf{w}_1\partial\mathbf{w}_2} & \cdots & \frac{\partial^2\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}}{\partial\mathbf{w}_1\partial\mathbf{w}_{2k}} \\ \frac{\partial^2\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}}{\partial\mathbf{w}_2\partial\mathbf{w}_1} & \frac{\partial^2\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}}{\partial\mathbf{w}_2^2} & \cdots & \frac{\partial^2\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}}{\partial\mathbf{w}_2\partial\mathbf{w}_{2k}} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}}{\partial\mathbf{w}_{2k}\partial\mathbf{w}_1} & \frac{\partial^2\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}}{\partial\mathbf{w}_{2k}\partial\mathbf{w}_2} & \cdots & \frac{\partial^2\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}}{\partial\mathbf{w}_{2k}^2} \end{bmatrix} \tag{49}$$

denotes the Hessian matrix of $\mathbf{y}^T\mathbf{P}^{\perp}\mathbf{y}$ evaluated at $\mathbf{w} = \hat{\mathbf{w}}$.

Next, let

$$\mathbf{B} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2j} \\ \frac{1}{2} & -\frac{1}{2j} \end{bmatrix} \tag{50}$$

and define $\mathbf{U}$ as the block diagonal $2k \times 2k$ matrix

$$\mathbf{U} = \begin{bmatrix} \mathbf{B} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B} \end{bmatrix}. \tag{51}$$

In addition, let

$$\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_1^*, \mathbf{e}_2, \mathbf{e}_2^*, \ldots, \mathbf{e}_k, \mathbf{e}_k^*] \tag{52}$$

where $\mathbf{e}_i$ is defined in (5), and $\mathbf{e}_i^*$ denotes the element by element conjugate of $\mathbf{e}_i$. We therefore have that

$$\mathbf{D} = \mathbf{E}\mathbf{U}. \tag{53}$$

Since $\mathbf{D}$ is real valued, $\mathbf{D}^T = \mathbf{D}^H = \mathbf{U}^H\mathbf{E}^H$, where $\mathbf{E}^H$ is the conjugate transpose of $\mathbf{E}$. As the column spaces of $\mathbf{D}$ and $\mathbf{E}$ are identical, the projection matrix $\mathbf{P}^{\perp}$ is also given by

$$\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{E}(\mathbf{E}^H\mathbf{E})^{-1}\mathbf{E}^H \tag{54}$$

where $\mathbf{E}^H\mathbf{E}$ is given by (55), shown at the bottom of the page. Evaluating the elements of $\mathbf{E}^H\mathbf{E}$, we have that for $i = l, \mathbf{e}_i^H\mathbf{e}_i = \mathbf{e}_i^T\mathbf{e}_i^* = ST$, whereas for $i \neq l, \mathbf{e}_i^H\mathbf{e}_l = O(1)$ and $\mathbf{e}_i^T\mathbf{e}_l^* = O(1)$, where we have used (44) and the separability of the 2-D exponentials.

Hence, from (55), (44), and Lemma 2, we have

$$\hat{\mathbf{E}}^H\hat{\mathbf{E}} = ST\mathbf{I}_{2k} + O(1) = \mathbf{E}^H\mathbf{E} \text{ a.s.}$$
$$|\hat{\mathbf{E}}^H\hat{\mathbf{E}}| = S^{2k}T^{2k}(1 + o(1)) \text{ a.s.} \tag{56}$$

where $\hat{\mathbf{E}}$ is the matrix $\mathbf{E}$ with $\mathbf{w}$ substituted by its ML estimate $\hat{\mathbf{w}}$.

To simplify notations, we will, in the following (without limiting the generality of the derivation) evaluate all required derivatives with respect to ω_1. Thus

$$\frac{\partial \mathbf{E}}{\partial \omega_1} = [\dot{\mathbf{e}}_1, \dot{\mathbf{e}}_1^*, \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}, \mathbf{0}] \tag{57}$$

where

$$\dot{\mathbf{e}}_1 = \Big[0, \ldots, 0, je^{j[\omega_1 + 0\nu_1]}, \ldots, je^{j[\omega_1 + (T-1)\nu_1]}, \ldots, 2je^{j[2\omega_1 + 0\nu_1]}, \ldots, (S-1)je^{j[(S-1)\omega_1 + (T-1)\nu_1]}\Big]^T \tag{58}$$

and

$$\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2} = [\ddot{\mathbf{e}}_1, \ddot{\mathbf{e}}_1^*, \mathbf{0}, \mathbf{0}, \ldots, \mathbf{0}, \mathbf{0}] \tag{59}$$

where

$$\ddot{\mathbf{e}}_1 = -\Big[0, \ldots, 0, e^{j[\omega_1 + 0\nu_1]}, \ldots, e^{j[\omega_1 + (T-1)\nu_1]}, \ldots, 4e^{j[2\omega_1 + 0\nu_1]}, \ldots, (S-1)^2 e^{j[(S-1)\omega_1 + (T-1)\nu_1]}\Big]^T. \tag{60}$$

Hence, using (55), we have (61) and (62), shown at the bottom of the page. Using (54), we have

$$\begin{aligned} S^{-3}T^{-1} & \left.\frac{\partial^2 \mathbf{y}^T\mathbf{P}^\perp\mathbf{y}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \\ &= -S^{-3}T^{-1}\mathbf{y}^T\left(\left.\frac{\partial^2\mathbf{E}}{\partial\omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\hat{\mathbf{E}}^H\right. \\ &\quad + \hat{\mathbf{E}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\left.\frac{\partial^2\mathbf{E}^H}{\partial\omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} + \hat{\mathbf{E}}\left.\frac{\partial^2(\mathbf{E}^H\mathbf{E})^{-1}}{\partial\omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\hat{\mathbf{E}}^H \\ &\quad + 2\left.\frac{\partial\mathbf{E}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\left.\frac{\partial\mathbf{E}^H}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \\ &\quad + 2\left.\frac{\partial\mathbf{E}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\left.\frac{\partial(\mathbf{E}^H\mathbf{E})^{-1}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\hat{\mathbf{E}}^H \\ &\quad \left. + 2\hat{\mathbf{E}}\left.\frac{\partial(\mathbf{E}^H\mathbf{E})^{-1}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\left.\frac{\partial\mathbf{E}^H}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\right)\mathbf{y} \\ &= -\left(H_1 + H_1^H + H_2 + H_3 + H_4 + H_4^H\right). \end{aligned} \tag{63}$$

Next, we evaluate each of these terms. From (56), we have that $(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1} = (1/(ST))(\mathbf{I}_{2k} + o(1))$ a.s. Since the limit as $\Psi(S,T) \to \infty$ of the product of any two sequences, where each

$$\mathbf{E}^H\mathbf{E} = \begin{bmatrix} \mathbf{e}_1^H\mathbf{e}_1 & \mathbf{e}_1^H\mathbf{e}_1^* & \mathbf{e}_1^H\mathbf{e}_2 & \mathbf{e}_1^H\mathbf{e}_2^* & \cdots & \mathbf{e}_1^H\mathbf{e}_k & \mathbf{e}_1^H\mathbf{e}_k^* \\ \mathbf{e}_1^T\mathbf{e}_1 & \mathbf{e}_1^T\mathbf{e}_1^* & \mathbf{e}_1^T\mathbf{e}_2 & \mathbf{e}_1^T\mathbf{e}_2^* & \cdots & \mathbf{e}_1^T\mathbf{e}_k & \mathbf{e}_1^T\mathbf{e}_k^* \\ \mathbf{e}_2^H\mathbf{e}_1 & \mathbf{e}_2^H\mathbf{e}_1^* & \mathbf{e}_2^H\mathbf{e}_2 & \mathbf{e}_2^H\mathbf{e}_2^* & \cdots & \mathbf{e}_2^H\mathbf{e}_k & \mathbf{e}_2^H\mathbf{e}_k^* \\ \mathbf{e}_2^T\mathbf{e}_1 & \mathbf{e}_2^T\mathbf{e}_1^* & \mathbf{e}_2^T\mathbf{e}_2 & \mathbf{e}_2^T\mathbf{e}_2^* & \cdots & \mathbf{e}_2^T\mathbf{e}_k & \mathbf{e}_2^T\mathbf{e}_k^* \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{e}_k^H\mathbf{e}_1 & \mathbf{e}_k^H\mathbf{e}_1^* & \mathbf{e}_k^H\mathbf{e}_2 & \mathbf{e}_k^H\mathbf{e}_2^* & \cdots & \mathbf{e}_k^H\mathbf{e}_k & \mathbf{e}_k^H\mathbf{e}_k^* \\ \mathbf{e}_k^T\mathbf{e}_1 & \mathbf{e}_k^T\mathbf{e}_1^* & \mathbf{e}_k^T\mathbf{e}_2 & \mathbf{e}_k^T\mathbf{e}_2^* & \cdots & \mathbf{e}_k^T\mathbf{e}_k & \mathbf{e}_k^T\mathbf{e}_k^* \end{bmatrix}. \tag{55}$$

$$\frac{\partial\mathbf{E}^H\mathbf{E}}{\partial\omega_1} = \begin{bmatrix} 0 & 2\mathbf{e}_1^H\dot{\mathbf{e}}_1^* & \dot{\mathbf{e}}_1^H\mathbf{e}_2 & \dot{\mathbf{e}}_1^H\mathbf{e}_2^* & \cdots & \dot{\mathbf{e}}_1^H\mathbf{e}_k & \dot{\mathbf{e}}_1^H\mathbf{e}_k^* \\ 2\mathbf{e}_1^T\dot{\mathbf{e}}_1 & 0 & \dot{\mathbf{e}}_1^T\mathbf{e}_2 & \dot{\mathbf{e}}_1^T\mathbf{e}_2^* & \cdots & \dot{\mathbf{e}}_1^T\mathbf{e}_k & \dot{\mathbf{e}}_1^T\mathbf{e}_k^* \\ \mathbf{e}_2^H\dot{\mathbf{e}}_1 & \mathbf{e}_2^H\dot{\mathbf{e}}_1^* & & & & & \\ \mathbf{e}_2^T\dot{\mathbf{e}}_1 & \mathbf{e}_2^T\dot{\mathbf{e}}_1^* & & & & & \\ \vdots & \vdots & & & \mathbf{0}_{(2k-2)\times(2k-2)} & & \\ \mathbf{e}_k^H\dot{\mathbf{e}}_1 & \mathbf{e}_k^H\dot{\mathbf{e}}_1^* & & & & & \\ \mathbf{e}_k^T\dot{\mathbf{e}}_1 & \mathbf{e}_k^T\dot{\mathbf{e}}_1^* & & & & & \end{bmatrix} \tag{61}$$

$$\frac{\partial^2\mathbf{E}^H\mathbf{E}}{\partial^2\omega_1} = \begin{bmatrix} 0 & 4\mathbf{e}_1^H\ddot{\mathbf{e}}_1^* & \ddot{\mathbf{e}}_1^H\mathbf{e}_2 & \ddot{\mathbf{e}}_1^H\mathbf{e}_2^* & \cdots & \ddot{\mathbf{e}}_1^H\mathbf{e}_k & \ddot{\mathbf{e}}_1^H\mathbf{e}_k^* \\ 4\mathbf{e}_1^T\ddot{\mathbf{e}}_1 & 0 & \ddot{\mathbf{e}}_1^T\mathbf{e}_2 & \ddot{\mathbf{e}}_1^T\mathbf{e}_2^* & \cdots & \ddot{\mathbf{e}}_1^T\mathbf{e}_k & \ddot{\mathbf{e}}_1^T\mathbf{e}_k^* \\ \mathbf{e}_2^H\ddot{\mathbf{e}}_1 & \mathbf{e}_2^H\ddot{\mathbf{e}}_1^* & & & & & \\ \mathbf{e}_2^T\ddot{\mathbf{e}}_1 & \mathbf{e}_2^T\ddot{\mathbf{e}}_1^* & & & & & \\ \vdots & \vdots & & & \mathbf{0}_{(2k-2)\times(2k-2)} & & \\ \mathbf{e}_k^H\ddot{\mathbf{e}}_1 & \mathbf{e}_k^H\ddot{\mathbf{e}}_1^* & & & & & \\ \mathbf{e}_k^T\ddot{\mathbf{e}}_1 & \mathbf{e}_k^T\ddot{\mathbf{e}}_1^* & & & & & \end{bmatrix}. \tag{62}$$

tends to a finite limit, is the product of the limits of the individual sequences

$$
\begin{aligned}
H_1 &= S^{-3}T^{-1}\mathbf{y}^T \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} (\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\hat{\mathbf{E}}^H\mathbf{y} \\
&= S^{-4}T^{-2}\left(\mathbf{y}^T \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \hat{\mathbf{E}}^H\mathbf{y}\right)(1+o(1)) \\
&= S^{-4}T^{-2}\left(\mathbf{a}^T\mathbf{D}^T \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \hat{\mathbf{E}}^H\mathbf{D}\mathbf{a}\right. \\
&\quad + \mathbf{a}^T\mathbf{D}^T \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \hat{\mathbf{E}}^H\mathbf{u} + \mathbf{u}^T \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \hat{\mathbf{E}}^H\mathbf{D}\mathbf{a} \\
&\quad \left. + \mathbf{u}^T \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \hat{\mathbf{E}}^H\mathbf{u}\right)(1+o(1)) \\
&= \big(M_{11} + M_{21} + M_{21}^T + M_{31}\big)(1+o(1)) \text{ a.s.}
\end{aligned} \tag{64}
$$

Let

$$
\begin{aligned}
\tilde{\mathbf{a}} = \mathbf{U}\mathbf{a} &= \left[\frac{C_1+G_1}{2j}, \frac{C_1-G_1}{2j}, \frac{C_2+G_2}{2j}, \frac{C_2-G_2}{2j}, \ldots \right. \\
&\quad \left. \frac{C_k+G_k}{2j}, \frac{C_k-G_k}{2j}\right]^T \\
&= [a_1, b_1, a_2, b_2, \ldots, a_k, b_k]^T.
\end{aligned} \tag{65}
$$

Rewriting the terms of (64), we have

$$
\begin{aligned}
M_{11} &= S^{-4}T^{-2}\mathbf{a}^T\mathbf{D}^T \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \hat{\mathbf{E}}^H\mathbf{D}\mathbf{a} \\
&= S^{-4}T^{-2}\mathbf{a}^T\mathbf{U}^H\mathbf{E}^H \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \hat{\mathbf{E}}^H\mathbf{E}\mathbf{U}\mathbf{a} \\
&= S^{-4}T^{-2}\tilde{\mathbf{a}}^H\mathbf{E}^H \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \hat{\mathbf{E}}^H\mathbf{E}\tilde{\mathbf{a}}
\end{aligned} \tag{66}
$$

and

$$
M_{21} = S^{-4}T^{-2}\tilde{\mathbf{a}}^H\mathbf{E}^H \left.\frac{\partial^2 \mathbf{E}}{\partial \omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \hat{\mathbf{E}}^H\mathbf{u}. \tag{67}
$$

From (56)–(60), (43), (44), and Lemma 2, we have

$$
\begin{aligned}
M_{11} &= S^{-4}T^{-2}\sum_{l=1}^{k}\left(a_l^*\mathbf{e}_l^H\hat{\ddot{\mathbf{e}}}_1 + b_l^*\mathbf{e}_l^T\hat{\ddot{\mathbf{e}}}_1\right) \\
&\quad \times \sum_{i=1}^{k}\left(a_i\hat{\mathbf{e}}_1^H\mathbf{e}_i + b_i\hat{\mathbf{e}}_1^H\mathbf{e}_i^*\right) \\
&\quad + S^{-4}T^{-2}\sum_{l=1}^{k}\left(a_l^*\mathbf{e}_l^H\hat{\ddot{\mathbf{e}}}_1^* + b_l^*\mathbf{e}_l^T\hat{\ddot{\mathbf{e}}}_1^*\right) \\
&\quad \times \sum_{i=1}^{k}\left(a_i\hat{\mathbf{e}}_1^T\mathbf{e}_i + b_i\hat{\mathbf{e}}_1^T\mathbf{e}_i^*\right) \\
&= -\sum_{l=1}^{k}S^{-3}T^{-1}\left(\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}a_l^*n^2e^{j[n(\hat{\omega}_1-\omega_l)+m(\hat{\nu}_1-\nu_l)]}\right. \\
&\quad \left. + \sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2b_l^*e^{j[n(\hat{\omega}_1+\omega_l)+m(\hat{\nu}_1+\nu_l)]}\right) \\
&\quad \times \sum_{i=1}^{k}S^{-1}T^{-1}\left(\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}a_ie^{-j[n(\hat{\omega}_1-\omega_i)+m(\hat{\nu}_1-\nu_i)]}\right. \\
&\quad \left. + \sum_{n=0}^{S-1}\sum_{m=0}^{T-1}b_ie^{-j[n(\hat{\omega}_1+\omega_i)+m(\hat{\nu}_1+\nu_i)]}\right) \\
&\quad - \sum_{l=1}^{k}S^{-3}T^{-1}\left(\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}a_l^*n^2e^{-l[n(\hat{\omega}_1+\omega_j)+m(\hat{\nu}_1+\nu_l)]}\right. \\
&\quad \left. + \sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2b_l^*e^{-j[n(\hat{\omega}_1-\omega_l)+m(\hat{\nu}_1-\nu_l)]}\right) \\
&\quad \times \sum_{i=1}^{k}S^{-1}T^{-1}\left(\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}a_ie^{j[n(\hat{\omega}_1+\omega_i)+m(\hat{\nu}_1+\nu_i)]}\right. \\
&\quad \left. + \sum_{n=0}^{S-1}\sum_{m=0}^{T-1}b_ie^{j[n(\hat{\omega}_1-\omega_i)+m(\hat{\nu}_1-\nu_i)]}\right) \\
&= -|a_1|^2\left(S^{-3}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2e^{j[n(\hat{\omega}_1-\omega_1)+m(\hat{\nu}_1-\nu_1)]}\right) \\
&\quad \times \left(S^{-1}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}e^{-j[n(\hat{\omega}_1-\omega_1)+m(\hat{\nu}_1-\nu_1)]}\right) \\
&\quad - |b_1|^2\left(S^{-3}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2e^{-j[n(\hat{\omega}_1-\omega_1)+m(\hat{\nu}_1-\nu_1)]}\right) \\
&\quad \times \left(S^{-1}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}e^{j[n(\hat{\omega}_1-\omega_1)+m(\hat{\nu}_1-\nu_1)]}\right) \\
&\quad + o(1) = -\frac{|a_1|^2+|b_1|^2}{3}(1+o(1)) \text{ a.s.}
\end{aligned} \tag{68}
$$

$$
\begin{aligned}
M_{21} &= S^{-4}T^{-2}\sum_{l=1}^{k}\left(a_l^*\mathbf{e}_l^H\hat{\ddot{\mathbf{e}}}_1 + b_l^*\mathbf{e}_l^T\hat{\ddot{\mathbf{e}}}_1\right)\hat{\mathbf{e}}_1^H\mathbf{u} \\
&\quad + S^{-4}T^{-2}\sum_{l=1}^{k}\left(a_l^*\mathbf{e}_l^H\hat{\ddot{\mathbf{e}}}_1^* + b_l^*\mathbf{e}_l^T\hat{\ddot{\mathbf{e}}}_1^*\right)\hat{\mathbf{e}}_1^T\mathbf{u} \\
&= -\sum_{l=1}^{k}S^{-3}T^{-1}\left(\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}a_l^*n^2e^{j[n(\hat{\omega}_1-\omega_l)+m(\hat{\nu}_1-\nu_l)]}\right. \\
&\quad \left. + \sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2b_l^*e^{j[n(\hat{\omega}_1+\omega_l)+m(\hat{\nu}_1+\nu_l)]}\right) \\
&\quad \times S^{-1}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}u(n,m)e^{-j[n\hat{\omega}_1+m\hat{\nu}_1]} \\
&\quad - \sum_{l=1}^{k}S^{-3}T^{-1}\left(\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}a_l^*n^2e^{-j[n(\hat{\omega}_1+\omega_l)+m(\hat{\nu}_1+\nu_l)]}\right. \\
&\quad \left. + \sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2b_l^*e^{-j[n(\hat{\omega}_1-\omega_l)+m(\hat{\nu}_1-\nu_l)]}\right) \\
&\quad \times S^{-1}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}u(n,m)e^{j[n\hat{\omega}_1+m\hat{\nu}_1]} \\
&= \left[-a_1^*\left(S^{-3}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2e^{j[n(\hat{\omega}_1-\omega_1)+m(\hat{\nu}_1-\nu_1)]}\right)\right.
\end{aligned}
$$

$$\times\left(S^{-1}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}u(n,m)e^{-j[n\hat{\omega}_1+m\hat{\nu}_1]}\right) - b_1^*\left(S^{-3}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2e^{-j[n(\hat{\omega}_1-\omega_1)+m(\hat{\nu}_1-\nu_1)]}\right) \times\left(S^{-1}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}u(n,m)e^{j[n\hat{\omega}_1+m\hat{\nu}_1]}\right)\Bigg] \times(1+o(1)) \text{ a.s.} \quad (69)$$

Using (44) and Lemma 3, we conclude that $M_{21}\to 0$ a.s. Similarly

$$\begin{aligned} M_{31} &= S^{-4}T^{-2}\mathbf{u}^T\left.\frac{\partial^2\mathbf{E}}{\partial\omega_i^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\hat{\mathbf{E}}^H\mathbf{u} \\ &= S^{-4}T^{-2}\mathbf{u}^T\hat{\ddot{\mathbf{e}}}_1\hat{\mathbf{e}}_1^H\mathbf{u}+S^{-4}T^{-2}\mathbf{u}^T\hat{\ddot{\mathbf{e}}}_1^*\hat{\mathbf{e}}_1^T\mathbf{u} \\ &= -\left(S^{-3}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2u(n,m)e^{j[n\hat{\omega}_1+m\hat{\nu}_1]}\right) \\ &\quad\times\left(S^{-1}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}u(n,m)e^{-j[n\hat{\omega}_1+m\hat{\nu}_1]}\right) \\ &\quad-\left(S^{-3}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2u(n,m)e^{-j[n\hat{\omega}_1+m\hat{\nu}_1]}\right) \\ &\quad\times\left(S^{-1}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}u(n,m)e^{j[n\hat{\omega}_1+m\hat{\nu}_1]}\right) \end{aligned} \quad (70)$$

and hence, $M_{31}\to 0$ a.s. By substituting (68)–(70) into (64), we have

$$H_1 = -\frac{|a_1|^2+|b_1|^2}{3}(1+o(1)) \text{ a.s.} \quad (71)$$

Next

$$\begin{aligned} H_2 &= S^{-3}T^{-1}\mathbf{y}^T\hat{\mathbf{E}}\left.\frac{\partial^2(\mathbf{E}^H\mathbf{E})^{-1}}{\partial\omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\hat{\mathbf{E}}^H\mathbf{y} \\ &= -S^{-3}T^{-1}\mathbf{y}^T\hat{\mathbf{E}}\frac{\partial}{\partial\omega_1} \\ &\quad\times\left.\left[(\mathbf{E}^H\mathbf{E})^{-1}\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}(\mathbf{E}^H\mathbf{E})^{-1}\right]\right|_{\mathbf{w}=\hat{\mathbf{w}}}\hat{\mathbf{E}}^H\mathbf{y} \\ &= -S^{-3}T^{-1}\left(\mathbf{y}^T\hat{\mathbf{E}}\left.\frac{\partial(\mathbf{E}^H\mathbf{E})^{-1}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\right. \\ &\quad\times\left.\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\hat{\mathbf{E}}^H\mathbf{y} \\ &\quad+\mathbf{y}^T\hat{\mathbf{E}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\left.\frac{\partial^2(\mathbf{E}^H\mathbf{E})}{\partial\omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\hat{\mathbf{E}}^H\mathbf{y} \\ &\quad+\mathbf{y}^T\hat{\mathbf{E}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\left.\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \\ &\quad\left.\times\left.\frac{\partial(\mathbf{E}^H\mathbf{E})^{-1}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\hat{\mathbf{E}}^H\mathbf{y}\right) \\ &= M_{12}+M_{22}+M_{12}^H. \end{aligned} \quad (72)$$

From (44) and Lemmas 2 and 3, we have

$$\begin{aligned} &S^{-1}T^{-1}\mathbf{y}^T\hat{\mathbf{e}}_l \\ &\quad= S^{-1}T^{-1}\tilde{\mathbf{a}}^H\mathbf{E}^H\hat{\mathbf{e}}_l+S^{-1}T^{-1}\mathbf{u}^T\hat{\mathbf{e}}_l \\ &\quad= S^{-1}T^{-1}\sum_{i=1}^{k}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left(a_i^*e^{j[n(\hat{\omega}_l-\omega_i)+m(\hat{\nu}_l-\nu_i)]}\right. \\ &\qquad\left.+b_i^*e^{j[n(\hat{\omega}_l+\omega_i)+m(\hat{\nu}_l+\nu_i)]}\right) \\ &\qquad+S^{-1}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}u(n,m)e^{j[n\hat{\omega}_l+m\hat{\nu}_l]} \\ &\quad= a_l^*(1+o(1)) \text{ a.s.} \end{aligned} \quad (73)$$

Similarly

$$\begin{aligned} S^{-1}T^{-1}\mathbf{y}^T\hat{\mathbf{e}}_l^* &= b_l^*(1+o(1)) \text{ a.s.} \\ S^{-2}T^{-1}\mathbf{y}^T\hat{\dot{\mathbf{e}}}_l &= \frac{a_l^*}{2}(1+o(1)) \text{ a.s.} \\ S^{-2}T^{-1}\mathbf{y}^T\hat{\dot{\mathbf{e}}}_l^* &= \frac{b_l^*}{2}(1+o(1)) \text{ a.s.} \\ S^{-3}T^{-1}\mathbf{y}^T\hat{\ddot{\mathbf{e}}}_l &= \frac{a_l^*}{3}(1+o(1)) \text{ a.s.} \\ S^{-3}T^{-1}\mathbf{y}^T\hat{\ddot{\mathbf{e}}}_l^* &= \frac{b_l^*}{3}(1+o(1)) \text{ a.s.} \end{aligned} \quad (74)$$

In addition, from (44) and Lemma 2

$$\begin{aligned} S^{-2}T^{-1}\hat{\dot{\mathbf{e}}}_i^T\hat{\mathbf{e}}_l^* &= S^{-2}T^{-1}\hat{\mathbf{e}}_l^H\hat{\dot{\mathbf{e}}}_i \\ &= S^{-2}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}ne^{j[n(\hat{\omega}_i-\hat{\omega}_l)+m(\hat{\nu}_i-\hat{\nu}_l)]} \\ &= o(1),\quad i\neq l \end{aligned} \quad (75)$$

and similarly

$$\begin{aligned} S^{-2}T^{-1}\hat{\dot{\mathbf{e}}}_i^H\hat{\mathbf{e}}_l^* &= S^{-2}T^{-1}\hat{\mathbf{e}}_l^H\hat{\dot{\mathbf{e}}}_i^* = o(1) \\ S^{-2}T^{-1}\hat{\dot{\mathbf{e}}}_i^H\hat{\mathbf{e}}_l &= S^{-2}T^{-1}\hat{\mathbf{e}}_l^T\hat{\dot{\mathbf{e}}}_i^* = o(1),\quad i\neq l \\ S^{-2}T^{-1}\hat{\dot{\mathbf{e}}}_i^T\hat{\mathbf{e}}_l &= S^{-2}T^{-1}\hat{\mathbf{e}}_l^T\hat{\dot{\mathbf{e}}}_i = o(1). \end{aligned} \quad (76)$$

Using its definition in (72)

$$\begin{aligned} M_{12} &= S^{-3}T^{-1}\mathbf{y}^T\hat{\mathbf{E}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\left.\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \\ &\quad(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\left.\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\hat{\mathbf{E}}^H\mathbf{y} \\ &= S^{-6}T^{-4}\mathbf{y}^T\hat{\mathbf{E}}\left.\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \\ &\quad\times\left.\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\hat{\mathbf{E}}^H\mathbf{y}(1+o(1)) \\ &= S^{-6}T^{-4}\left|\mathbf{y}^T\hat{\mathbf{E}}\left.\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\right|^2(1+o(1)) \\ &= \left(\left|2S^{-3}T^{-2}\mathbf{y}^T\hat{\dot{\mathbf{e}}}_1^*\hat{\dot{\mathbf{e}}}_1^T\hat{\mathbf{e}}_1\right.\right. \\ &\quad\left.+S^{-3}T^{-2}\sum_{l=2}^{k}\left(\mathbf{y}^T\hat{\mathbf{e}}_l\hat{\mathbf{e}}_l^H\hat{\dot{\mathbf{e}}}_1+\mathbf{y}^T\hat{\mathbf{e}}_l^*\hat{\mathbf{e}}_l^T\hat{\dot{\mathbf{e}}}_1\right)\right|^2 \\ &\quad+\left|2S^{-3}T^{-2}\mathbf{y}^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\hat{\dot{\mathbf{e}}}_1^*\right. \\ &\quad\left.+S^{-3}T^{-2}\sum_{l=2}^{k}\left(\mathbf{y}^T\hat{\mathbf{e}}_l\hat{\mathbf{e}}_l^H\hat{\dot{\mathbf{e}}}_1^*+\mathbf{y}^T\hat{\mathbf{e}}_l^*\hat{\mathbf{e}}_l^T\hat{\dot{\mathbf{e}}}_1^*\right)\right|^2 \end{aligned}$$

$$+\sum_{l=2}^{k}\left|S^{-3}T^{-2}\mathbf{y}^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\hat{\mathbf{e}}_l+S^{-3}T^{-2}\mathbf{y}^T\hat{\mathbf{e}}_1^*\hat{\mathbf{e}}_1^T\hat{\mathbf{e}}_l\right|^2 + \sum_{l=2}^{k}\left|S^{-3}T^{-2}\mathbf{y}^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\hat{\mathbf{e}}_l^*+S^{-3}T^{-2}\mathbf{y}^T\hat{\mathbf{e}}_1^*\hat{\mathbf{e}}_1^T\hat{\mathbf{e}}_l^*\right|^2\Bigg) \times(1+o(1)). \tag{77}$$

Since by (75) and (76) all squared terms on the of (77) tend to zero a.s. as $\Psi(S,T)\to\infty, M_{12}\to 0$ a.s.

Similarly to (76)

$$\begin{aligned}S^{-3}T^{-1}\hat{\mathbf{e}}_i^T\hat{\mathbf{e}}_l^* &= S^{-3}T^{-1}\hat{\mathbf{e}}_l^H\hat{\mathbf{e}}_i\\ &= S^{-3}T^{-1}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}n^2e^{l[n(\hat{\omega}_i-\hat{\omega}_l)+m(\hat{\nu}_i-\hat{\nu}_l)]}\\ &= o(1),\quad i\neq l\\ S^{-3}T^{-1}\hat{\mathbf{e}}_i^H\hat{\mathbf{e}}_l^* &= S^{-3}T^{-1}\hat{\mathbf{e}}_l^H\hat{\mathbf{e}}_i^* = o(1)\\ S^{-3}T^{-1}\hat{\mathbf{e}}_i^H\hat{\mathbf{e}}_l &= S^{-3}T^{-1}\hat{\mathbf{e}}_l^T\hat{\mathbf{e}}_i^* = o(1),\quad i\neq l\\ S^{-3}T^{-1}\hat{\mathbf{e}}_i^T\hat{\mathbf{e}}_l &= S^{-3}T^{-1}\hat{\mathbf{e}}_l^T\hat{\mathbf{e}}_i = o(1).\end{aligned} \tag{78}$$

Hence

$$\begin{aligned}M_{22} &= -S^{-3}T^{-1}\mathbf{y}^T\hat{\mathbf{E}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\left.\frac{\partial^2(\mathbf{E}^H\mathbf{E})}{\partial\omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \times(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\hat{\mathbf{E}}^H\mathbf{y}\\ &= -S^{-5}T^{-3}\mathbf{y}^T\hat{\mathbf{E}}\left.\frac{\partial^2(\mathbf{E}^H\mathbf{E})}{\partial\omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\hat{\mathbf{E}}^H\mathbf{y}(1+o(1))\\ &= -S^{-5}T^{-3}\Bigg[4\mathbf{y}^T\hat{\mathbf{e}}_1^*\hat{\mathbf{e}}_1^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\mathbf{y}\\ &\quad+\sum_{l=2}^{k}\left(\mathbf{y}^T\hat{\mathbf{e}}_l\hat{\mathbf{e}}_l^H\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\mathbf{y}+\mathbf{y}^T\hat{\mathbf{e}}_l^*\hat{\mathbf{e}}_l^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\mathbf{y}\right)\\ &\quad+4\mathbf{y}^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\hat{\mathbf{e}}_1^*\hat{\mathbf{e}}_1^T\mathbf{y}+\sum_{l=2}^{k}\Big(\mathbf{y}^T\hat{\mathbf{e}}_l\hat{\mathbf{e}}_l^H\hat{\mathbf{e}}_1^*\hat{\mathbf{e}}_1^T\mathbf{y}\\ &\quad+\mathbf{y}^T\hat{\mathbf{e}}_l^*\hat{\mathbf{e}}_l^T\hat{\mathbf{e}}_1^*\hat{\mathbf{e}}_1^T\mathbf{y}\Big)+\sum_{l=2}^{k}\Big(\mathbf{y}^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\hat{\mathbf{e}}_l\hat{\mathbf{e}}_l^H\mathbf{y}\\ &\quad+\mathbf{y}^T\hat{\mathbf{e}}_1^*\hat{\mathbf{e}}_1^T\hat{\mathbf{e}}_l\hat{\mathbf{e}}_l^H\mathbf{y}+\mathbf{y}^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\hat{\mathbf{e}}_l^*\hat{\mathbf{e}}_l^T\mathbf{y}\\ &\quad+\mathbf{y}^T\hat{\mathbf{e}}_1^*\hat{\mathbf{e}}_1^T\hat{\mathbf{e}}_l^*\hat{\mathbf{e}}_l^T\mathbf{y}\Big)\Bigg](1+o(1)).\end{aligned} \tag{79}$$

Since, by (78), all squared terms on the right-hand side of (79) tend to zero a.s. as $\Psi(S,T)\to\infty, M_{22}\to 0$ a.s. Hence, as $\Psi(S,T)\to\infty, H_2\to 0$ a.s.

Similarly, using (73)–(76), we have

$$\begin{aligned}H_3 &= 2S^{-3}T^{-1}\mathbf{y}^T\left.\frac{\partial\mathbf{E}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1} \times\left.\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\hat{\mathbf{E}}^H\mathbf{y}\\ &= 2S^{-5}T^{-3}\mathbf{y}^T\left.\frac{\partial\mathbf{E}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}} \times\left.\frac{\partial(\mathbf{E}^H\mathbf{E})}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\hat{\mathbf{E}}^H\mathbf{y}(1+o(1))\\ &= 2S^{-5}T^{-3}\Big[2\mathbf{y}^T\hat{\dot{\mathbf{e}}}_1^*\hat{\mathbf{e}}_1^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\mathbf{y}+2\mathbf{y}^T\hat{\mathbf{e}}_1\hat{\mathbf{e}}_1^H\hat{\dot{\mathbf{e}}}_1^*\hat{\mathbf{e}}_1^T\mathbf{y}\\ &\quad+\sum_{l=2}^{k}\Big(\mathbf{y}^T\hat{\mathbf{e}}_1\hat{\dot{\mathbf{e}}}_1^H\hat{\mathbf{e}}_l\hat{\mathbf{e}}_l^H\mathbf{y}+\mathbf{y}^T\hat{\dot{\mathbf{e}}}_1^*\hat{\dot{\mathbf{e}}}_1^T\hat{\mathbf{e}}_l\hat{\mathbf{e}}_l^H\mathbf{y}\\ &\quad+\mathbf{y}^T\hat{\mathbf{e}}_1\hat{\dot{\mathbf{e}}}_1^H\hat{\mathbf{e}}_l^*\hat{\mathbf{e}}_l^T\mathbf{y}+\mathbf{y}^T\hat{\dot{\mathbf{e}}}_1^*\hat{\dot{\mathbf{e}}}_1^T\hat{\mathbf{e}}_l^*\hat{\mathbf{e}}_l^T\mathbf{y}\Big)\Big](1+o(1)).\end{aligned} \tag{80}$$

As $\Psi(S,T)\to\infty, H_3\to 0$ a.s.

In addition

$$\begin{aligned}H_4 &= 2S^{-3}T^{-1}\mathbf{y}^T\left.\frac{\partial\mathbf{E}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}(\hat{\mathbf{E}}^H\hat{\mathbf{E}})^{-1}\left.\frac{\partial\mathbf{E}^H}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\mathbf{y}\\ &= 2S^{-4}T^{-2}\left|\mathbf{y}^T\left.\frac{\partial\mathbf{E}}{\partial\omega_1}\right|_{\mathbf{w}=\hat{\mathbf{w}}}\right|^2(1+o(1))\\ &= 2(|S^{-2}T^{-1}\mathbf{y}^T\hat{\mathbf{e}}_1|^2+|S^{-2}T^{-1}\mathbf{y}^T\hat{\mathbf{e}}_1^*|^2)(1+o(1))\\ &= \frac{|a_1|^2+|b_1|^2}{2}(1+o(1))\text{ a.s.}\end{aligned} \tag{81}$$

Let $\mathcal{C}$ denote again some "generic" constant whose exact value is of no importance to us in the current context. Substituting the above results into (63), we conclude that

$$S^{-3}T^{-1}\left.\frac{\partial^2\mathbf{y}^T\mathbf{P}^\perp\mathbf{y}}{\partial\omega_1^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} = \mathcal{C}(1+o(1))\text{ a.s.} \tag{82}$$

Deriving similar asymptotic expressions for all the other normalized elements of $\hat{\mathcal{H}}_{\mathrm{ML}}$, we have, as $\Psi(S,T)\to\infty$

$$\begin{aligned}S^{-3}T^{-1}\left.\frac{\partial^2\mathbf{y}^T\mathbf{P}^\perp\mathbf{y}}{\partial\omega_i^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} &= \mathcal{C}(1+o(1))\text{ a.s.}\\ S^{-1}T^{-3}\left.\frac{\partial^2\mathbf{y}^T\mathbf{P}^\perp\mathbf{y}}{\partial\nu_i^2}\right|_{\mathbf{w}=\hat{\mathbf{w}}} &= \mathcal{C}(1+o(1))\text{ a.s.}\\ S^{-3}T^{-1}\left.\frac{\partial^2\mathbf{y}^T\mathbf{P}^\perp\mathbf{y}}{\partial\omega_i\partial\omega_l}\right|_{\mathbf{w}=\hat{\mathbf{w}}} &\to 0\text{ a.s.}\\ S^{-1}T^{-3}\left.\frac{\partial^2\mathbf{y}^T\mathbf{P}^\perp\mathbf{y}}{\partial\nu_i\partial\nu_l}\right|_{\mathbf{w}=\hat{\mathbf{w}}} &\to 0\text{ a.s.}\\ S^{-2}T^{-2}\left.\frac{\partial^2\mathbf{y}^T\mathbf{P}^\perp\mathbf{y}}{\partial\omega_i\partial\nu_l}\right|_{\mathbf{w}=\hat{\mathbf{w}}} &\to 0\text{ a.s.}\\ S^{-2}T^{-2}\left.\frac{\partial^2\mathbf{y}^T\mathbf{P}^\perp\mathbf{y}}{\partial\nu_i\partial\omega_l}\right|_{\mathbf{w}=\hat{\mathbf{w}}} &\to 0\text{ a.s.}\\ S^{-2}T^{-2}\left.\frac{\partial^2\mathbf{y}^T\mathbf{P}^\perp\mathbf{y}}{\partial\omega_i\partial\nu_i}\right|_{\mathbf{w}=\hat{\mathbf{w}}} &= \mathcal{C}(1+o(1))\text{ a.s.}\end{aligned} \tag{83}$$

Finally, let $\mathbf{1_k}$ denote a k-dimensional vector, such that all its entries are 1, and let us define the diagonal matrices $\boldsymbol{\mathcal{D}}_1 = \operatorname{diag}\{S^{-2}\mathbf{1_k}, S^{-1}T^{-1}\mathbf{1_k}\}$, and $\boldsymbol{\mathcal{D}}_2 = \operatorname{diag}\{S^{-1}T^{-1}\mathbf{1_k}, T^{-2}\mathbf{1_k}\}$. Since $\hat{\mathcal{H}}_{\mathrm{ML}}$ can be rewritten as

$$\hat{\mathcal{H}}_{\mathrm{ML}} = \boldsymbol{\mathcal{D}}_1^{-1}\boldsymbol{\mathcal{D}}_1\hat{\mathcal{H}}_{\mathrm{ML}}\boldsymbol{\mathcal{D}}_2\boldsymbol{\mathcal{D}}_2^{-1} \tag{84}$$

we conclude by substituting (83) into (84) that

$$|\hat{\mathcal{H}}_{\mathrm{ML}}| = S^{4k}T^{4k}\mathcal{C}(1+o(1)) \tag{85}$$

Hence

$$|\hat{\mathbf{H}}_{\mathrm{ML}}| = \mathcal{C}(\mathbf{y}^T\hat{\mathbf{P}}^\perp\mathbf{y})^{-2k}S^{4k}T^{4k}(1+o(1))\text{ a.s.} \tag{86}$$

■

REFERENCES

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. AC-19, no. 6, pp. 716–723, Dec. 1974.

[2] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. New York: Wiley, 1992.

[3] T. J. I'A. Bromwich, *An Introduction to the Theory of Infinite Series*, Second ed. London, U.K.: MacMillan and Co. LTD, 1965.

[4] N. G. De Bruijn, *Asymptotic Methods in Analysis*, Third ed. Amsterdam, The Netherlands: North-Holland, 1970.

[5] P. M. Djuric, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Signal Process.*, vol. 44, no. 7, pp. 1744–1751, Jul. 1996.

[6] J. M. Francos, A. Z. Meiri, and B. Porat, "A Wold-like decomposition of 2-D discrete homogeneous random fields," *Ann. Appl. Prob.*, vol. 5, pp. 248–260, 1995.

[7] J. M. Francos, A. Narashimhan, and J. W. Woods, "Maximum-likelihood parameter estimation of textures using a Wold-decomposition based model," *IEEE Trans. Image Process.*, vol. 4, no. 12, pp. 1655–1666, Dec. 1995.

[8] J.-J. Fuchs, "Estimating the number of sinusoids in additive white noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 12, pp. 1846–1853, Dec. 1988.

[9] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*. New York: Academic, 1980.

[10] E. J. Hannan, "The estimation of the order an ARMA process," *Ann. Statist*, vol. 8, pp. 1071–1081, 1980.

[11] ——, "Estimating the dimension of a linear system," *J. Multivariate Anal.*, vol. 11, pp. 459–472, 1981.

[12] S. He, "Parameter estimation of hidden periodic model in random fields," *Sci. China Ser. A*, vol. 42, pp. 238–245, 1998.

[13] L. Kavalieris and E. J. Hannan, "Determining the number of terms in a trigonometric regression," *J. Time Series Anal.*, vol. 15, pp. 613–625, 1994.

[14] S. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJL: Prentice-Hall, 1988.

[15] M. A. Kliger and J. M. Francos, "Strong consistency of a family of model order selection rules for estimating the parameters of 2-D sinusoids in white noise," , submitted for publication.

[16] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with application to target feature extraction," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 281–295, Feb. 1996.

[17] B. G. Quinn, "Estimating the number of terms in a sinusoidal regression," *J. Time Series Anal.*, vol. 10, pp. 71–75, 1989.

[18] C. R. Rao, L. C. Zhao, and B. Zhou, "Maximum likelihood estimation of 2-D superimposed exponential signals," *IEEE Trans. Signal Process.*, vol. 42, no. 7, pp. 1795–1802, Jul. 1994.

[19] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, pp. 468–478, 1978.

[20] ——, *Stochastic Complexity in Statistical Inquiry*, Singapore: World Scientific, 1989.

[21] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, pp. 461–464, 1978.

[22] P. Stoica, P. Eykhoff, P. Janssen, and T. Soderstrom, "Model-structure selection by cross-validation," *Int. J. Control*, vol. 43, pp. 1841–1878, 1986.

[23] R. Stoica, J. Zerubia, and J. M. Francos, "The two-dimensional wold decomposition for segmentation and indexing in image libraries," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Seattle, WA, 1998.

[24] V. U. Reddy and L. S. Biradar, "SVD based information theoretic criteria for detection of the number of damped/undamped sinusoids and their performance analysis," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2872–2881, Sep. 1993.

[25] M. Wax and T. Kailath, "Detection of signal by information theoretic criteria," in *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, Apr. 1985, pp. 397–392.

[26] D. V. Widder, *The Laplace Transform*. Princeton, NJ: Princeton Univ. Press, 1972.

[27] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in presence of white noise," *J. Multivariate Anal.*, vol. 20, pp. 1–25, 1986.

Mark Kliger (S'04) was born in Kishinev, Republic of Moldova, in 1973. He received the B.Sc. degree in electrical and computer engineering in 1996 and the M.Sc. degree (*magna cum laude*) in electrical and computer engineering in 2001, both from the Ben-Gurion University, Beer-Sheva, Israel, where he is currently pursuing the Ph.D. degree.

From 1996 to 1999, he was with the Operations Research Laboratory, Israeli Defense Forces. His current research interests are in parametric modeling and estimation of 2-D random fields, random fields theory, image and video modeling and indexing, and texture analysis and synthesis.

Mr. Kliger is a recipient of the Israeli Council for Higher Education scholarship for outstanding Ph.D. students (2001 to 2005).

Joseph M. Francos (SM'97) was born on November 6, 1959, in Tel-Aviv, Israel. He received the B.Sc. degree in computer engineering in 1982 and the D.Sc. degree in electrical engineering in 1991, both from the Technion—Israel Institute of Technology, Haifa.

From 1982 to 1987, he was with the Signal Corps Research Laboratories, Israeli Defense Forces. From 1991 to 1992, he was with the Department of Electrical Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, as a Visiting Assistant Professor. During 1993, he was with Signal Processing Technology, Palo Alto, CA. In 1993, he joined the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva, Israel, where he is now an Associate Professor. He also held visiting positions at the Massachusetts Institute of Technology Media Laboratory, Cambridge; at the Electrical and Computer Engineering Department, University of California, Davis; at the Electrical Engineering and Computer Science Department, University of Illinois, Chicago; and at the Electrical Engineering Department, University of California, Santa Cruz. His current research interests are in parametric modeling and estimation of 2-D random fields, random fields theory, parametric modeling and estimation of nonstationary signals, space-time coding, image modeling and indexing, and texture analysis and synthesis.

Dr. Francos served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2001.