Parameter Estimation of Two-Dimensional Moving Average Random Fields

Joseph M. Francos, *Senior Member, IEEE*, and Benjamin Friedlander, *Fellow, IEEE*

***Abstract*—This paper considers the problem of estimating the parameters of two-dimensional (2-D) moving average random (MA) fields. We first address the problem of expressing the covariance matrix of nonsymmetrical half-plane, noncausal, and quarter-plane MA random fields in terms of the model parameters. Assuming the random field is Gaussian, we derive a closed-form expression for the Cramér–Rao lower bound (CRLB) on the error variance in jointly estimating the model parameters. A computationally efficient algorithm for estimating the parameters of the MA model is developed. The algorithm initially fits a 2-D autoregressive model to the observed field and then uses the estimated parameters to compute the MA model. A maximum-likelihood algorithm for estimating the MA model parameters is also presented. The performance of the proposed algorithms is illustrated by Monte-Carlo simulations and is compared with the Cramér–Rao bound.**

***Index Terms*— Maximum likelihood, moving average random fields, parameter estimation, random fields.**

I. Introduction

THE PROBLEM of estimating the parameters of a two-dimensional (2-D) real-valued discrete and homogeneous moving average (MA) random field from a single observed realization of it is of great theoretical and practical importance. For example, it arises in the problem of estimating the parameters of the purely indeterministic component of natural textures in images [11] as well as in image segmentation and restoration problems (e.g., [12]).

More specifically, in [11], we presented a texture model based on the 2-D Wold-type decomposition of homogeneous random fields [2]. In this framework, the texture field is assumed to be a realization of a regular homogeneous random field, which can have a mixed spectral distribution. The texture is represented as a sum of purely indeterministic, harmonic, and a countable number of evanescent fields. The harmonic and evanescent components of the field result in the structural attributes of the observed realization, whereas the purely indeterministic component is the structureless, "random looking" component of the texture field.

It is shown in [3] (see also [2]) that any 2-D purely indeterministic random field has a unique white innovations-driven nonsymmetrical half plane (NSHP) MA representation. In general, the MA model support has infinite dimensions. However, in the texture modeling problem, we have found (see, e.g., [10] and [11]) that in many cases, the spectral density function of the purely indeterministic component is smooth and has small dynamic range (i.e., no sharp peaks). This property suggests that the purely indeterministic component can be well modeled using a *finite support* NSHP MA model. On the other hand, modeling the purely indeterministic component by a 2-D autoregressive (AR) model may require larger supports, and hence, a less compact parameterization of the field is obtained. Thus, in those cases where the covariance function of the purely indeterministic component of the field rapidly decays to zero, a finite support NSHP MA model would generally provide a more compact representation of the purely indeterministic field. We note that many of the existing texture analysis and synthesis algorithms employ 2-D AR models for texture modeling (see, e.g., [7]–[9]). These AR models produce efficient parameterization of the purely indeterministic field when its spectral density function contains high peaks and has large dynamic range.

The general problem of estimating the parameters of random fields has received considerable attention. Most approaches for estimating the parameters of purely indeterministic random fields concentrate on fitting 2-D AR models to the observed field. In general, three types of AR models that differ in the model support definition are in use. These are the noncausal (NC), nonsymmetrical half-plane (NSHP), and quarter-plane (QP) AR models. Least squares solution of the set of 2-D normal equations that corresponds to each of the different models is a method widely used in various image processing applications like image restoration and segmentation. A Levinson-type algorithm for solving the set of 2-D normal equations of a continuous support NSHP AR model is derived in [5]. A recent analysis of the problem of estimating the parameters of 2-D noncausal Gauss–Markov random fields can be found in [6]. The asymptotic Cramér–Rao bound for the parameters of a Gaussian purely indeterministic field was derived by Whittle [4]. More recently, this general derivation was specialized for the case of noncausal AR models and NSHP AR models in [13].

In this paper, we concentrate on finding estimation algorithms for 2-D MA random fields and on establishing bounds on the achievable estimation accuracy of the MA model parameters, given a finite dimensional observed realization. We propose a computationally efficient algorithm for estimating the parameters of MA random fields using a *finite dimension,*

Manuscript received January 17, 1997; revised January 17, 1998. This work was supported by the Office of Naval Research under Contract N00014-95-1-0912. The associate editor coordinating the review of this paper and approving it for publication was Prof. Luis F. Chaparro.

J. M. Francos is with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva, Israel.

B. Friedlander is with the Department of Electrical and Computer Engineering, University of California, Davis, CA 95616 USA.

Publisher Item Identifier S 1053-587X(98)05225-8.

single observed realization of this field. The algorithm is an extension to two-dimensions of Durbin's "MA by AR" method [1] for estimating the parameters of scalar MA processes. The algorithm has two stages. In the first stage, a 2-D NSHP AR model is fit to the observed field using a least squares solution of the 2-D normal equations or alternatively by using a finite support version [10] of Marzetta's [5] Levinson-type algorithm. In the second stage, the estimated parameters of the AR model are used to compute the parameters of the MA model through a least squares solution of a system of linear equations. The overall algorithm is computationally efficient.

We also address here the problem of expressing the covariance matrix of the observed field in terms of the MA model parameters. Then, assuming the MA field is Gaussian, we employ this result to establish bounds on the achievable accuracy in jointly estimating the parameters of the MA modeled purely indeterministic random field. We derive closed form exact expression for the Cramér-Rao lower bound (CRLB) on the achievable estimation accuracy. Using the expressions of the covariance matrix in terms of the MA model parameters, we then derive a maximum likelihood algorithm for these. The previously derived "MA by AR" algorithm is used for initialization of the multidimensional search involved in the maximum likelihood estimation (MLE) algorithm. Since the MLE method requires an iterative and computationally intensive procedure, it becomes computationally prohibitive even for moderate size data fields. However, as we show in this paper, as the data size increases, the "MA by AR" algorithm becomes less biased and, therefore, offers an increasingly attractive alternative to ML estimation.

In [14], we consider the general problem of establishing bounds on the achievable accuracy in jointly estimating the parameters of a real valued, 2-D, homogeneous random field with mixed spectral distribution from a single observed realization of it. However, in [14], we restricted our attention to the case in which the purely indeterministic component of the random field is a white noise field. Thus, the derivation presented here provides a generalization of the lower bound derived in [14] for the case of an arbitrary purely indeterministic component.

The paper is organized as follows. In Section II, we consider the problem of representing the covariance matrix of the observed MA field in terms of the MA model parameters. The result is derived first for an NSHP MA model and then extended to the cases of noncausal MA model and quarter-plane model. In Section III, a closed-form expression for the CRB on the error variance in jointly estimating the MA model parameters is derived. In Section IV, we develop the computationally efficient "MA by AR" estimation algorithm. Section V presents the ML algorithm for estimating the MA model parameters. In Section VI, we present some numerical examples. The performance of the proposed algorithms is illustrated by Monte Carlo simulations and is compared with the Cramér–Rao bound. We investigate the effects of both the data size and the dimensions of the support of the approximating AR model on the bias and error variance of the proposed algorithms. In Section VII, we present some concluding remarks.

II. The Parametric Representation of the MA Field and Its Covariance Matrix

Let $\{y(n,m),(n,m) \in \mathcal{Z}^2\}$, be a real-valued, regular, and homogeneous random field, such that its spectral distribution function is absolutely continuous. Let us define a total order on the discrete lattice such that

$$(i,j) \preceq (s,t) \text{ iff } (i,j) \in \{\{(k,\ell)|k=s, \ell \leq t\} \cup \{(k,\ell)|k<s, -\infty < \ell < \infty\}\}. \quad (1)$$

Then, $y(n,m)$ can be uniquely represented by

$$y(n,m) = \sum_{(0,0)\preceq(k,\ell)} b(k,\ell)u(n-k,m-\ell) \quad (2)$$

where $\Sigma_{(0,0)\preceq(k,\ell)}\, b^2(k,\ell) < \infty; b(0,0)=1$, and $\{u(n,m)\}$ is the innovations field of $\{y(n,m)\}$ with respect to the total-order definition (1), [3]. $\{u(n,m)\}$ is a white noise field. The random field $\{y(n,m)\}$ is called *purely indeterministic* random field. We therefore conclude that the most general model of any regular random field, whose spectral distribution function is absolutely continuous, is the innovations-driven, NSHP support MA model (2).

In practice, the observed random field is of finite dimensions. Hence, let $\{y(n,m),(n,m) \in D\}$, where $D = \{(i,j)|0 \leq i \leq S-1, 0 \leq j \leq T-1\}$, be the observed random field. The MA model (2) is, in general, of infinite dimensions. In this paper, we restrict our attention to MA models of finite-dimensional NSHP support. Next, we elaborate on expressing the covariance matrix of the observed 2-D MA random field in terms of the model parameters.

Assumption 1: The purely indeterministic field is a real-valued MA field, whose model is given by (2) with $(k,\ell) \in S_{N,M}$, where $S_{N,M} = \{(i,j)|i=0, 0 \leq j \leq M\} \cup \{(i,j)|1 \leq i \leq N, -M \leq j \leq M\}$, and N, M are *a priori* known. The driving noise of the MA model is a zero mean, real-valued white noise field with variance σ^2. Thus, (2) is replaced by

$$y(n,m) = \sum_{(k,\ell)\in S_{N,M}} b(k,\ell)u(n-k,m-\ell). \quad (3)$$

The parameter vector of the observed field $\{y(n,m)\}$ is given by

$$\boldsymbol{\theta} = [\sigma^2, b(0,1), \cdots, b(0,M), b(1,-M), \cdots, b(1,M) \cdots, b(N,-M), \cdots, b(N,M)]^T. \quad (4)$$

Let us stack the columns of the observed field into the vector form

$$\boldsymbol{y} = [y(S-1,T-1), \cdots, y(S-1,0), \cdots, y(1,T-1) \cdots, y(1,0), y(0,T-1), \cdots, y(0,0)]^T. \quad (5)$$

Similarly, let the driving noise vector be defined by

$$\boldsymbol{u} = [u(S-1,T+M-1), \cdots, u(S-1,-M) \cdots, u(0,T+M-1), \cdots, u(0,-M) \cdots, u(-N,T+M-1), \cdots, u(-N,-M)]^T. \quad (6)$$

Let $\mathbf{0}_k$ denote a k-dimensional row vector of zeros. In addition, let

$$\begin{aligned}\boldsymbol{b}_0 &= [\mathbf{0}_M, 1, b(0,1), \cdots, b(0,M), \mathbf{0}_{T-1}]\\ \boldsymbol{b}_1 &= [b(1,-M), \cdots, b(1,0), \cdots, b(1,M), \mathbf{0}_{T-1}]\\ &\vdots \qquad\qquad \vdots\\ \boldsymbol{b}_N &= [b(N,-M), \cdots, b(N,0), \cdots, b(N,M)]\end{aligned} \tag{7}$$

and

$$\boldsymbol{b} = [\boldsymbol{b}_0, \boldsymbol{b}_1, \cdots, \boldsymbol{b}_N]. \tag{8}$$

Note that $\boldsymbol{b}$ is a $(T+2M)\cdot(N+1)-(T-1)$-dimensional row vector.

Define the $T \times (T+2M)\cdot(N+1)$ banded Toeplitz matrix

$$[\overline{\boldsymbol{B}}]_{i,j} = \begin{cases}\boldsymbol{b}(j-i+1), & j \geq i\\ 0, & i < j\end{cases} \tag{9}$$

where $\boldsymbol{b}(i) = 0$ for $i < 0$, and $i > (T+2M)\cdot(N+1)-(T-1)$. Finally, we define the $ST \times (T+2M)(S+N)$ block matrix in (10), shown at the bottom of the page.

Thus, we can rewrite the observations (3) in the form $\boldsymbol{y} = \boldsymbol{Bu}$. The covariance matrix of the observed field is given in terms of the MA model parameters by

$$\boldsymbol{\Gamma} = \sigma^2 \boldsymbol{B}\boldsymbol{B}^T. \tag{11}$$

Note that (10) and (11) are made valid for any type of support of the MA model simply by redefining $\boldsymbol{b}, \boldsymbol{u}$, and $\boldsymbol{B}$. In the following, we list two important examples, namely, the noncausal and quarter-plane models.

The Noncausal MA Model: Consider the 2-D MA model of (3), where we redefine its support so that $S_{N,M} = \{(i,j)| -N \leq i \leq N, -M \leq j \leq M\}$, and $b(0,0) = 1$. Let

$$\begin{aligned}\boldsymbol{b}_{-N} &= [b(-N,-M), \cdots, b(-N,0), \cdots, b(-N,M), \mathbf{0}_{T-1}]\\ &\vdots \qquad\qquad \vdots\\ \boldsymbol{b}_0 &= [b(0,-M), \cdots, b(0,0), \cdots, b(0,M), \mathbf{0}_{T-1}]\\ &\vdots \qquad\qquad \vdots\\ \boldsymbol{b}_N &= [b(N,-M), \cdots, b(N,0), \cdots, b(N,M)]\end{aligned} \tag{12}$$

and

$$\boldsymbol{b} = [\boldsymbol{b}_{-N}, \cdots, \boldsymbol{b}_0, \cdots, \boldsymbol{b}_N]. \tag{13}$$

Here, $\boldsymbol{b}$ is a $(T+2M)\cdot(2N+1)-(T-1)$-dimensional row vector. Define $\overline{\boldsymbol{C}}$ to be the $T \times (T+2M)\cdot(2N+1)$ banded Toeplitz matrix

$$[\overline{\boldsymbol{C}}]_{i,j} = \begin{cases}\boldsymbol{b}(j-i+1), & j \geq i\\ 0, & i < j\end{cases} \tag{14}$$

where $\boldsymbol{b}(i) = 0$ for $i < 0$, and $i > (T+2M)\cdot(2N+1)-(T-1)$. Finally, $\boldsymbol{B}$ is an $ST \times (T+2M)(S+2N)$ block matrix defined similarly to (10) with $\overline{\boldsymbol{B}}$ replaced by $\overline{\boldsymbol{C}}$.

The Quarter Plane MA Model: Consider the 2-D MA model of (3), where we redefine its support so that $S_{N,M} = \{(i,j)|0 \leq i \leq N, 0 \leq j \leq M\}$, and $b(0,0) = 1$. Let

$$\begin{aligned}\boldsymbol{b}_0 &= [b(0,0), \cdots, b(0,M), \mathbf{0}_{T-1}]\\ &\vdots \qquad\qquad \vdots\\ \boldsymbol{b}_N &= [b(N,0), \cdots, b(N,M)]\end{aligned} \tag{15}$$

and

$$\boldsymbol{b} = [\boldsymbol{b}_0, \cdots, \boldsymbol{b}_N]. \tag{16}$$

In the case of QP support, $\boldsymbol{b}$ is a $(T+M)\cdot(N+1)-(T-1)$-dimensional row vector. Define $\overline{\boldsymbol{D}}$ to be the $T \times (T+M)\cdot(N+1)$ banded Toeplitz matrix

$$[\overline{\boldsymbol{D}}]_{i,j} = \begin{cases}\boldsymbol{b}(j-i+1), & j \geq i\\ 0, & i < j\end{cases} \tag{17}$$

where $\boldsymbol{b}(i) = 0$ for $i < 0$, and $i > (T+M)\cdot(N+1)-(T-1)$. Finally, $\boldsymbol{B}$ is the $ST \times (T+M)(S+N)$ block matrix in (18), shown at the bottom of the page.

III. The Cramér–Rao Bound on the MA Model Parameters

Assume that the driving noise of the NSHP MA model is a zero mean, real-valued Gaussian white noise field with variance σ^2. Hence, the observed field $\{y(n,m)\}$ is also Gaussian. The general expression for the Fisher information

$$\boldsymbol{B} = \begin{bmatrix}\overline{\boldsymbol{B}} & \mathbf{0}_{T\times(T+2M)} & \cdots & & \mathbf{0}_{T\times(T+2M)}\\ \mathbf{0}_{T\times(T+2M)} & \overline{\boldsymbol{B}} & \mathbf{0}_{T\times(T+2M)} & \cdots & \mathbf{0}_{T\times(T+2M)}\\ & & \ddots & & \\ \mathbf{0}_{T\times(T+2M)} & \cdots & & \overline{\boldsymbol{B}} & \mathbf{0}_{T\times(T+2M)}\\ \mathbf{0}_{T\times(T+2M)} & \cdots & & \mathbf{0}_{T\times(T+2M)} & \overline{\boldsymbol{B}}\end{bmatrix}. \tag{10}$$

$$\boldsymbol{B} = \begin{bmatrix}\overline{\boldsymbol{D}} & \mathbf{0}_{T\times(T+M)} & \cdots & & \mathbf{0}_{T\times(T+M)}\\ \mathbf{0}_{T\times(T+M)} & \overline{\boldsymbol{D}} & \mathbf{0}_{T\times(T+M)} & \cdots & \mathbf{0}_{T\times(T+M)}\\ & & \ddots & & \\ \mathbf{0}_{T\times(T+M)} & \cdots & & \overline{\boldsymbol{D}} & \mathbf{0}_{T\times(T+M)}\\ \mathbf{0}_{T\times(T+M)} & \cdots & & \mathbf{0}_{T\times(T+M)} & \overline{\boldsymbol{D}}\end{bmatrix} \tag{18}$$

matrix (FIM) of a real, zero mean, Gaussian process is given by (e.g., [16])

$$[\boldsymbol{J}(\boldsymbol{\theta})]_{i,j} = \frac{1}{2}\,\mathrm{tr}\left\{\boldsymbol{\Gamma}^{-1}\,\frac{\partial \boldsymbol{\Gamma}}{\partial \theta_i}\,\boldsymbol{\Gamma}^{-1}\,\frac{\partial \boldsymbol{\Gamma}}{\partial \theta_j}\right\} \tag{19}$$

where $\boldsymbol{\Gamma}$ is the observation vector covariance matrix, and $[\boldsymbol{J}(\boldsymbol{\theta})]_{i,j}$ denotes the (i,j) entry of the matrix $\boldsymbol{J}$.

Note that

$$\frac{\partial \boldsymbol{b}}{\partial b(k,\ell)} = \boldsymbol{e}_{k(T+2M)+M+1+\ell} \tag{20}$$

where $\boldsymbol{e}_{k(T+2M)+M+1+\ell}$ is a $(T+2M)\cdot(N+1)-(T-1)$-dimensional row vector whose $k(T+2M)+M+1+\ell$ element equals one, whereas all its other elements are zero. Hence

$$\frac{\partial \overline{\boldsymbol{B}}}{\partial b(k,\ell)} = \overline{\boldsymbol{U}}_{(k,\ell)} \tag{21}$$

where $\overline{\boldsymbol{U}}_{(k,\ell)}$ is the up-shift matrix

$$[\overline{\boldsymbol{U}}_{(k,\ell)}]_{i,j} = \begin{cases} 1, & j-i = k(T+2M)+M+\ell \\ 0, & \text{otherwise.} \end{cases} \tag{22}$$

Taking the partial derivatives of $\boldsymbol{\Gamma}$ with respect to the MA model parameters, we have

$$\frac{\partial \boldsymbol{\Gamma}}{\partial b(k,\ell)} = \sigma^2\,[\boldsymbol{U}_{(k,\ell)}\boldsymbol{B}^T + \boldsymbol{B}\boldsymbol{U}_{(k,\ell)}^T], \quad (k,\ell)\in S_{N,M}\setminus\{(0,0)\} \tag{23}$$

where $\boldsymbol{U}_{(k,\ell)}$ is defined in (24), shown at the bottom of the page. In addition

$$\frac{\partial \boldsymbol{\Gamma}}{\partial \sigma^2} = \frac{1}{\sigma^2}\,\boldsymbol{\Gamma}. \tag{25}$$

Substituting (11), (23), and (25) into (19), we obtain a closed-form expression for the FIM of 2-D Gaussian MA random fields.

In many cases, we are interested not only in estimating the MA model parameters but in estimating some function of these parameters, such as the spectral density function of the observed field. Having estimated the model parameters $\boldsymbol{\theta}$, the spectral density of the field can be computed using its known functional dependence on the (estimated) parameters

$$S(e^{j2\pi\omega}, e^{j2\pi\nu}) = \sigma^2 |\mathcal{B}(e^{j2\pi\omega}, e^{j2\pi\nu})|^2 \tag{26}$$

where

$$\mathcal{B}(e^{j2\pi\omega}, e^{j2\pi\nu}) = \sum_{(k,\ell)\in S_{N,M}} b(k,\ell)e^{-j2\pi(k\omega+\ell\nu)}. \tag{27}$$

Next, we derive the CRB on the spectral density of the field. The spectral density function of the MA field is a differentiable function of the parameter vector $\boldsymbol{\theta}$. Hence, e.g., [15], the CRB on $S(e^{j2\pi\omega}, e^{j2\pi\nu})$ is related to the CRB of $\boldsymbol{\theta}$ by

$$\mathrm{CRB}(S(e^{j2\pi\omega}, e^{j2\pi\nu})) = \mathbf{W}^T\mathrm{CRB}(\boldsymbol{\theta})\boldsymbol{W} \tag{28}$$

where the column vector $\boldsymbol{W}$ is defined by

$$\boldsymbol{W} = \frac{\partial S(e^{j2\pi\omega}, e^{j2\pi\nu})}{\partial \boldsymbol{\theta}}. \tag{29}$$

Taking the partial derivatives of $S(e^{j2\pi\omega}, e^{j2\pi\nu})$ with respect to the parameters in $\boldsymbol{\theta}$, we have

$$\frac{\partial S(e^{j2\pi\omega}, e^{j2\pi\nu})}{\partial \sigma^2} = \frac{1}{\sigma^2}\,S(e^{j2\pi\omega}, e^{j2\pi\nu}), \tag{30}$$

$$\frac{\partial S(e^{j2\pi\omega}, e^{j2\pi\nu})}{\partial b(k,\ell)} = 2\sigma^2\,\mathrm{Re}\{\mathcal{B}(e^{j2\pi\omega}, e^{j2\pi\nu})e^{j2\pi(k\omega+\ell\nu)}\}. \tag{31}$$

IV. 2-D Moving Average Parameter Estimation

The parameter estimation algorithm that we present in this section is an extension to two dimensions of the algorithm proposed by Durbin [1] for estimating the parameters of scalar MA processes. The idea is to fit a NSHP AR model to the observed field and then using the estimated AR parameters to estimate the MA model parameters.

It was shown by Whittle [4] that any purely indeterministic random field whose spectral density is analytic in some neighborhood of the unit bicircle and strictly positive on the unit bicircle can be represented by a NSHP AR model of generally infinite dimensions. This result was later extended and was shown to hold even under milder conditions [3]. Hence, any 2-D purely indeterministic MA random field that satisfies the foregoing conditions can be fit with a NSHP AR model. Since parameter estimation algorithms of 2-D AR random fields are available (e.g., [5]), we employ such an algorithm as the first step of the proposed procedure for estimating the parameters of the MA field.

Let $S_{P,Q}$ be defined similarly to $S_{N,M}$, and let $S_{P,Q}\setminus\{(0,0)\}$ be the NSHP support of the MA field AR model. In general, $S_{P,Q}$ is of infinite dimensions. In practice, we must choose finite values for P and Q, and hence, an approximation error is introduced. It is obvious that such a method is necessarily inconsistent, even if the covariance function of the observed field is *a priori* known since no MA field can be exactly modeled by a finite support AR model. However, the bias of the estimates can be made arbitrarily small by sufficiently increasing the support of the AR model $S_{P,Q}$. Therefore, we choose P and Q such that $P \gg N$ and $Q \gg M$, i.e., the finite support of the AR model is chosen to be much larger than that of the MA model. More specifically, let the

$$\boldsymbol{U}_{(k,\ell)} = \begin{bmatrix} \overline{\boldsymbol{U}}_{(k,\ell)} & \mathbf{0}_{T\times(T+2M)} & \cdots & & \mathbf{0}_{T\times(T+2M)} \\ \mathbf{0}_{T\times(T+2M)} & \overline{\boldsymbol{U}}_{(k,\ell)} & \mathbf{0}_{T\times(T+2M)} & \cdots & \mathbf{0}_{T\times(T+2M)} \\ & & \ddots & & \\ \mathbf{0}_{T\times(T+2M)} & \cdots & & \overline{\boldsymbol{U}}_{(k,\ell)} & \mathbf{0}_{T\times(T+2M)} \\ \mathbf{0}_{T\times(T+2M)} & \cdots & & \mathbf{0}_{T\times(T+2M)} & \overline{\boldsymbol{U}}_{(k,\ell)} \end{bmatrix}. \tag{24}$$

2-D finite support MA model of the data be given by (3), and let the approximated finite support NSHP AR model of the same field be given by

$$y(n,m) = -\sum_{(k,\ell)\in S_{P,Q}\setminus\{(0,0)\}} a(k,\ell)y(n-k,m-\ell) + u(n,m). \tag{32}$$

Define

$$\mathcal{B}(z_1,z_2) = \sum_{(k,\ell)\in S_{N,M}} b(k,\ell)z_1^{-k}z_2^{-\ell} \tag{33}$$

and

$$\mathcal{A}(z_1,z_2) = \sum_{(k,\ell)\in S_{P,Q}} a(k,\ell)z_1^{-k}z_2^{-\ell} \tag{34}$$

where $a(0,0)=1$. We therefore have the approximate relation

$$\mathcal{A}(z_1,z_2)\mathcal{B}(z_1,z_2) \approx 1. \tag{35}$$

Let

$$\tilde{\boldsymbol{b}} = [b(0,1),\cdots,b(0,M),b(1,-M),\cdots,b(1,M) \cdots,b(N,-M),\cdots,b(N,M)]^T. \tag{36}$$

Similarly, let

$$\boldsymbol{a}_0 = [a(0,1),\cdots,a(0,Q),\boldsymbol{0}_M]^T \tag{37}$$

$$\boldsymbol{a}_1 = [a(1,-(Q-1)),\cdots,a(1,0),\cdots,a(1,Q),\boldsymbol{0}_M]^T \tag{38}$$

$$\boldsymbol{a}_k = [\boldsymbol{0}_M,a(k,-Q),\cdots,a(k,0),\cdots,a(k,Q),\boldsymbol{0}_M]^T \quad 2\le k\le P \tag{39}$$

and

$$\boldsymbol{a} = [\boldsymbol{a}_0^T,\boldsymbol{a}_1^T,\cdots,\boldsymbol{a}_P^T,\boldsymbol{0}^T_{[2(Q+M)+1]N}]^T. \tag{40}$$

We can now set the following linear system of equations by equating the coefficients of identical powers of $z_1^{-k}z_2^{-\ell}$

$$\boldsymbol{A}\tilde{\boldsymbol{b}} + \boldsymbol{a} = \boldsymbol{e} \tag{41}$$

where $\boldsymbol{e}$ is the approximation error vector, and we have $\boldsymbol{A}$ as the block matrix, shown in (42) at the bottom of the page. Each of the blocks of $\boldsymbol{A}$ is a Toeplitz matrix. The structure of the blocks is given as

$$\boldsymbol{A}_0 = \begin{bmatrix} \multicolumn{4}{c}{\boldsymbol{0}_{Q\times(2M+1)}} \\ 1 & 0 & \cdots & 0 \\ a(0,1) & 1 & \cdots & 0 \\ \vdots & & \ddots & 0 \\ \vdots & \vdots & & 1 \\ \vdots & & & \vdots \\ a(0,Q) & & \cdots & a(0,Q-2M) \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & a(0,Q) \end{bmatrix} \tag{43}$$

and for $1\le k\le P$

$$\boldsymbol{A}_k = \begin{bmatrix} a(k,-Q) & 0 & \cdots & 0 \\ a(k,-(Q-1)) & a(k,-Q) & \cdots & 0 \\ \vdots & & \ddots & 0 \\ \vdots & \vdots & & a(k,-Q) \\ \vdots & & & \vdots \\ a(k,Q) & & \cdots & a(k,Q-2M) \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & a(k,Q) \end{bmatrix}. \tag{44}$$

Note that the matrices $\boldsymbol{A}_k, 0\le k\le P$ are all $[2(Q+M)+1]\times(2M+1)$-dimensional matrices. In addition, let $\bar{\boldsymbol{A}}_0$ be the $(2Q+M)\times(2M+1)$ sub-block of $\boldsymbol{A}_0$ consisting of its $2Q+M$ lower rows, and let $\boldsymbol{A}_0^0$ be the $(Q+M)\times M$ lower right sub-block of $\boldsymbol{A}_0$. Similarly, let $\bar{\boldsymbol{A}}_k, 1\le k\le P$ be the $(2Q+M)\times M$ lower right sub-block of $\boldsymbol{A}_k$, and

$$\boldsymbol{A}_k^0 = \begin{bmatrix} \boldsymbol{0}_{(M+1)\times M} \\ \bar{\boldsymbol{A}}_k \end{bmatrix}. \tag{45}$$

The MA model parameters can now be found by minimizing the sum of the squared approximation error. The solution to this linear least squares problem is

$$\tilde{\boldsymbol{b}} = -(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{a}. \tag{46}$$

In the actual solution for the MA model parameters, the parameters of the AR model $\{a(k,\ell)\}$ are replaced by their

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_0^0 & \boldsymbol{0}_{(Q+M)\times(2M+1)} & \boldsymbol{0}_{(Q+M)\times(2M+1)} & \cdots & \boldsymbol{0}_{(Q+M)\times(2M+1)} \\ \bar{\boldsymbol{A}}_1 & \bar{\boldsymbol{A}}_0 & \boldsymbol{0}_{(2Q+M)\times(2M+1)} & \cdots & \boldsymbol{0}_{(2Q+M)\times(2M+1)} \\ \boldsymbol{A}_2^0 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \cdots & \boldsymbol{0}_{[2(Q+M)+1]\times(2M+1)} \\ \vdots & \vdots & & \ddots & \vdots \\ \vdots & \vdots & & & \boldsymbol{0}_{[2(Q+M)+1]\times(2M+1)} \\ \boldsymbol{A}_N^0 & \boldsymbol{A}_{N-1} & \cdots & \cdots & \boldsymbol{A}_0 \\ \vdots & \vdots & & & \vdots \\ \boldsymbol{A}_P^0 & \boldsymbol{A}_{P-1} & \cdots & \cdots & \boldsymbol{A}_{P-N} \\ \boldsymbol{0}_{[2(Q+M)+1]\times M} & \boldsymbol{A}_P & \cdots & \cdots & \boldsymbol{A}_{P-N-1} \\ \boldsymbol{0}_{[2(Q+M)+1]\times M} & \boldsymbol{0}_{[2(Q+M)+1]\times(2M+1)} & \boldsymbol{A}_P & \cdots & \boldsymbol{A}_{P-N-2} \\ \vdots & \vdots & & \ddots & \vdots \\ \boldsymbol{0}_{[2(Q+M)+1]\times M} & \boldsymbol{0}_{[2(Q+M)+1]\times(2M+1)} & \cdots & & \boldsymbol{A}_P \end{bmatrix}. \tag{42}$$

estimated values, which have been obtained by solving the corresponding set of 2-D normal equations using the *estimated* covariance function.

Finally, note that the proposed algorithm is derived using no *a priori* assumptions regarding the probability density function of the observed field. It is, therefore, applicable to Gaussian MA fields, as well as to non-Gaussian ones.

V. The Maximum Likelihood Estimator

The main advantage of the "MA by AR" algorithm of the previous section is that it requires only the solution of two sets of linear systems of equations [one to estimate the AR parameters by solving the set of 2-D normal equations and the other the solution (46) to (41)]. In particular, there is no need for an iterative solution. However, as indicated in Section IV, the estimates are biased and inconsistent. Hence, improved estimation algorithms are required in cases where the performance of the "MA by AR" algorithm is not acceptable. The "MA by AR" algorithm can then serve to initialize a more sophisticated algorithm. One such estimator is the maximum likelihood estimator (MLE) for Gaussian MA fields, which we derive in this section.

Since the observed field $\{y(n,m)\}$ is Gaussian, the log-likelihood function of the observations is given by

$$\log p(\boldsymbol{y};\boldsymbol{\theta}) = -\frac{ST}{2}\log(2\pi) - \frac{1}{2}\log(\det\boldsymbol{\Gamma}) - \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{y}. \tag{47}$$

The MLE of the field parameters is found by maximizing $\log p(\boldsymbol{y};\boldsymbol{\theta})$ with respect to the MA model parameters. Since this objective function is highly nonlinear in the problem parameters, the maximization problem cannot be solved analytically, and we must resort to numerical methods. In order to avoid the enormous computational burden of an exhaustive search, we used the following two-step procedure. In the first stage, we obtain a suboptimal initial estimate for the parameters of the MA model by using the algorithm described in Section IV. In the second stage, we refine these initial estimates by an iterative numerical maximization of the log likelihood function. In our experiments, we used the Broyden–Fletcher–Goldfarb–Shanno (BFGS) quasi-Newton optimization method [17], [18]. This algorithm requires evaluation of the first derivative of the objective function at each iteration.

Next, we derive expressions for the first derivatives of the objective function (47) with respect to $\boldsymbol{\theta}$. In general

$$\frac{\partial\log p(\boldsymbol{y};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i} = -\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial\boldsymbol{\theta}_i}\right\} + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial\boldsymbol{\theta}_i}\boldsymbol{\Gamma}^{-1}\boldsymbol{y} \tag{48}$$

where $\boldsymbol{\theta}_i$ is the ith element of $\boldsymbol{\theta}$.

Hence, using (11), (23), and (25), we obtain

$$\frac{\partial\log p(\boldsymbol{y};\boldsymbol{\theta})}{\partial\sigma^2} = -\frac{ST}{2\sigma^2} + \frac{1}{2\sigma^2}\boldsymbol{y}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{y} \tag{49}$$

$$\frac{\partial\log p(\boldsymbol{y};\boldsymbol{\theta})}{\partial b(k,\ell)} = -\frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial b((k,\ell)}\right\} + \frac{1}{2}\boldsymbol{y}^T\boldsymbol{\Gamma}^{-1}\frac{\partial\boldsymbol{\Gamma}}{\partial b(k,\ell)}\boldsymbol{\Gamma}^{-1}\boldsymbol{y}. \tag{50}$$

TABLE I
Squared Roots of the CRB on the Parameters of the Two Fields

	Field I		*Field II*	
	Parameters	$(CRB)^{\frac{1}{2}}$	*Parameters*	$(CRB)^{\frac{1}{2}}$
σ^2	1	0.04778	1	0.06444
$b(0,1)$	-0.9	0.02318	0.9	0.04441
$b(1,-1)$	0.1	0.03171	-0.2	0.04156
$b(1,0)$	-0.5	0.04063	0.5	0.04337
$b(1,1)$	0.4	0.03343	0.1	0.04248

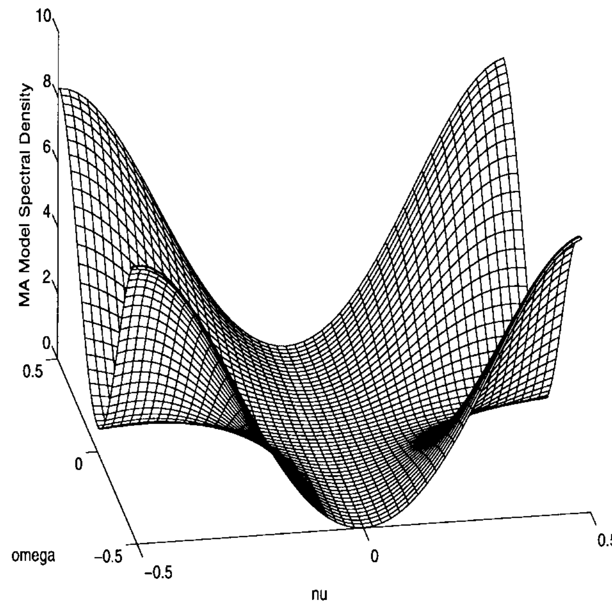

Fig. 1. Spectral density function of Field I.

As is well known, this type of iterative optimization procedure converges to a local maximum and does not guarantee global optimality, unless the initial estimate is sufficiently close to the global optimum. As we show in Section VI, the initial estimates provided by the "MA by AR" algorithm proposed in Section IV appear to provide a good initial starting point (i.e., one that leads to convergence to the global maximum). The performance of the ML algorithm is discussed in more detail in Section VI and is compared with the CRB, which was derived earlier.

VI. Numerical Examples

To gain more insight into the performance of the proposed algorithms relatively to the CRB, we resort to numerical evaluation of some specific examples. We present several such examples, which illustrate the dependence of the algorithms bias and error variance on the dimensions of the support of the approximating AR model and on the size of the observed data field.

Example 1: Consider the two NSHP MA fields with support $S_{1,1}$, whose parameters are listed in Table I. In this example, we evaluate the CRB on the error variance in estimating the models parameters, as well as the bound on the error variance in estimating the spectral densities of the two fields. The dimensions of the observed field, for both models, are relatively small: $S = T = 30$. The squared roots of the CRB on the parameters of the two models are listed in Table I.

The spectral density function of Field I is depicted in Fig. 1, and the CRB on the error variance in estimating the spectral

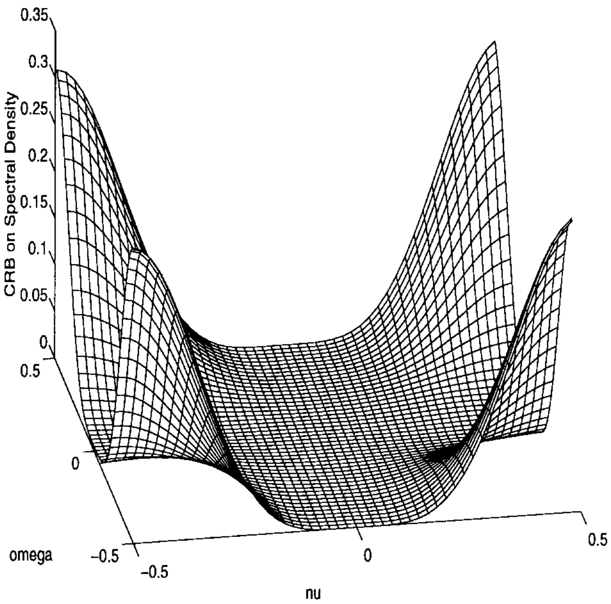

Fig. 2. CRB on the spectral density function of Field I.

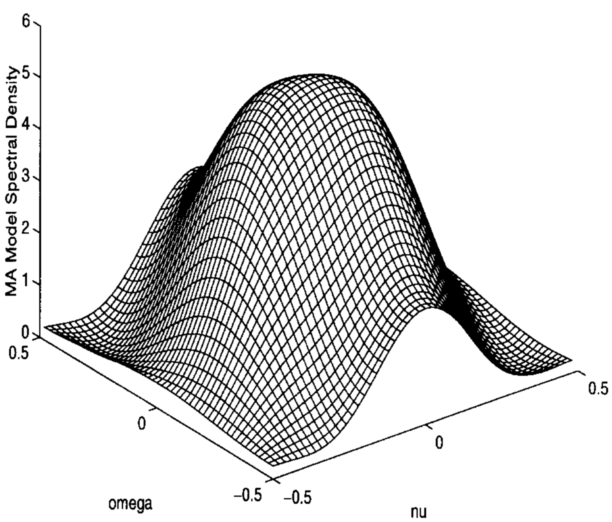

Fig. 3. Spectral density function of Field II.

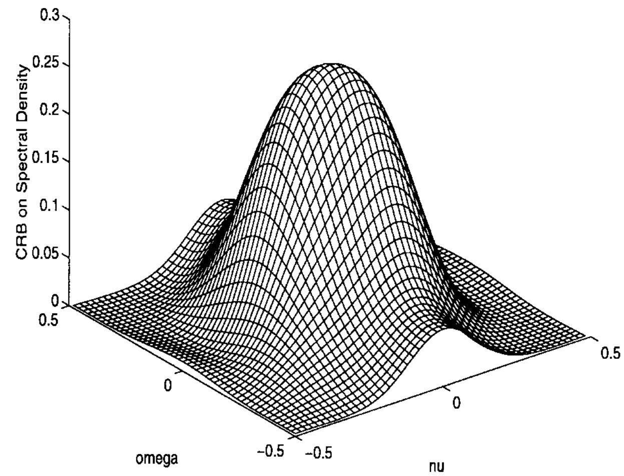

Fig. 4. CRB on the spectral density function of Field II.

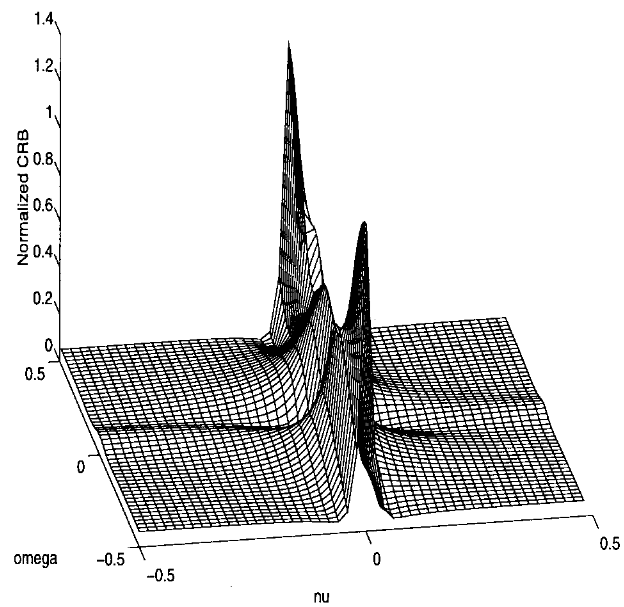

Fig. 5. Ratio of the square root of the CRB to the spectral density function for Field I.

density is depicted in Fig. 2. The spectral density function of Field II is depicted in Fig. 3, and the bound on the error variance in estimating the density is depicted in Fig. 4. Note that in both examples, the shape of the bound as a function of frequency matches the shape of the MA field spectral density function. In order to further investigate the dependence of the CRB on the shape of the spectral density, we depict in Fig. 5 the normalized CRB for Field I, i.e., the ratio of the squared root of the CRB to the spectral density function of the MA field. Note that the estimation of the MA field spectral density function is relatively less accurate in frequency regions where the spectral density function is close to zero than in regions of higher spectral density.

Example 2: Consider Field I—the NSHP MA field with support $S_{1,1}$ from Example 1. Using 100 Monte Carlo runs, we investigate the performance of the proposed "MA by AR" estimation algorithm, as well as that of the ML algorithm, which is initialized using the "MA by AR" algorithm. Three approximating NSHP AR models with different supports are considered. Table II lists the bias and standard deviation of the estimates for each of the supports, as well as the square root of the CRB. Note, however, that since the "MA by AR" estimation algorithm is found to be biased, the parameters estimation error variances are not comparable with the CRB, which is the lower bound on any unbiased estimator of the model parameters.

From the results summarized in Table II, we conclude that for the "MA by AR" algorithm, increasing the dimensions of the approximating AR model support reduces the bias of the estimated MA model parameters and increases the standard deviation of the estimation error. The overall effect of increasing the dimensions of the approximating AR model support is a smaller mean squared error. The ML algorithm is slightly biased, due to the small dimensions of the observed field. However, its bias is considerably lower than that of the

TABLE II
ESTIMATION RESULTS OF FIELD I MA MODEL PARAMETERS USING DIFFERENT NSHP AR SUPPORTS, FOR THE "MA BY AR" AND THE MAXIMUM LIKELIHOOD ALGORITHMS

Parameters		$\sigma^2 = 1$	$b(0,1) = -0.9$	$b(1,-1) = 0.1$	$b(1,0) = -0.5$	$b(1,1) = 0.4$
	$(CRB)^{\frac{1}{2}}$	0.04778	0.02318	0.03171	0.04063	0.03343
$(P,Q) = (4,4)$	bias	0.16245	0.22799	-0.02720	0.04957	-0.12292
	std	0.05488	0.01808	0.02803	0.03238	0.02794
	bias ML	-0.04633	-0.04943	0.00702	-0.00954	0.02309
	std ML	0.08869	0.08403	0.03502	0.04089	0.04433
$(P,Q) = (7,7)$	bias	0.07848	0.17654	-0.02872	0.04631	-0.10275
	std	0.04936	0.02305	0.02980	0.03647	0.03181
	bias ML	-0.05614	-0.05808	0.00802	-0.00965	0.02515
	std ML	0.08268	0.08599	0.03449	0.03945	0.04073
$(P,Q) = (10,10)$	bias	0.01150	0.16728	-0.02953	0.04987	-0.10046
	std	0.05323	0.02452	0.03025	0.03776	0.03129
	bias ML	-0.05411	-0.05605	0.00584	-0.01213	0.02477
	std ML	0.09143	0.08758	0.03491	0.03907	0.04243

TABLE III
ESTIMATION RESULTS OF FIELD I MA MODEL PARAMETERS FOR DIFFERENT DATA SIZES, USING THE "MA BY AR" ALGORITHM

Parameters		$S = 30$, $T = 30$		$S = 100$, $T = 100$	
		bias	std	bias	std
σ^2	1	0.01150	0.05323	0.05509	0.01644
$b(0,1)$	-0.9	0.16728	0.02452	0.08664	0.00523
$b(1,-1)$	0.1	-0.02953	0.03025	-0.01317	0.00903
$b(1,0)$	-0.5	0.04987	0.03776	0.01500	0.01123
$b(1,1)$	0.4	-0.10046	0.03129	-0.04679	0.00842

"MA by AR" algorithm, which is used for its initialization. Hence, the mean squared error (MSE) in estimating the model parameters using the ML algorithm is smaller than the MSE of the "MA by AR" algorithm.

Note that since the dimensions of the observed data field are relatively small, increasing P and Q so that $(P,Q) = (10,10)$, rather than (7, 7), does not have a considerable effect on the bias nor on the variance of the estimation error for the "MA by AR" algorithm. Since the "MA by AR" algorithm provides the initial conditions for the ML algorithm, it is clear that no improvement in the accuracy of the ML algorithm can be expected. The reason for the lack of performance gain when P and Q are increased from $(P,Q) = (7,7)$ to $(P,Q) = (10,10)$ is the large error in estimating the covariance function of the field using small data size. In order to further illustrate this point, we have repeated the experiment using 100 Monte Carlo runs for a 100×100 observed field. The support of the approximating AR model was chosen to be $S_{10,10}$. The estimation results for both data sizes are compared in Table III. It is clear that the bias and variance of the estimates drop sharply when the dimensions of the observed field are larger since the estimates of the covariance function are more accurate.

Finally, we note that in cases where the data size is not very small, the ML algorithm computational and storage requirements (due to the dimensions of the covariance matrix) make it impractical. On the other hand, the "MA by AR" algorithm is considerably less complex with respect to both the computations and storage requirements. Since its bias decreases with the increase in the data size, the "MA by AR" algorithm is much more useful in these cases.

VII. CONCLUSIONS

In this paper, we studied the problem of estimating the parameters of 2-D MA random fields. We first addressed the problem of expressing the covariance matrix of various types of MA random fields in terms of the model parameters. This derivation was then employed to derive a maximum likelihood algorithm and the CRB on the error variance in jointly estimating the model parameters for Gaussian MA fields. A suboptimal "MA by AR" algorithm for estimating the parameters of the MA model was developed. This algorithm has low computational complexity but is biased. The bias decreases when increasing the dimensions of the observed field or the dimensions of the approximating NSHP AR model support. It was demonstrated that the "MA by AR" estimator is a good choice for implementing the initialization phase of the maximum likelihood algorithm. Furthermore, as the data size increases, the maximum likelihood method becomes computationally prohibitive, whereas the "MA by AR" algorithm becomes less biased and, therefore, offers an increasingly attractive alternative to ML estimation.

REFERENCES

[1] J. Durbin, "Efficient estimation of parameters in moving average models," *Biometrika*, vol. 46, pp. 306–316, 1959.
[2] 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.
[3] H. Helson and D. Lowdenslager, "Prediction theory and Fourier series in several variables," *Acta Math.*, vol. 99, pp. 165–202, 1958.
[4] P. Whittle, "On stationary processes in the plane," *Biometrika*, vol. 41, pp. 434–449, 1954.
[5] T. L. Marzetta, "Two-dimensional linear prediction: Autocorrelation arrays, minimum-phase prediction error filters and reflection coefficient arrays, *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 725–733, 1980.
[6] R. Cusani and E. Baccarelli, "Identification of 2-D noncausal Gauss-Markov random fields," *IEEE Trans. Signal Processing*, vol. 44, pp. 759–764, 1996.
[7] R. Chellappa and R. L. Kashyap, "Texture synthesis using 2-D noncausal autoregressive models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 194–203, 1985.
[8] F. C. Jeng and J. W. Woods, "Texture discrimination using doubly stochastic gaussian random fields," *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Glasgow, U.K., 1989, pp. 1675–1678.
[9] C. Bouman and B. Liu, "Multiple resolution segmentation of textured images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 99–113, 1991.

[10] J. M. Francos, A. Z. Meiri, and B. Porat, "A unified texture model based on a 2-D wold like decomposition," *IEEE Trans. Signal Processing*, vol. 41, pp. 2665–2678, 1993.
[11] J. M. Francos, A. Narasimhan, and J. W. Woods, "Maximum likelihood parameter estimation of textures using a wold decomposition based model," *IEEE Trans. Image Processing*, vol. 4, pp. 1655–1666, 1995.
[12] R. Krishnamurthy, J. W. Woods, and J. M. Francos, "Adaptive restoration of textured images with mixed spectra using a generalized wiener filter," *IEEE Trans. Image Processing*, vol. 5, pp. 648–652, 1996.
[13] A. J. Isaksson, "Analysis of identified 2-D noncausal models," *IEEE Trans. Inform. Theory*, vol. 39, pp. 525–534, 1993.
[14] J. M. Francos, "Bounds on the accuracy of estimating the parameters of discrete homogeneous random fields with mixed spectral distributions," *IEEE Trans. Inform. Theory*, vol. 43, pp. 908–922, 1997.
[15] C. R. Rao, *Linear Statistical Inference and Its Applications*. New York: Wiley, 1965.
[16] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[17] David G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed. Reading, MA: Addison-Wesley, 1984.
[18] R. Fletcher, *Practical Methods of Optimization: Vol. 1—Unconstrained Optimization*. New York: Wiley, 1980.

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 1990, 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 a Senior Lecturer. He also held visiting positions at the Massachusetts Institute of Technology Media Laboratory, Cambridge, and at the Electrical and Computer Engineering Department, University of California, Davis. 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, image modeling, and texture analysis and synthesis.

Benjamin Friedlander (S'74–M'76–SM'82–F'87) received the B.Sc. and the M.Sc. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, in 1968 and 1972, respectively, and the Ph.D. degree in electrical engineering and the M.Sc. degree in statistics from Stanford University, Stanford, CA, in 1976.

From 1976 to 1985, he was with Systems Control Technology, Inc., Palo Alto, CA. From November 1985 to July 1988, he was with Saxpy Computer Corporation, Sunnyvale, CA. Currently, he is with the University of California, Davis and Signal Processing Technology, Ltd., Palo Alto.

Dr. Friedlander is the recipient of the 1983 ASSP Senior Award, the 1985 Award for the Best Paper of the Year from the European Association for Signal Processing (EURASIP), and the 1989 Technical Achievement Award of the Signal Processing Society.