Noncausal Gauss Markov Random Fields: Parameter Structure and Estimation

Nikhil Balram and José M. F. Moura, *Senior Member, IEEE*

Abstract—**The parameter structure of noncausal homogeneous Gauss Markov random fields (GMRF) defined on finite lattices is studied. For first-order (nearest neighbor) and a special class of second-order fields, we provide a complete characterization of the parameter space and a fast implementation of the maximum likelihood (ML) estimator of the field parameters. For general higher order fields, tight bounds for the parameter space are presented and an efficient procedure for ML estimation is described. Experimental results illustrate the application of the approach presented and the viability of the present method in fitting noncausal models to 2-D data.**

Index Terms—**Random fields, noncausal, Gauss Markov random fields, maximum likelihood parameter estimation, parameter space in random fields.**

I. Introduction

NONCAUSAL Markov Random Fields (MRF) are an important paradigm in multidimensional signal processing. In this paper, we study the parameter structure of noncausal Gauss MRF's (GMRF) defined on finite lattices. GMRF's on lattices may result from sampling spatially dependent phenomena described for example by an elliptic equation or may correspond to a snapshot in a marching problem as with the discretization of a parabolic partial differential equation (PDE). Examples of applications include modeling circulation and transport phenomena in Physical Oceanography, e.g., steady state problems like Munk's model or Stommel's problem [31, ch. 3]. Another area where GMRF's have found widespread use is in image processing.

Multidimensional signal processing, and in particular two dimensional (2-D) processing where the two indexing variables are spatial coordinates, raises interesting challenges that go beyond the "curse" of higher dimensions. The difficulties we will deal with in this paper stem from the absence of causality or unilateral structure for 2-D fields. For 1-D signal processing, where time is the variable of interest, causality is natural, leading to expedient processing algorithms that are recursive like in the Kalman–Bucy filtering theory. In order to handle higher dimensions and to extend 1-D recursive filters to 2-D, many researchers have compromised the noncausality of the field by assuming various causality constraints, e.g., the Markov Mesh models [1] or the nonsymmetric half plane fields [38].

Working directly with general noncausal MRF's has found renewed interest since the work of [13], [16], [26]. These and other authors have addressed enhancement and segmentation issues in the context of image processing. The optimal processing algorithms are iterative, based upon the Metropolis algorithm [27] and simulated annealing [22]. These procedures are computationally very expensive. More recent work has developed suboptimal or practical alternatives to simulated annealing, see, for example, [10], [11]. In earlier work, by assuming periodic boundary conditions for the finite lattice, researchers have derived optimal fast transform filtering algorithms that exploit the circulant structure of toroidal field matrices, see, for example, [9].

In [29], [30], we have developed a recursive framework for noncausal finite lattice GMRF's with nonperiodic boundary conditions, which enables the use of fast optimal recursive techniques like Kalman–Bucy filtering without compromising the noncausality of the field model. This framework is based on the existence of an equivalent *spatially varying* one-sided representation for every noncausal finite lattice GMRF. A more complete understanding of this framework, and its application to 2-D signal processing, requires further the study of the parameter structure of these fields and the problem of parameter estimation.

Parameter estimation with noncausal MRF's is very much an open problem. An associated problem that also remains open is the specification of the parameter space. Again, the case of toroidal (periodic) fields has been well studied, see, for example, [7], [20], [23], [24], for necessary and sufficient conditions on the parameter space, and, e.g., [7] for optimal parameter estimation. Alternatively, in the case of fields defined on infinite lattices, parameter space descriptions are available, see, for example, [23], [24], while parameter estimation is treated, for example, in [35] and the references therein. In contrast, for nonperiodic finite lattices, barring special cases, notably that of isotropic (single parameter) first-order fields with the so-called free boundary conditions [4], [32], in general, the parameter space specification is limited to a loose sufficient condition, while optimal parameter estimation is hampered by the high cost associated with the computation of the partition function. See [8] for an overview of parameter estimation procedures for finite and infinite lattice GMRF's.

In this paper, we consider parameter estimation for homogeneous *noncausal* finite lattice GMRF's of *arbitrary order* with nonperiodic boundary conditions (b.c.). For first-order fields with various b.c. and a certain class of second-order

Manuscript received April 18, 1991; revised November 20, 1992. This work was supported in part by ONR Grant # N00014-91-J-1001. This work was presented in part at the 25th Annual Conference Information Science and Systems, Baltimore, MD, March 20, 1991.

N. Balram is with IBM, 1000 N. W. 51st Street, Boca Raton FL 33432.

J. M. F. Moura is with the Department Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213.

IEEE Log Number 920458.

fields, we derive exactly the parameter space of noncausal GMRF's. We design for these fields a simple and fast implementation for the maximum likelihood (ML) estimator of the field parameters. The basic difficulty with ML estimators lies in the computation of the partition function (normalizing constant of the associated probability function). Because we establish explicitly the eigenstructure of the inverse of the field covariance matrix, we have a closed analytical expression for the partition function. For general higher order noncausal fields, we provide tight upper bounds on the parameter space and develop an efficient algorithm for ML estimation. The algorithm uses the recursive framework for noncausal GMRF's [30] to overcome the problems that have made ML estimation for noncausal fields computationally intractable.

This paper derives for nonperiodic fields the parameter space specification and optimal parameter estimation procedures that complement the work done by other authors for periodic fields, e.g., [7], [8], [20], [24]. It is interesting to note that while for periodic fields the work of these authors has led to algorithms which are usually of the transform type, our approach utilizes Kalman–Bucy recursive algorithms.

As a final comment, we would like to say that there is a strong connection between the structure of the matrices we deal with and the theory of Perron–Frobenius for nonnegative matrices, e.g., [37].

The organization of the paper is as follows. In Section II, we present the main terminology and notation that we use. In Section III, we discuss briefly the framework underlying our work which focuses on the inverse of the covariance matrix, rather than the field covariance itself. In Section IV, we study the eigenstructure of the inverse covariance matrix. In Section V, the results from the previous section are applied to derive necessary and sufficient conditions for the parameter spaces of first-order fields and a special class of second-order fields, and to study sufficient conditions for higher order fields. In Section VI, we present the ML estimation procedure for fields of arbitrary order. A simpler ML procedure for first-order and a special class of second-order fields is presented in Section VII. Section VIII explains modifications that enable these procedures to perform well on noisy data, while in Section IX, experimental verification is provided. Finally, Section X concludes the paper. To facilitate the discussion all proofs are relegated to the Appendix.

II. Terminology and Notation

A finite $N \times M$ lattice, L, is defined as a set of sites, (i,j) such that $L = \{(i,j) : 1 \le i \le N, 1 \le j \le M\}$. We define the neighborhood order using an Euclidean distance based measure, as done, for example, in [13], [16].

Definition 1: A pth-order neighborhood is defined on an $N \times M$ lattice, L, as the set of neighbors of site (i,j)

$$S_p^{i,j} = \{(k,l) : 0 < (i-k)^2 + (j-l)^2 \le D_p\}, \quad (1)$$

where D_p is an increasing function of p that represents the square of the Euclidean distance between a site and its furthest neighbor.

Usually, e.g., [13], [16], D_p takes on values 1, 2, 4, 5, 8, 9, and so on, for $p = 1, 2, 3, 4, 5, 6, \cdots$. See Fig. 1(a) in Section IV for a hierarchical sequence of neighborhoods produced by this definition. Note that a neighborhood set of order p includes all the neighbors of sets of order 1 to $p-1$.

The symbol $\otimes$ represents the Kronecker product, also known sometimes as the direct product or the tensor product, see, for example, [14]. The Kronecker product $A \otimes B$ of two matrices, $A_{N\times N}$ and $B_{M\times M}$, is defined as the $NM \times NM$ matrix formed by placing at each location (i,j) in A, a copy of B scaled by the (i,j)th element of A.

We introduce below matrices that are relevant in the sequel. The vector $\vec{e}_i$ is the ith unit vector. Let

$$K_M^1 = \begin{bmatrix} 0 & 0 & . & . & . \\ 1 & 0 & 0 & . & . \\ 0 & 1 & 0 & 0 & . \\ . & . & . & . & . \\ . & . & 0 & 1 & 0 \end{bmatrix},$$

$$H_M^1 = \begin{bmatrix} 0 & 1 & 0 & . & . \\ 1 & 0 & 1 & 0 & . \\ . & . & . & . & . \\ . & 0 & 1 & 0 & 1 \\ . & . & 0 & 1 & 0 \end{bmatrix},$$

i.e., $K_M^1 = [\vec{e}_2 \cdots \vec{e}_M, \vec{0}]$, $H_M^1 = K_M^1 + (K_M^1)^T$. More generally,

$$(K_M^j)^T = \begin{array}{c} \begin{matrix} \mathbf{1} & \mathbf{2} & . & \mathbf{j+1} & \cdots & \mathbf{M} \end{matrix} \\ \begin{bmatrix} 0 & 0 & . & 1 & 0 & . \\ 0 & 0 & . & . & 1 & 0 \\ . & . & . & . & . & . \\ . & . & . & . & 0 & 1 \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 0 \end{bmatrix} \end{array} \begin{matrix} \mathbf{1} \\ \mathbf{2} \\ . \\ \boldsymbol{M-j} \\ . \\ \boldsymbol{M} \end{matrix} \quad (2)$$

$$= [\vec{e}_{j+1} \cdots \vec{e}_M, \vec{0} \cdots \vec{0}]^T, \quad (3)$$

and

$$H_M^j = \begin{array}{c} \begin{matrix} \mathbf{1} & \mathbf{2} & . & \mathbf{j+1} & . & . & . & . & \mathbf{M} \end{matrix} \\ \begin{bmatrix} 0 & 0 & . & 1 & 0 & . & . & . & . & . \\ 0 & 0 & . & 0 & 1 & 0 & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ 1 & 0 & . & . & . & 0 & 1 & 0 & . & . \\ 0 & 1 & . & . & . & . & . & 1 & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & 1 & . & . & . & . & . & 1 \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & 1 & . & . & 0 & 0 \end{bmatrix} \end{array} \begin{matrix} \mathbf{1} \\ \mathbf{2} \\ . \\ \boldsymbol{j+1} \\ . \\ \boldsymbol{M-j} \\ . \\ \boldsymbol{M} \end{matrix} \quad (4)$$

$$= K_M^j + (K_M^j)^T, \quad (5)$$

where the rows and columns of these matrices are indicated along the top and side, respectively. The jth upper diagonal of $(K_M^j)^T$ and the jth lower and upper diagonals of H_M^j are composed of 1's, while all others are zero.

We define I_N as the N dimensional identity matrix, $\underline{0}$ as a matrix of zeros whose dimensions are determined by the

context, and the following set of four $NM \times NM$ symmetric matrices, which we call *interaction matrices*, that will be used in the sequel to represent the structure of higher order fields:

1) I_H^j*:*

$$I_H^j = \begin{bmatrix} H_M^j & \underline{0} & . & . & . \\ \underline{0} & H_M^j & \underline{0} & . & . \\ . & . & . & . & . \\ . & . & . & H_M^j & \underline{0} \\ . & . & . & \underline{0} & H_M^j \end{bmatrix} \tag{6}$$

$$= I_N \otimes H_M^j. \tag{7}$$

2) H_i^I*:*

$$H_i^I = \begin{bmatrix} \underline{0} & \underline{0} & . & I_M & \underline{0} & . & . & . & . & . \\ \underline{0} & \underline{0} & . & \underline{0} & I_M & \underline{0} & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ I_M & \underline{0} & . & . & . & \underline{0} & I_M & \underline{0} & . & . \\ \underline{0} & I_M & . & . & . & . & . & I_M & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & I_M & . & . & . & . & . & I_M \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & I_M & . & . & \underline{0} & \underline{0} \end{bmatrix} \tag{8}$$

$$= H_N^i \otimes I_M. \tag{9}$$

The matrix H_i^I has the structure of H_N^i with each 0 replaced by a $M \times M$ matrix of zeros, and each 1 replaced by the $M \times M$ identity matrix I_M, i.e., the ith block diagonals above and below the main block diagonal are composed of copies of I_M.

3) $P_{i,j}$*:* (see (10) and (11) at the bottom of the page.) To aid the discussion, the block rows and columns of $P_{i,j}$ are labeled along the top and side of the matrix, respectively. $P_{i,j}$ has the structure of the decomposition of H_N^i, defined similarly to (5), with the 0's in K_N^i being replaced by $M \times M$ matrices of zeros, the 1's being replaced by K_M^j, and the resulting matrix being added to its symmetric counterpart. The result is a block matrix with only the ith block diagonals above and below the main block diagonal being nonzero, and these have copies of K_M^j and $(K_M^j)^T$, respectively, along them.

4) $P_{i,j,T}$*:* (see (12) and (13) at the bottom of the page.)The block structure of $P_{i,j,T}$ is similar to $P_{i,j}$ except that it has $(K_M^j)^T$ in place of K_M^j and vice versa.

At the element level, the four matrices defined in (6)–(13) have only one nonzero diagonal above the main diagonal and one, its symmetric counterpart, below it. In the case of H_i^I, the nonzero diagonals are the iMth diagonals above and below the main diagonal, and all the entries of these diagonals are 1's. For I_H^j, $P_{i,j}$, $P_{i,j,T}$, the nonzero diagonals are, respectively, the jth, $(iM + j)$th, and $(iM - j)$th, diagonals above and below the main diagonal. For these three matrices, not all the entries in the nonzero diagonals are 1's. Instead, the nonzero diagonals have strings of $(M - j)$ consecutive 1's alternating with strings of j consecutive 0's.

$$P_{i,j} = \begin{matrix} \\ \end{matrix} \begin{matrix} \mathbf{1} & \mathbf{2} & . & \boldsymbol{i+1} & . & . & \mathbf{2}\boldsymbol{i}+\mathbf{1} & . & . & \boldsymbol{N} \end{matrix}$$

$$= \begin{bmatrix} \underline{0} & \underline{0} & . & (K_M^j)^T & \underline{0} & . & . & . & . & . \\ \underline{0} & \underline{0} & . & \underline{0} & (K_M^j)^T & \underline{0} & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ K_M^j & \underline{0} & . & . & . & \underline{0} & (K_M^j)^T & \underline{0} & . & . \\ \underline{0} & K_M^j & . & . & . & . & . & (K_M^j)^T & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & K_M^j & . & . & . & . & . & (K_M^j)^T \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & K_M^j & . & . & \underline{0} & \underline{0} \end{bmatrix} \begin{matrix} \mathbf{1} \\ \mathbf{2} \\ . \\ \boldsymbol{i+1} \\ . \\ . \\ \boldsymbol{N-i} \\ . \\ \boldsymbol{N} \end{matrix} \tag{10}$$

$$= K_N^i \otimes K_M^j + (K_N^i)^T \otimes (K_M^j)^T, \tag{11}$$

$$P_{i,j,T} = \begin{matrix} \\ \end{matrix} \begin{matrix} \mathbf{1} & \mathbf{2} & . & \boldsymbol{i+1} & . & . & \mathbf{2}\boldsymbol{i}+\mathbf{1} & . & . & \boldsymbol{N} \end{matrix}$$

$$\begin{bmatrix} \underline{0} & \underline{0} & . & K_M^j & \underline{0} & . & . & . & . & . \\ \underline{0} & \underline{0} & . & \underline{0} & K_M^j & \underline{0} & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ (K_M^j)^T & \underline{0} & . & . & . & \underline{0} & K_M^j & \underline{0} & . & . \\ \underline{0} & (K_M^j)^T & . & . & . & . & . & K_M^j & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & (K_M^j)^T & . & . & . & . & . & K_M^j \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & (K_M^j)^T & . & . & \underline{0} & \underline{0} \end{bmatrix} \begin{matrix} \mathbf{1} \\ \mathbf{2} \\ . \\ \boldsymbol{i+1} \\ . \\ . \\ \boldsymbol{N-i} \\ . \\ \boldsymbol{N} \end{matrix} \tag{12}$$

$$= K_N^i \otimes (K_M^j)^T + (K_N^i)^T \otimes K_M^j. \tag{13}$$

III. Underlying Framework

We discuss now briefly the approach we use in this paper to derive the desired results. In recent work [29], [30], we have constructed a framework for characterizing noncausal finite $(N \times M)$ lattice GMRF's. This framework is based on the structure and properties of the inverse of the field covariance matrix, which we call the potential matrix. For zero mean Gauss fields the exponent of the joint Gauss probability density function (pdf) can be written as:

$$U(X) = -\frac{1}{2\sigma^2}\vec{X}^T A\vec{X}, \tag{14}$$

where $\vec{X}$ is the $NM \times 1$ vector of the $N \times M$ field variables which we arrange in lexicographic order, σ is a positive scaling constant, and A is an $NM \times NM$ matrix that is a scaled version of the inverse of the field covariance matrix. When, besides being Gaussian, the field is (noncausal) Markovian of order p, A is a highly sparse and structured matrix, block banded, with blocks that are themselves banded, see [30] for details. The elements of A represent the field interactions between a lattice site and its neighbors. These parameters are also called field potentials, reflecting the equivalence between MRF's and Gibbs fields from Statistical Mechanics, [4]. The specification of A for an arbitrary pth-order GMRF is provided in [30]. We consider in this work nondegenerate GMRF's, i.e., those with positive definite (pd) covariance matrices. We study fields with potentials that are independent of the site locations. Such fields are said to be *homogeneous* or *spatially invariant*. At the boundaries, homogeneity is interpreted to mean that the same boundary condition rule, along with the same boundary potentials, is applied everywhere. The boundary conditions are usually drawn from the PDE literature. Examples are:

1) *Free or Dirichlet b.c.:* Here, the field values of the off-lattice neighbors of a boundary pixel are set to zero; consequently the boundary pixels have fewer neighbors. These boundary conditions lead to the *autonormal* models considered in [4]. Since this term was later applied in [5] to all GMRF's, in order to avoid confusion in this paper, we will call these *Dirichlet* fields.
2) *Asymmetric Neumann b.c.:* The off-lattice neighbors of a boundary pixel are assumed to have the same intensity as the pixel. We call this the *variational* field. A field with similar potentials has been used in [26] to represent quadratic intensities.
3) *Symmetric Neumann b.c.:* The off-lattice neighbors of a boundary pixel are assumed to have the same intensity values as the pixels that are their mirror images when reflected across the x and y axes (defined with the boundary pixel as their origin). We call this the *symmetric* field.

Here, we investigate the structure of the potential matrix further; in particular we consider the decomposition of A into the sum of symmetric matrices whose eigenvalues can be derived. In the case of first-order fields with the boundary conditions previously mentioned, and a special second-order field, we choose decompositions in which the eigenvalues of the component matrices can be combined to provide the eigenvalues of A as functions of the field parameters. This, in turn, enables the specification of the valid parameter space via a set of necessary and sufficient conditions on the parameters. More generally, for higher order fields, we derive a recursive procedure that constructs a lower bound for the eigenvalues of the potential matrix of a pth-order field in terms of the corresponding bound for the $(p-1)$th-order field and the eigenvalues of the symmetric matrices which contain the new component of the field interactions. This bound provides a sufficient condition, for the parameter space, which is shown to always be tighter than the usual one that is derived by imposing the strict diagonal dominance condition on A.

IV. Eigenstructure of the Potential Matrix A

For a nondegenerate noncausal GMRF, the parameter space Ω where the field interaction parameters (potentials) can take values is defined by the positive definite condition on the field covariance matrix. To determine this space Ω is in general a very difficult problem. Parametrizing the covariance matrix elements explicitly is usually itself a formidable task. In contrast, the parametrization of the inverse of the covariance matrix, i.e., the potential matrix, is much simpler.

A complete definition of Ω results if we parametrize the smallest eigenvalue of A in terms of the potentials. This is possible for first-order fields and certain classes of higher order fields. For more general higher order fields, we derive a lower bound on the minimum eigenvalue of A, as we will see later.

In this section, we present our results on the eigenstructure of A, which are then used in subsequent sections to define the parameter space Ω and develop algorithms for ML estimation. For convenience, hereafter, whenever we refer to a field of given order, we index the potential matrix and its blocks accordingly, e.g., A_1 for first-order fields, A_2 for second-order fields, and so on.

A. First-Order Fields

For first-order fields, we can derive exact analytic expressions for the eigenvalues of A by applying the properties of the Kronecker product. We first discuss the case of Dirichlet fields since it is the simplest of the three introduced in Section III. Let β_{h_1} and β_{v_1} represent the horizontal and vertical nearest neighbor field interactions, respectively.

Dirichlet fields: The potential matrix is [29], [30]

$$A_1 = \begin{bmatrix} B_1 & C_1 & \underline{0} & . & . & . \\ C_1 & B_1 & C_1 & \underline{0} & . & . \\ \underline{0} & C_1 & B_1 & C_1 & \underline{0} & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & \underline{0} & C_1 & B_1 & C_1 \\ . & . & . & \underline{0} & C_1 & B_1 \end{bmatrix} = I_N \otimes B_1 + S_N \otimes C_1, \tag{15}$$

where the blocks on the main diagonal, B_1, reflect the structure

of A_1,

$$B_1 = \begin{bmatrix} 1 & -\beta_{h_1} & 0 & . & . & . \\ -\beta_{h_1} & 1 & -\beta_{h_1} & 0 & . & . \\ 0 & -\beta_{h_1} & 1 & -\beta_{h_1} & 0 & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & 0 & -\beta_{h_1} & 1 & -\beta_{h_1} \\ . & . & . & 0 & -\beta_{h_1} & 1 \end{bmatrix} = I_M - \beta_{h_1} S_M, \tag{16}$$

and C_1 is diagonal,

$$C_1 = -\beta_{v_1} I_M. \tag{17}$$

The matrices S_N, S_M, are of the form

$$S_K = \begin{bmatrix} 0 & 1 & 0 & . & . & . \\ 1 & 0 & 1 & 0 & . & . \\ 0 & 1 & 0 & 1 & 0 & . \\ . & . & . & . & . & . \\ . & . & 0 & 1 & 0 & 1 \\ . & . & . & 0 & 1 & 0 \end{bmatrix} = H_K^1, \qquad K = N, M, \tag{18}$$

where H_k^1 was defined in Section II. The eigenvalues of S_K are well known, see, for example, [18], and are provided here in the following lemma.

Lemma 1 (Dirichlet Fields), e.g., [18]: The eigenvalues of S_K are

$$\lambda_k(S_K) = 2\cos\frac{k\pi}{K+1}, \qquad 1 \le k \le K. \tag{19}$$

We now consider the other two fields.

Variational Fields: The potential matrix for these fields is a perturbation of the potential matrix for free b.c. (Dirichlet fields). See [29] for the specification. It turns out that we can cast it in the form of (15)–(17), if we redefine S_N and S_M to be as follows:

$$S_K = \begin{bmatrix} 1 & 1 & 0 & . & . & . \\ 1 & 0 & 1 & 0 & . & . \\ 0 & 1 & 0 & 1 & 0 & . \\ . & . & . & . & . & . \\ . & . & 0 & 1 & 0 & 1 \\ . & . & . & 0 & 1 & 1 \end{bmatrix} = H_K^1 + \vec{e}_1\vec{e}_1^T + \vec{e}_K\vec{e}_K^T, \qquad K = N, M. \tag{20}$$

The eigenvalues for this matrix can be derived by simple manipulation of the transform matrix for the Discrete Cosine Transform. The results are provided in the following lemma while the details of the derivation can be found in the Appendix.

Lemma 2 (Variational Field): The eigenvalues of S_K are

$$\lambda_k(S_K) = 2\cos\frac{(k-1)\pi}{K}, \qquad 1 \le k \le K. \tag{21}$$

Finally, we deal with the symmetric field.

Symmetric Fields: As in the case of variational fields, the potential matrix for these fields can be expressed in the form of (15)–(17) if we redefine S_K, $(K = N, M)$, as

$$S_K = \begin{bmatrix} 0 & 2 & 0 & . & . & . & . \\ 2 & 0 & 1 & 0 & . & . & . \\ 0 & 1 & 0 & 1 & 0 & . & . \\ . & . & . & . & . & . & . \\ . & . & 0 & 1 & 0 & 1 & 0 \\ . & . & . & 0 & 1 & 0 & 2 \\ . & . & . & . & 0 & 2 & 0 \end{bmatrix}, \qquad K = N, M. \tag{22}$$

The structure of the eigenvalues for the symmetric field S_K is more cumbersome than for the previous cases, requiring the solution of a transcendental equation as shown in Lemma 3, which is stated next, assuming $K > 5$.

Lemma 3 (Symmetric Field): The eigenvalues of S_K are

$$\lambda_k(S_K) = \begin{cases} 2\cos\theta_k & \text{when } k \ne k_1, k_2, k_3, k_4 \\ 2\cosh\theta_{k_1} & \text{when } k = k_1 \\ 2\cosh\theta_{k_2} & \text{when } k = k_2 \\ -2\cosh\theta_{k_1} & \text{when } k = k_3 \\ -2\cosh\theta_{k_2} & \text{when } k = k_4 \end{cases} \tag{23}$$

where $\{\theta_k, k \ne k_1, k_2, k_3, k_4\}$ are the $K-4$ solutions of the transcendental equation,

$$[(K+1)\theta_k - k\pi] + 2\tan^{-1}\frac{3\sin 2\theta_k}{1-3\cos 2\theta_k} = 0, \qquad 1 \le k \le K, \tag{24}$$

and θ_{k_1}, θ_{k_2}, $-\theta_{k_1}$, $-\theta_{k_2}$ are the 4 solutions of the hyperbolic equation,

$$\sinh(K+1)\theta - 6\sinh(K-1)\theta + 9\sinh(K-3)\theta = 0. \tag{25}$$

This lemma is proved by finding a recurrence relation for the characteristic polynomial of S_K (given by (22)) for arbitrary $K > 5$. This recurrence relation can be specified in terms of the characteristic polynomial of the matrix in (18) for which five different closed form solutions are possible. One of these leads to Chebyshev polynomials of the second kind, e.g., [34], and equation (24), two others lead to a trivial result, and the final two lead to (25). The detailed proof is presented in the Appendix. For each $K \le 5$, the structure of the matrix S_K and its eigenvalues is different and must be derived as a special case.

Eigenvalues of A_1: The eigenvalues given in Lemmas 1, 2, and 3 can be used in conjunction with the Kronecker product specification of A_1 in (15)–(17) to specify the eigenvalues of A_1 as functions of the field parameters.

Lemma 4 (Eigenvalues of A_1): The eigenvalues of the potential matrix A_1 for a first-order homogeneous GMRF are given by:

$$\lambda_{i,j}(A_1) = 1 - \beta_{v_1}\lambda_i(S_N) - \beta_{h_1}\lambda_j(S_M), \qquad 1 \le i \le N,\ 1 \le j \le M \tag{26}$$

where the eigenvalues, $\{\lambda_k(S_K)\}_{k=1}^K$, of S_K, for $K = M, N$, are given above by lemmas 1, 2, or 3, respectively, for the Dirichlet, variational, or symmetric fields.

The details of the proof are in the Appendix. This result relies on the fact that the potential matrix can be expressed in the form of a sum of Kronecker products where each term (i.e., each Kronecker product) has the same set of eigenvectors. Consequently, the eigenvalues are obtained by simply adding the corresponding eigenvalues of each term. In general, it is not possible to decompose A for higher order fields into a sum of Kronecker product terms such that each term has the same set of eigenvectors, therefore the argument used in the proof of the previous result does not generalize, barring special cases one of which is discussed next.

B. Second-Order Field with Symmetric Diagonal Interactions

Consider a second-order field with free b.c. and symmetric diagonal interactions, see Fig. 1, i.e., $\beta_{ld_{11}} = \beta_{rd_{11}} = \beta_{d_{11}}$, which means the left and right diagonal interactions are constrained to be identical. This is a three parameter $(\beta_{h_1}, \beta_{v_1}, \beta_{d_{11}})$ field, falling between the two parameter first-order field and the fully general four parameter $(\beta_{h_1}, \beta_{v_1}, \beta_{ld_{11}}, \beta_{rd_{11}})$ second-order field, see Fig. 1. For this field, the potential matrix can be expressed in the form of (15),

$$A_2 = I_N \otimes B_2 + S_N \otimes C_2, \tag{27}$$

with $B_2 = B_1$ given by (16), S_K by (18) and C_2 by

$$C_2 = -\beta_{v_1} I_M - \beta_{d_{11}} S_M. \tag{28}$$

Substituting the structure of B_2 and C_2 into (27), we get a sum of Kronecker product terms each having the same set of eigenvectors. Hence, the eigenvalues of A_2 are obtained by adding the corresponding ones for each term. The result is given below. The details can be found in the Appendix.

Lemma 5 (Symmetric Diagonal Interactions): The eigenvalues of the potential matrix A_2 given by (16), (18), (27), and (28) for the second-order Dirichlet field with symmetric diagonal interactions, are given by:

$$\lambda_{i,j}(A_2) = 1 - 2\beta_{v_1} \cos\frac{i\pi}{N+1} - 2\beta_{h_1} \cos\frac{j\pi}{M+1} - 4\beta_{d_{11}} \cos\frac{i\pi}{N+1} \cos\frac{j\pi}{M+1} \tag{29}$$

for $1 \le i \le N$, $1 \le j \le M$.

C. Higher Order Fields

Pairwise Interactions: We consider here Dirichlet fields of second-order or higher. The notation for the pairwise interaction potentials for fields up to order 6 is given in Fig. 1. The four types of pairwise interactions are the following.

a) *Horizontal (within same row):* These are represented by β_{h_j}, with the two interacting pixels being j columns apart.

b) *Vertical (within the same column):* These are represented by β_{v_i}, with the two interacting pixels being i rows apart.

c) *Left (NW-SE) diagonal:* These are represented by $\beta_{ld_{ij}}$, with the two interacting pixels being i rows and j columns apart.

d) *Right (NE-SW) diagonal:* These are represented by $\beta_{rd_{ij}}$, with the two interacting pixels being i rows and j columns apart.

Next, we introduce some notation. Define β_τ as the coefficient of interaction τ (e.g., $\beta_{ld_{11}}$ is the coefficient for interaction $\tau = ld_{11}$), η_p as the set of interactions for a pth-order field (e.g., $\eta_2 = \{h_1, v_1, ld_{11}, rd_{11}\}$, $\eta_3 = \{h_1, v_1, ld_{11}, rd_{11}, h_2, v_2\}$), and Γ_p as the set of new interactions in the pth-order field, i.e., $\Gamma_p = \eta_p - \eta_{p-1}$, e.g., $\Gamma_3 = \{h_2, v_2\}$.

Structure of A_p: Here, we paraphrase the discussion in [30], providing only the information relevant to the issues in this paper. See [30] for a detailed discussion of the structure of A_p. For Dirichlet fields, A_p is a block banded, block Toeplitz matrix with each block being banded and Toeplitz as well. The order of the field determines the structure of A_p. It has N rows of $M \times M$ blocks, with each block containing the interactions between two rows in the field. Thus, the blocks in the ith block row represent the interactions between row i and all the other rows in the lattice. The (i,j)th block contains the interactions between the ith and jth rows. The block banded nature of A_p is the result of the Markovianity of the field, since any row i only has nonzero interactions with rows that lie within its neighborhood structure. To make things clearer, we discuss the structure of A_p in the context of a third-order field, relating the structure back to Fig. 1. From Fig. 1, we see that for third-order fields the pixels in row i are unrelated to pixels in rows further than $i+2$ and $i-2$. Consequently, only the first two blocks on either side of the (i,i) block are nonzero, i.e., A_3 has block bandwidth 2. The symmetry of the neighborhood definition produces a corresponding symmetry in A_p. The block structure of A_3 is given next.

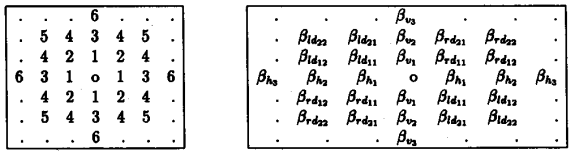
$$A_3 = \begin{bmatrix} B_3 & C_3 & D_3 & \underline{0} & . & . & . \\ C_3^T & B_3 & C_3 & D_3 & \underline{0} & . & . \\ D_3^T & C_3^T & B_3 & C_3 & D_3 & \underline{0} & . \\ \underline{0} & D_3^T & C_3^T & B_3 & C_3 & D_3 & \underline{0} \\ . & . & . & . & . & . & . \\ . & \underline{0} & D_3^T & C_3^T & B_3 & C_3 & D_3 \\ . & . & \underline{0} & D_3^T & C_3^T & B_3 & C_3 \\ . & . & . & \underline{0} & D_3^T & C_3^T & B_3 \end{bmatrix} \tag{30}$$

$$= I_M \otimes B_3 + K_N^1 \otimes C_3^T + (K_N^1)^T \otimes C_3 + K_N^2 \otimes D_3^T + (K_N^2)^T \otimes D_3. \tag{31}$$

Because of the symmetry of A_p, we only need to study the blocks that lie after the (i,i) block in any block row i. Consider the interactions between row i and row $i+2$ which is the furthest row that i is related to (see Fig. 1). The block D_3 contains the interactions between the pixels in row i and those in row $i+2$. From Fig. 1, we see that the only interaction between the pixels of these two rows is the vertical interaction represented by β_{v_2}. Consequently, D_3 is diagonal,

$$D_3 = -\beta_{v_2} I_M. \tag{32}$$

Next, we consider the interactions between rows i and $i+1$. From Fig. 1, there are three pairwise interactions for each pixel in row i. These are: the left diagonal interaction ($\beta_{ld_{11}}$),



(a) (b)

Fig. 1. (a) Hierarchical sequence of neighborhoods, (b) Coefficients for pair interactions.

the right diagonal interaction ($\beta_{rd_{11}}$), and the nearest neighbor vertical interaction (β_{v_1}). Therefore, C_3 is tridiagonal,

$$C_3 = \begin{bmatrix} -\beta_{v_1} & -\beta_{ld_{11}} & 0 & . & . & . \\ -\beta_{rd_{11}} & -\beta_{v_1} & -\beta_{ld_{11}} & 0 & . & . \\ 0 & -\beta_{rd_{11}} & -\beta_{v_1} & -\beta_{ld_{11}} & 0 & . \\ . & . & . & . & . & . \\ . & . & 0 & -\beta_{rd_{11}} & -\beta_{v_1} & -\beta_{ld_{11}} \\ . & . & . & 0 & -\beta_{rd_{11}} & -\beta_{v_1} \end{bmatrix}. \tag{33}$$

$$= -\beta_{v_1} I_M - \beta_{rd_{11}} K_M^1 - \beta_{ld_{11}} (K_M^1)^T. \tag{34}$$

Finally, we consider the self interactions of row i, i.e., the interactions within the row. From Fig. 1, we see that there are four pairwise interactions for each pixel, excluding its interaction with itself. These interactions are: two nearest neighbor horizontal interactions (β_{h_1}), and two horizontal interactions with pixels that are two columns away (β_{h_2}). Including the self interaction, B_3 is pentadiagonal,

$$B_3 = \begin{bmatrix} 1 & -\beta_{h_1} & -\beta_{h_2} & 0 & 0 & . & . \\ -\beta_{h_1} & 1 & -\beta_{h_1} & -\beta_{h_2} & 0 & . & . \\ -\beta_{h_2} & -\beta_{h_1} & 1 & -\beta_{h_1} & -\beta_{h_2} & 0 & . \\ 0 & -\beta_{h_2} & -\beta_{h_1} & 1 & -\beta_{h_1} & -\beta_{h_2} & . \\ . & . & . & . & . & . & . \\ . & 0 & -\beta_{h_2} & -\beta_{h_1} & 1 & -\beta_{h_1} & -\beta_{h_2} \\ . & . & 0 & -\beta_{h_2} & -\beta_{h_1} & 1 & -\beta_{h_1} \\ . & . & . & 0 & -\beta_{h_2} & -\beta_{h_1} & 1 \end{bmatrix}. \tag{35}$$

$$= I_M - \beta_{h_1} H_M^1 - \beta_{h_2} H_M^2. \tag{36}$$

Generalizing this argument to any order is straightforward. When we go up in order, we add new diagonals within the nonzero blocks for new interactions between the same rows (or within the same row in the case of self interaction blocks B_p), and new blocks for interactions between previously unrelated rows.

Now we are in a position to define A_p in general terms. The potential matrix, A_p, can be represented compactly as

$$A_p = I_{NM} - \sum_{\tau \in \eta_p} \beta_\tau \Delta_p^\tau, \tag{37}$$

where η_p was defined previously as the set of interactions for the pth-order field, while the matrices Δ_p^τ place the interaction parameters, β_τ, at the appropriate locations in A_p. The structure of the Δ_p^τ is explored in the following lemma:

Lemma 6: For any p, $\tau \in \eta_p$, Δ_p^τ is one of the 4 interaction matrices, I_H^j, H_i^I, $P_{i,j}$, or $P_{i,j,T}$, defined by (7), (9), (11), and (13), respectively.

The proof follows from a formalization of the discussion at the beginning of this section, in which the types of interactions are broken down into horizontal, vertical, left and right diagonal, represented by β_{h_j}, β_{v_i}, $\beta_{ld_{ij}}$, and $\beta_{rd_{ij}}$, respectively. It is straightforward to verify from (6)–(13) that the matrices required to place these parameters at the appropriate locations in A_p are respectively, I_H^j, H_i^I, $P_{i,j}$, and $P_{i,j,T}$.

Order Recursive Expression for A_p: From (37), the representation of A_p can be put in the recursive form,

$$A_p = A_{p-1} + \Delta_p, \qquad p \geq 1, \tag{38}$$

where A_{p-1} is the potential matrix for the $(i-1)$th-order field, $A_0 = I_{NM}$, and Δ_p contains the sum of the new interactions, i.e.,

$$\Delta_p = -\sum_{\tau \in \Gamma_p} \beta_\tau \Delta_p^\tau, \tag{39}$$

with Γ_p as defined earlier. A few examples follow. Using Fig. 1 and the discussion following Lemma 6, we get,

$$A_1 = A_0 + [-(\beta_{h_1} I_H^1 + \beta_{v_1} H_1^I)], \tag{40}$$

$$A_2 = A_1 + [-(\beta_{ld_{11}} P_{1,1} + \beta_{rd_{11}} P_{1,1,T})], \tag{41}$$

$$A_3 = A_2 + [-(\beta_{h_2} I_H^2 + \beta_{v_2} H_2^I)], \tag{42}$$

$$A_4 = A_3 + [-(\beta_{ld_{12}} P_{1,2} + \beta_{rd_{12}} P_{1,2,T} + \beta_{ld_{21}} P_{2,1} + \beta_{rd_{21}} P_{2,1,T})], \tag{43}$$

$$A_5 = A_4 + [-(\beta_{ld_{22}} P_{2,2} + \beta_{rd_{22}} P_{2,2,T})]. \tag{44}$$

In the next section, we will derive a lower bound for the smallest eigenvalue of A_p for arbitrary p. Construction of this bound requires knowledge of the spectral norms of the interaction matrices.

Spectral Norms of the Interaction Matrices: Let $\|.\|_s$ be the matrix spectral norm, e.g., [37], and $\lfloor . \rfloor$ the floor function. Define

$$M_j = \begin{cases} \frac{M}{j}, & \text{if } M \bmod j = 0, \\ \lfloor \frac{M}{j} \rfloor + 1, & \text{if } M \bmod j \neq 0, \end{cases} \tag{45}$$

and similarly for N_i, and

$$Q_{N,M}^{i,j} = \min(N_i, M_j). \tag{46}$$

Lemma 7: The spectral norms of the interaction matrices I_H^j, H_i^I, $P_{i,j}$, and $P_{i,j,T}$, defined in (6)–(13), are given by

$$\|I_H^j\|_s = 2\cos\frac{\pi}{M_j + 1}, \tag{47}$$

$$\|H_i^I\|_s = 2\cos\frac{\pi}{N_i + 1}, \tag{48}$$

$$\|P_{i,j}\|_s = 2\cos\frac{\pi}{Q_{N,M}^{i,j} + 1}, \tag{49}$$

$$\|P_{i,j,T}\|_s = 2\cos\frac{\pi}{Q_{N,M}^{i,j} + 1}. \tag{50}$$

The first two norms are derived by using the properties of Kronecker products and the eigenvalues of H_K^k from [15].

The other two are obtained by using graph theory to find the largest block that is obtained when the corresponding matrices are permuted into block diagonal form. A summarized version of the proof is provided in the appendix. The complete proof may be found in [3].

The results we have derived for the eigenstructure of A_p are applied in the next section to the problem of specification of the field parameter space.

V. Parameter Space for Noncausal GMRF's

The parameter space for a nondegenerate GMRF is derived from the positive definite condition on the field covariance matrix or its inverse. This translates into

$$\sigma^2 > 0,$$

which is assumed in the sequel, and the positivity of the eigenvalues of A,

$$\lambda_l(A) > 0, \qquad 1 \le l \le NM, \tag{51}$$

or equivalently

$$\lambda_{\min}(A) > 0, \tag{52}$$

where $\lambda_{\min}(A)$ is the smallest eigenvalue of A.

In Sections IV-A and IV-B, we derived the eigenstructure of A for first-order fields and second-order Dirichlet fields with symmetric diagonal interactions. Applying (52) we can then determine completely Ω for these fields. This is done in Section V-A and Section V-B. For arbitrary higher order Dirichlet fields, we derive a lower bound for $\lambda_{\min}$ that serves as a sufficient condition for the valid parameter space. This sufficient condition is shown to be tighter than the one derived by imposing the strict diagonal dominance sufficient condition for $A_p = [a_{ij}^p] > 0$, see, for example, [37], which corresponds to

$$S_p = \min_i \{1 - \sum_{j \ne i} |a_{ij}^p|\} > 0, \tag{53}$$

where we have simplified the definition by using the fact that the potential matrix for a Dirichlet field has a unit main diagonal.

A. First-Order Fields

We use Lemma 4 to specify the parameter space for each of the first-order fields considered in Section IV-A. The parameter space is derived in each case by considering the form of $\lambda_{\min}(A_1)$ for each of the four possible combinations of $\pm$ signs of the pair of coefficients $(\beta_{h_1}, \beta_{v_1})$.

Result 1 (Dirichlet Field): The valid parameter space for the first-order Dirichlet field is defined by

$$|\beta_{v_1}| \cos\frac{\pi}{N+1} + |\beta_{h_1}| \cos\frac{\pi}{M+1} < \frac{1}{2}. \tag{54}$$

The necessary and sufficient conditions in (54) correspond to a region in $(\beta_{h_1}, \beta_{v_1})$ space bounded by four lines which intersect with the β_{h_1} and β_{v_1} axes at points

$$\left\{(\frac{1}{2\cos\frac{\pi}{M+1}}, 0), (0, \frac{1}{2\cos\frac{\pi}{N+1}}), (-\frac{1}{2\cos\frac{\pi}{M+1}}, 0), (0, -\frac{1}{2\cos\frac{\pi}{N+1}})\right\},$$

see Fig. 2.

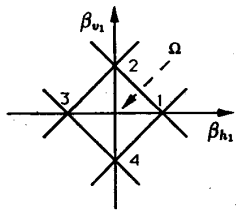


Fig. 2. Parameter space for first-order fields is the region bounded by the four lines whose $(\beta_{h_1}, \beta_{v_1})$ axes intercepts are at points 1–4.

Result 2 (Variational Field): The parameter space for the first-order variational field is defined by the set of inequalities:

$$\beta_{v_1} + \beta_{h_1} < \frac{1}{2}, \tag{55}$$

$$\beta_{v_1} - \beta_{h_1} \cos\frac{\pi}{M} < \frac{1}{2}, \tag{56}$$

$$\beta_{h_1} - \beta_{v_1} \cos\frac{\pi}{N} < \frac{1}{2}, \tag{57}$$

$$-\beta_{v_1} \cos\frac{\pi}{N} - \beta_{h_1} \cos\frac{\pi}{M} < \frac{1}{2}. \tag{58}$$

The parameter space is bounded by four lines that intersect the axes at

$$\left\{(\frac{1}{2}, 0), (0, \frac{1}{2}), (-\frac{1}{2\cos\frac{\pi}{M}}, 0), (0, -\frac{1}{2\cos\frac{\pi}{N}})\right\},$$

see Fig. 2.

Result 3 (Symmetric Field): The parameter space for the first-order symmetric field is defined by

$$|\beta_{v_1}| \cosh\theta_N^* + |\beta_{h_1}| \cosh\theta_M^* < \frac{1}{2}, \tag{59}$$

with

$$\theta_K^* = \max\{\theta_{k1}, \theta_{k2}\}, \tag{60}$$

where $\pm\theta_{k1}$, $\pm\theta_{k2}$ are the four solutions of the hyperbolic equation (25).

For symmetric fields, the corresponding intercepts are

$$\left\{(\frac{1}{2\cosh\theta_M^*}, 0), (0, \frac{1}{2\cosh\theta_N^*}), (-\frac{1}{2\cosh\theta_M^*}, 0), (0, -\frac{1}{2\cosh\theta_N^*})\right\}.$$

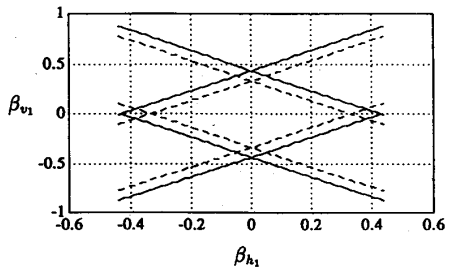


Fig. 3. Comparison of actual parameter space for first-order symmetric fields (region enclosed by the four solid lines) versus region defined by strict diagonal dominance sufficient condition (region enclosed by the four dashed lines) for lattice size $N = M = 16$.

Strict Diagonal Dominance Sufficient Condition: The strict diagonal dominance sufficient condition for Dirichlet and variational fields is

$$|\beta_{v_1}| + |\beta_{h_1}| < \frac{1}{2}. \tag{61}$$

This is derived in a straightforward manner from the structure of the corresponding potential matrices, from (15)–(17) and respectively, (18) and (20). The region defined by (61) is always subsumed (assuming finite N, M) by the regions defined respectively by (54) and (55)–(58) for the Dirichlet and variational fields.

In the case of symmetric fields, the strict diagonal dominance condition is not (61) but instead

$$|\beta_{v_1}| + |\beta_{h_1}| < \frac{1}{3}. \tag{62}$$

For these fields, a considerable portion of the parameter space is excluded by the sufficient condition in (62). The difference between the region defined by (59) and the sufficient space defined by (62) is illustrated by an example. Consider a first-order symmetric field defined on a 16×16 lattice, i.e., $N = M = 16$. For these values of N and M, equation (25) was solved numerically and θ_{16}^{*} in (60) was computed to be 2.3090. The parameter space defined by (59) is plotted in Fig. 3 as the region in $(\beta_{h_1}, \beta_{v_1})$ space enclosed by the four solid lines which are obtained by replacing the inequalities in (59) with equalities. The corresponding sufficient condition space is plotted in the same figure as the region enclosed by the four dashed lines that are obtained by replacing the inequality in (62) by an equality. The plots in Fig. 3 demonstrate that a significant portion of the parameter space is excluded by the strict diagonal dominance sufficient condition. Hence, the results presented above are of great importance to parameter estimation procedures because they allow the optimization algorithm the freedom of the entire valid space rather than just the portion of it defined by the sufficient conditions.

B. Second-Order Field with Symmetric Diagonal Interactions

The necessary and sufficient conditions for the valid parameter space for the second-order Dirichlet field with symmetric diagonal interactions are obtained from Lemma 5. The eigenvalues of A_2 form a set of planes in the 3-D space $(\beta_{v_1}, \beta_{h_1}, \beta_{d_{11}})$, with each plane having only two independent intercepts. The minimum set of planes that bound the valid region in the 3-D space are obtained by applying the basic principles of 3-D coordinate geometry.

Result 4: The parameter space for the second-order Dirichlet field with symmetric diagonal interactions is defined by the set of inequalities:

$$\beta_{v_1} \cos\frac{\pi}{N+1} + \beta_{h_1} \cos\frac{\pi}{M+1} + 2\beta_{d_{11}} \cos\frac{\pi}{N+1} \cos\frac{\pi}{M+1} < \frac{1}{2}, \tag{63}$$

$$\beta_{v_1} \cos\frac{\pi}{N+1} - \beta_{h_1} \cos\frac{\pi}{M+1} - 2\beta_{d_{11}} \cos\frac{\pi}{N+1} \cos\frac{\pi}{M+1} < \frac{1}{2}, \tag{64}$$

$$\beta_{h_1} \cos\frac{\pi}{M+1} - \beta_{v_1} \cos\frac{\pi}{N+1} - 2\beta_{d_{11}} \cos\frac{\pi}{N+1} \cos\frac{\pi}{M+1} < \frac{1}{2}, \tag{65}$$

$$-\beta_{v_1} \cos\frac{\pi}{N+1} - \beta_{h_1} \cos\frac{\pi}{M+1} + 2\beta_{d_{11}} \cos\frac{\pi}{N+1} \cos\frac{\pi}{M+1} < \frac{1}{2}. \tag{66}$$

In comparison, the strict diagonal dominance sufficient condition is

$$|\beta_{h_1}| + |\beta_{v_1}| + 2|\beta_{d_{11}}| < \frac{1}{2}, \tag{67}$$

which excludes a considerable portion of the valid parameter space. For example, consider $N = M = 32$, $\beta_{h_1} = \beta_{v_1} = 0.3$, $\beta_{d_{11}} = -0.25$. For these parameter values, the left-hand side of (67) evaluates to 0.9 which is well outside the region defined by the condition in (67), i.e., the point $(0.3, 0.3, -0.25)$ is excluded by the sufficient condition. On the other hand, substituting these values into equations (63)–(66) we see that this point is in fact inside the valid parameter space.

C. Higher Order Fields

In this subsection, we consider second or higher order Dirichlet fields. For arbitrary higher order fields, we do not have the eigenstructure of A_p, or $\lambda_{\min}(A_p)$. We use, however, the structure of A_p to derive a sufficient condition for the parameter space Ω. The main result of this section is a recursive specification for a sufficient condition, L_p, that is always tighter than the corresponding strict diagonal dominance condition, S_p. Our sufficient condition is derived by using the recursive specification of A_p given in (38) and (39) to bound $\lambda_{\min}(A_p)$ from below, i.e.,

$$\lambda_{\min}(A_p) \geq L_p, \tag{68}$$

whereupon

$$L_p > 0 \tag{69}$$

becomes a sufficient condition for the parameter space of A_p.

The strict diagonal dominance condition can be represented compactly as follows. Let $\|.\|_\infty$ be the matrix infinity norm, e.g., [37].

Lemma 8: The strict diagonal dominance sufficient condition for a pth-order Dirichlet field is given by

$$S_p > 0, \tag{70}$$

where

$$S_p = S_{p-1} - \sum_{\tau \in \Gamma_p} |\beta_\tau| \|\Delta_p^\tau\|_\infty, \qquad p \geq 1, S_0 = 1, \tag{71}$$

and Γ_p and Δ_p^τ are as defined in Section IV-C.

This is a straightforward relationship that comes from the discussion in Section IV-C. Recall that in the decomposition of A_p given in (37) each Δ_p^τ places the coefficients for only the interaction τ in the appropriate locations in A_p. Consequently, for any two different interactions, τ_1 and τ_2, the corresponding matrices $\Delta_p^{\tau_1}$, and $\Delta_p^{\tau_1}$ have no nonzero elements in overlapping locations. Keeping this in mind, we apply the definition of S_p, from (53), to (38) and (39) to get the above result. Barring degenerate cases, corresponding to lattices that are so small that there is no site with the full complement of neighbors, all the matrices Δ_p^τ have at least one row (column) with two 1's in it, which is the most they can have, and (71) simplifies to

$$S_p = S_{p-1} - 2 \sum_{\tau \in \Gamma_p} |\beta_\tau|, \tag{72}$$

or

$$S_p = 1 - 2 \sum_{\tau \in \eta_p} |\beta_\tau|, \tag{73}$$

which is the usual form this condition appears in, in the literature. In the sequel, when we discuss the strict diagonal dominance sufficient condition for a field of given order, we assume that the lattice size is above this threshold. We now establish a similar recursion for the lower bound L_p for arbitrary order p.

Result 5: The sufficient condition, L_p, for the pth-order Dirichlet field is given by

$$L_p = L_{p-1} - \sum_{\tau \in \Gamma_p} |\beta_\tau| \|\Delta_p^\tau\|_s, \qquad p \geq 1, L_0 = 1, \tag{74}$$

where L_{p-1} is the sufficient condition for the $(p-1)$th-order field and Γ_p and Δ_p^τ are as defined in Section IV-C.

This result is derived by using the property [25, theorem 5, p. 205] that the smallest eigenvalue of the sum of symmetric matrices is always greater than or equal to the sum of the smallest eigenvalues of each of these matrices. This property is applied to the recursive specification in (38) and (39) to produce the desired result. More details are provided in the Appendix. For the sake of concreteness, we apply (40)–(44), (74), and Lemma 7 to get the following results up to fifth-order fields.

Result 6: The sufficient conditions, L_p for $p = 1, 2, 3, 4, 5$ are given by

$$L_1 = 1 - 2[|\beta_{h_1}| \cos \frac{\pi}{M+1} + |\beta_{v_1}| \cos \frac{\pi}{N+1}], \tag{75}$$

$$L_2 = L_1 - 2[(|\beta_{ld_{11}}| + |\beta_{rd_{11}}|) \cos \frac{\pi}{Q_{N,M}^{1,1} + 1}], \tag{76}$$

$$L_3 = L_2 - 2[|\beta_{h_2}| \cos \frac{\pi}{M_2+1} + |\beta_{v_2}| \cos \frac{\pi}{N_2+1}], \tag{77}$$

$$L_4 = L_3 - 2[(|\beta_{ld_{12}}| + |\beta_{rd_{12}}|) \cos \frac{\pi}{Q_{N,M}^{1,2} + 1} + (|\beta_{ld_{21}}| + |\beta_{rd_{21}}|) \cos \frac{\pi}{Q_{N,M}^{2,1} + 1}], \tag{78}$$

$$L_5 = L_4 - 2[(|\beta_{ld_{22}}| + |\beta_{rd_{22}}|) \cos \frac{\pi}{Q_{N,M}^{2,2} + 1}]. \tag{79}$$

□

Note that L_1 is in fact necessary and sufficient. Quantities M_j, N_i and $Q_{N,M}^{i,j}$ were introduced earlier in Section IV-C. The corresponding strict diagonal dominance sufficient conditions for the same fields are:

$$S_1 = 1 - 2[|\beta_{h_1}| + |\beta_{v_1}|], \tag{80}$$

$$S_2 = S_1 - 2[|\beta_{ld_{11}}| + |\beta_{rd_{11}}|], \tag{81}$$

$$S_3 = S_2 - 2[|\beta_{h_2}| + |\beta_{v_2}|], \tag{82}$$

$$S_4 = S_3 - 2[|\beta_{ld_{12}}| + |\beta_{rd_{12}}| + |\beta_{ld_{21}}| + |\beta_{rd_{21}}|], \tag{83}$$

$$S_5 = S_4 - 2[|\beta_{ld_{22}}| + |\beta_{rd_{22}}|]. \tag{84}$$

It is straightforward to verify that S_1 to S_5 are strictly smaller than their counterparts L_1 to L_5. This is always the case, as stated in the following result.

Result 7: The sufficient condition (69) with L_p obtained from (74) is always tighter than the corresponding one, in (70), (71), obtained by imposing the strict diagonal dominance condition on A_p, i.e.,

$$\forall p \geq 1, \quad L_p > S_p. \tag{85}$$

This is proved by comparing (74) and (71). The details are provided in the Appendix.

VI. Maximum Likelihood Parameter Estimation

In most applications, the field potentials, $\beta_{h_1}, \beta_{v_1}, \cdots$, are unknown and have to be estimated. Let θ be the vector containing the field potentials. These are the parameters of the potential matrix, $A_p(\theta)$. The joint probability density function (pdf) of a GMRF, defined on a $N \times M$ lattice, is the multivariate Gauss,

$$P(X) = \frac{1}{(2\pi\sigma^2)^{NM/2}} |A_p(\theta)|^{1/2} exp\{-\frac{1}{2\sigma^2} \vec{X}^T A_p(\theta) \vec{X}\}. \tag{86}$$

The negative log likelihood function, scaled by $(1/NM)$ for convenience, is given by

$$L(X/\theta, \sigma^2) = \frac{1}{2} \ln \sigma^2 - \frac{1}{2NM} \cdot \ln |A_p(\theta)| + \frac{1}{2\sigma^2 NM} \vec{X}^T A_p(\theta) \vec{X}. \tag{87}$$

The ML estimates for θ and σ^2 are obtained by minimizing the negative log likelihood function previously defined.

ML estimation for noncausal fields is complicated by several factors. A major problem is the prohibitive cost of computing the likelihood function or some scaled monotonic function of

it, e.g., (87). This is a direct consequence of the complexity of $|A_p(\theta)|^{-\frac{1}{2}}$ which is, apart from a factor, the partition function from Statistical Mechanics. In addition, the function gradient which is central to an efficient search of the parameter space is usually expensive to compute. A third complication is the fact that constrained optimization is required because of the restrictions on the parameter space.

In this section, we present an ML parameter estimation procedure for noncausal finite lattice GMRF's of *arbitrary order* that uses the recursive structure of these fields [30] to provide computationally practical solutions to the above mentioned problems. For the sake of clarity, the procedure is discussed in the context of Dirichlet fields, but extension to other fields is simply a matter of applying the corresponding field structure from [30]. As a prelude to the parameter estimation, we comment on the convexity of the function defined in (87).

Convexity of $L(.)$: It is straightforward to establish the convexity of $L(.)$ over the valid parameter space of $(\theta, \frac{1}{\sigma^2})$ by inspection of the three terms in (87). The convexity of the middle term,

$$-\frac{1}{2NM}\ln|A_p(\theta)|,$$

follows from the fact that the function $(-\log|A|)$ is convex over the space of positive definite matrices, see, for example, [25, p. 222]. The other two terms are immediately identifiable as convex functions. Since ML estimates commute with nonlinear functions, the convexity of $L(.)$ over the valid parameter space of $(\theta, \frac{1}{\sigma^2})$ also ensures that optimization in (θ, σ^2) is unhindered by the problem of local extrema.

Estimation Procedure: Define

$$\theta = [\beta_\tau, \tau \in \eta_p], \tag{88}$$

where η_p was defined in Section IV-C as the set of neighborhood interactions for a pth-order field. Using the structure of A_p for Dirichlet fields, from (37), in (87), we get,

$$L(X/\theta,\sigma^2) = \frac{1}{2}\ln\sigma^2 - \frac{1}{2NM}\ln|A_p(\theta)| + \frac{1}{2\sigma^2}[S_x - 2\sum_{\tau\in\eta_p}\beta_\tau X_\tau], \tag{89}$$

where S_x and X_τ are defined as

$$S_x = \frac{1}{NM}\vec{X}^T\vec{X} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}x_{i,j}^2 \tag{90}$$

$$X_\tau = \frac{1}{2NM}\vec{X}^T\Delta_p^\tau\vec{X} = \begin{cases} \frac{1}{NM}\sum_{r=1}^{N-i}\sum_{s=1}^{M} x_{r,s}x_{r+i,s}, & \text{for } \tau = v_i, \\ \frac{1}{NM}\sum_{r=1}^{N}\sum_{s=1}^{M-j} x_{r,s}x_{r,s+j}, & \text{for } \tau = h_j, \\ \frac{1}{NM}\sum_{r=1}^{N-i}\sum_{s=1}^{M-j} x_{r,s}x_{r+i,s+j}, & \text{for } \tau = ld_{ij}, \\ \frac{1}{NM}\sum_{r=1}^{N-i}\sum_{s=j+1}^{M} x_{r,s}x_{r+i,s-j}, & \text{for } \tau = rd_{ij}. \end{cases} \tag{91}$$

The quantity S_x is interpreted as the average power in the given field sample, while X_τ may be interpreted as the sample correlation of lag τ.

The gradient of the function in (89) is given by

$$\frac{\partial L}{\partial\sigma^2} = \frac{1}{2\sigma^2} - \frac{1}{2(\sigma^2)^2}[S_x - 2\sum_{\tau\in\eta_p}\beta_\tau X_\tau], \tag{92}$$

and for each interaction τ in the set η_p,

$$\frac{\partial L}{\partial\beta_\tau} = \frac{1}{2NM}\text{trace}\,(A_p^{-1}\Delta_p^\tau) - \frac{X_\tau}{\sigma^2}, \tag{93}$$

where the first term in (93) comes from

$$\begin{aligned}\frac{\partial(\ln|A_p|)}{\partial\beta_\tau} &= \text{trace}\left(\frac{\partial(\ln|A_p|)}{\partial A_p}\frac{\partial A_p}{\partial\beta_\tau}\right) \\ &\quad \text{(applying the chain rule)}, \\ &= \text{trace}\,(A_p^{-1}\frac{\partial A_p}{\partial\beta_\tau}) \text{(from [14, p. 75])}, \\ &= \text{trace}\,(-A_p^{-1}\Delta_p^\tau) \\ &\quad \text{(taking the partial of (37))}.\end{aligned} \tag{94}$$

Only (92) can be solved explicitly for the ML estimate of σ^2. Setting (92) to zero we get

$$\widehat{\sigma^2} = S_x - 2\sum_{\tau\in\eta_p}\widehat{\beta_\tau}X_\tau, \tag{95}$$

which, when substituted into (89), produces the function (after we drop the $\hat{}$),

$$L(X/\theta,\sigma^2(\theta)) = \frac{1}{2}\ln[S_x - 2\sum_{\tau\in\eta_p}\beta_\tau X_\tau] - \frac{1}{2NM}\ln|A_p(\theta)| + \frac{1}{2}, \tag{96}$$

which is minimized over the valid space of θ to get the ML estimates.

The ML estimation procedure presented here is based on the (Polak–Ribiere) conjugate gradient search method, which is a well studied unconstrained optimization procedure, see, for example, [33]. The parameter estimates are obtained by an iterative process that consists of a series of 1-D line minimizations along a sequence of conjugate directions generated using the function gradient, with each line minimization being preceded by a procedure that brackets the line minima. We will now discuss the problems mentioned earlier, namely, the computational burden imposed by repeated evaluations of the function and its gradient and the complications induced by the constraints on the parameter space, and outline our solutions to these.

1) Partition Function: Computation of the partition function is required for every evaluation of $L(.)$. Direct computation of $|A_p(\theta)|$ when the eigenvalues of A_p are unavailable, as is usually the case for higher order fields, is extremely costly since A_p is $NM \times NM$. A computationally practical alternative is provided by the recursive framework in [30] which includes the computation of the Cholesky factor of A_p via a matrix Riccati iteration that converges at a geometric rate. As a result, the determinant of A_p can be computed expeditely, to any specified degree of precision, as a function of the determinants of the transient iterates and the steady state solution of the Riccati equation.

2) Parameter Space Constraints: The constrained optimization problem of locating the parameter estimates within the valid section of θ space can be converted into an unconstrained problem by imposing the parameter space constraints in the bracketing procedure that precedes each line minimization. In other words, the bracketing procedure is constrained to place the bracketing points within the valid parameter space. One way of doing this is to assign a large penalty term to the function whenever it is evaluated outside the valid region, see [36]. Since the line minimization algorithm only searches the portion of the line between the bracketing points, it can continue to operate as an unconstrained optimization procedure. The exact bounds on the parameter space are in general unknown for higher order fields. Therefore, we use instead the fundamental definition of the parameter space as the region of parameter values for which A_p is positive definite (pd). The positive definiteness of A_p is automatically checked during the computation of $\ln|A_p(\theta)|$ because the Riccati equation previously mentioned diverges if $A_p(\theta)$ is not pd. The estimation procedure can be speeded up by restricting the search to the space defined by the sufficient conditions derived in Section V-C. Here, the tighter bounds derived by us enable more of the valid parameter space to be included in the search area than the previously available strict diagonal dominance sufficient condition bounds, see results 5 and 7 in Section V-C.

3) Function Gradient: The gradient of $L(.)$ is required to guide the iterative search for the optimal parameter estimates. The expression given in (93) for the gradient is computationally intractable because it involves the inversion of $A_p(\theta)$, which is an $NM \times NM$ matrix. An alternate approach to computing the gradient at a given point $(\theta^{(i)}, \sigma^{2(i)})$ in the parameter space is obtained next by interpreting (93) as a statistical quantity. The following lemma is derived from (93) by using the structure of A_p and the fact that A_p^{-1} is a scaled version of the field covariance matrix.

Lemma 9: For each interaction τ in the set η_p,

$$\frac{\partial L}{\partial \beta_\tau} = \frac{E\{X_\tau\} - X_\tau}{\sigma^2}, \tag{97}$$

where $E\{.\}$ is the expectation operator.

The details of the proof are in the Appendix. The form of (97) illustrates the fact that the ML estimates, located at the extremum of $L(.)$, correspond to the parameter values that match the theoretical covariance of the field to the sample covariance.

4) Computation of $E\{X_\tau\}$: The ML estimation procedure is an iterative search of the parameter space that is propagated by the present estimates $(\theta^{(i)}, \sigma^{2(i)})$ being updated by a step whose direction and magnitude are determined by the gradient at the present location, i.e., $(\theta^{(i)}, \sigma^{2(i)})$, in the parameter space. From Lemma 9, evaluation of the gradient requires the computation of $E\{X_\tau\}$, the ensemble average of the sample correlation of lag τ. We compute this quantity by using a Monte Carlo procedure that generates a large number (K) of samples of the field with parameters $(\theta^{(i)}, \sigma^{2(i)})$ and estimates $E\{X_\tau\}$ from these samples as

$$\hat{E}\{X_\tau\} = \frac{1}{K}\sum_{k=1}^{K} X_\tau^{(k)}, \tag{98}$$

where $X_\tau^{(k)}$ is the correlation of lag τ computed from field sample # k. The generation of field samples is computationally intensive if relaxation methods are applied. We use instead the equivalent one–sided (spatially varying) representation for noncausal fields [30] which provides the means for fast synthesis of samples of any noncausal finite lattice GMRF.

Summary: To summarize, the ML estimation algorithm for fields of arbitrary order is an iterative search of the valid parameter space using the well-known (Polak–Ribiere) conjugate gradient search method that consists of a series of 1-D line minimizations along a sequence of conjugate directions obtained from the gradient. The constraints on the field parameter space are incorporated into the bracketing procedure that precedes each line minimization. The gradient at any point $(\theta^{(i)}, \sigma^{2(i)})$ is obtained by sampling the field for those parameter values and applying Lemma 9. This is a novel approach which is made possible in a practical sense by the recursive framework in [30] that enables rapid synthesis of field samples. Experimental verification is provided in Section IX.

The estimation procedure presented in this section made no assumptions regarding the eigenstructure of the potential matrix. In the case of fields where this information is available, the additional knowledge can be used to simplify the parameter estimation as well as reduce the computational cost. This is discussed in the next section.

VII. ML Estimation for First-Order and Special Second-Order Fields

In Sections IV-A and IV-B, respectively, we derived the eigenstructure of A_p for first-order fields and second-order Dirichlet fields with symmetric diagonal interactions. In this section, we show how this information enables the parametrization of the likelihood function and its gradient in terms of the field parameters, leading to simpler, faster, and more direct means of computing these quantities than in the general case discussed in the previous section. In addition, since exact bounds on the parameter space were derived in Sections V–A and V–B, the constrained optimization is easily handled.

First-Order Fields: In [7], the likelihood function for isotropic $(\beta_{h_1} = \beta_{v_1})$ toroidal fields was expressed in terms of the field parameters by applying the known eigenvalues of the block circulant matrix A_1. Here, we take a more general approach to handle nonisotropic fields with nonperiodic b.c.

For a first-order field,

$$\theta = [\beta_{h_1}, \beta_{v_1}]. \tag{99}$$

From (26), the determinant of A_1 is

$$|A_1| = \prod_{i=1}^{N}\prod_{j=1}^{M}(1 - \beta_{v_1}\lambda_i(S_N) - \beta_{h_1}\lambda_j(S_M)). \tag{100}$$

The quadratic, $\vec{X}^T A_1 \vec{X}$, can be written using the structure of A_1 from (15)–(17) as

$$\frac{1}{NM}\vec{X}^T A_1 \vec{X} = S_x - 2\beta_{h_1} X_{h_1} - 2\beta_{v_1} X_{v_1} \tag{101}$$

where S_x is the average power of the sample, defined in (90), and

$$X_{h_1} = \frac{1}{2NM}\vec{X}^T (I_N \otimes S_M)\vec{X} = \frac{1}{NM}\left(\sum_{i=1}^{N}\sum_{j=1}^{M-1} x_{i,j}x_{i,j+1} + X_{h_1}^b\right) \tag{102}$$

$$X_{v_1} = \frac{1}{2NM}\vec{X}^T (S_N \otimes I_M)\vec{X} = \frac{1}{NM}\left(\sum_{i=1}^{N-1}\sum_{j=1}^{M} x_{i,j}x_{i+1,j} + X_{v_1}^b\right), \tag{103}$$

with $X_{h_1}^b$ and $X_{v_1}^b$ representing correction terms that depend on the choice of b.c. For the fields with S_K given by (18), (20), and (22),

$$X_{h_1}^b = \begin{cases} 0, & \text{for Dirichlet fields} \\ \frac{1}{2}\sum_{i=1}^{N}(x_{i,1}^2 + x_{i,M}^2), & \text{for variational fields} \\ \sum_{i=1}^{N}(x_{i,1}x_{i,2} + x_{i,M-1}x_{i,M}), & \text{for symmetric fields} \end{cases} \tag{104}$$

$$X_{v_1}^b = \begin{cases} 0, & \text{for Dirichlet fields} \\ \frac{1}{2}\sum_{j=1}^{M}(x_{1,j}^2 + x_{N,j}^2), & \text{for variational fields} \\ \sum_{j=1}^{M}(x_{1,j}x_{2,j} + x_{N-1,j}x_{N,j}), & \text{for symmetric fields} \end{cases} \tag{105}$$

The quantities X_{h_1} and X_{v_1} may be interpreted, respectively, as the horizontal and vertical nearest-neighbor correlations in the field sample. In this context, the boundary condition dependent portion of X_{h_1} and X_{v_1} can be interpreted as adding the correlations due to pixel pairs that are produced by the choice of boundary conditions. Thus, for example, in the case of free b.c. there is no correction term because the off-lattice neighbors of the boundary pixels are assumed to be zero, hence the correlations due to these pairs are zero.

Substituting (99)–(101) into (87) we get,

$$L(X/\beta_{h_1},\beta_{v_1},\sigma^2) = \frac{1}{2}\ln\sigma^2 - \frac{1}{2NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\ln(1-\beta_{v_1}\lambda_i(S_N) - \beta_{h_1}\lambda_j(S_M)) + \frac{1}{2\sigma^2}[S_x - 2\beta_{h_1}X_{h_1} - 2\beta_{v_1}X_{v_1}]. \tag{106}$$

As an illustration, the log likelihood function, i.e., the negative of $L(.)$ without the scaling, for a 16 × 16 Dirichlet field with parameters $\beta_{h_1} = 0.2$, $\beta_{v_1} = 0.15$, and $\sigma^2 = 10$, is plotted in $(\beta_{h_1},\beta_{v_1})$ space in Fig. 4. As expected from the discussion at the beginning of Section VI, this is a concave function with a well-defined global maximum. The minimum of the negative log likelihood function is obtained by taking the partial derivative with respect to each of the parameters and setting these to zero:

$$\frac{\partial L}{\partial\beta_{h_1}}\Big|_{\widehat{\beta_{h_1}},\widehat{\beta_{v_1}},\widehat{\sigma^2}} = \frac{1}{2NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{\lambda_j(S_M)}{(1-\widehat{\beta_{v_1}}\lambda_i(S_N) - \widehat{\beta_{h_1}}\lambda_j(S_M))} - \frac{X_{h_1}}{\widehat{\sigma^2}} = 0, \tag{107}$$

$$\frac{\partial L}{\partial\beta_{v_1}}\Big|_{\widehat{\beta_{h_1}},\widehat{\beta_{v_1}},\widehat{\sigma^2}} = \frac{1}{2NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\frac{\lambda_i(S_N)}{(1-\widehat{\beta_{v_1}}\lambda_i(S_N) - \widehat{\beta_{h_1}}\lambda_j(S_M))} - \frac{X_{v_1}}{\widehat{\sigma^2}} = 0, \tag{108}$$

$$\frac{\partial L}{\partial\sigma^2}\Big|_{\widehat{\beta_{h_1}},\widehat{\beta_{v_1}},\widehat{\sigma^2}} = \frac{1}{2\widehat{\sigma^2}} - \frac{1}{2(\widehat{\sigma^2})^2}[S_x - 2\widehat{\beta_{h_1}}X_{h_1} - 2\widehat{\beta_{v_1}}X_{v_1}] = 0, \tag{109}$$

where only (109) has an explicit solution

$$\widehat{\sigma^2} = [S_x - 2\widehat{\beta_{h_1}}X_{h_1} - 2\widehat{\beta_{v_1}}X_{v_1}]. \tag{110}$$

We substitute σ^2 from (110) (dropping the $\widehat{\ }$) in $L(.)$ in (106), to get the function

$$L(X/\theta,\sigma^2(\theta)) = \frac{1}{2}\ln(S_x - 2\beta_{h_1}X_{h_1} - 2\beta_{v_1}X_{v_1}) - \frac{1}{2NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\ln(1-\beta_{v_1}\lambda_i(S_N) - \beta_{h_1}\lambda_j(S_M)) + \frac{1}{2}, \tag{111}$$

which is minimized over the space of θ to get the ML estimates.

Second-Order Field with Symmetric Diagonal Interactions: The negative log likelihood function for the second-order Dirichlet field with diagonal interactions is derived by substituting (16), (18), (27)–(29) into (87):

$$L(X/\theta,\sigma^2) = -\frac{1}{2NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\ln(1 - 2\beta_{v_1}\cos\frac{i\pi}{N+1} - 2\beta_{h_1}\cos\frac{j\pi}{M+1} - 4\beta_{d_{11}}\cos\frac{i\pi}{N+1}\cos\frac{j\pi}{M+1}) + \frac{1}{2}\ln\sigma^2 + \frac{1}{2\sigma^2}[S_x - 2\beta_{h_1}X_{h_1} - 2\beta_{v_1}X_{v_1} - 2\beta_{d_{11}}X_{d_{11}}], \tag{112}$$

where S_x, X_{h_1} and X_{v_1} are defined similarly to (90), (102), and (103), respectively, with boundary correction terms $X_{h_1}^b$ and $X_{v_1}^b$ being zero for the Dirichlet field and

$$X_{d_{11}} = \frac{1}{2NM}\vec{X}^T(S_N \otimes S_M)\vec{X} = \frac{1}{NM}\left(\sum_{i=1}^{N-1}\sum_{j=1}^{M-1} x_{i,j}x_{i+1,j+1} + \sum_{i=1}^{N-1}\sum_{j=2}^{M} x_{i,j}x_{i+1,j-1}\right). \tag{113}$$

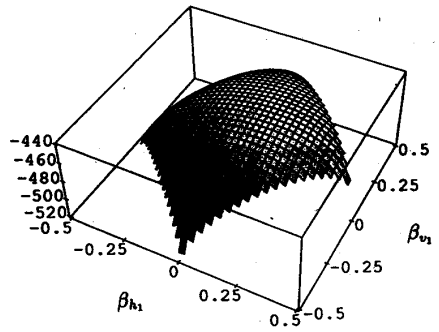


Fig. 4. Log likelihood function for first-order Dirichlet field with parameters $\beta_{h_1} = 0.2$, $\beta_{v_1} = 0.15$, $\sigma^2 = 10$, $N = M = 16$, plotted as a function of $(\beta_{h_1}, \beta_{v_1})$.

As was done for first-order fields, the gradient is derived, providing an explicit solution for $\widehat{\sigma^2}$ which is substituted into (112) to obtain a function of θ alone, analogous to (111) for first-order fields. Since no additional insight is provided by the details, we leave them out.

Estimation Procedure: The Polak–Ribiere conjugate gradient search method, e.g., [33], that was discussed in Section VI is applied directly to this problem. As explained in the previous section, this search procedure consists of a series of 1-D line minimizations along a sequence of conjugate directions generated using the function gradient, with each line minimization being preceded by a procedure that brackets the line minima. In contrast to the case for fields of arbitrary order, discussed previously, the parametrization of $L(.)$ and the gradient explicitly in terms of the field parameters provides the means for fast computation of the ML estimates, the only remaining problem being the constraints on the parameter space. This is taken care of by using the necessary and sufficient conditions for the parameter space, respectively, (54), (55)–(58), and (59) for first-order Dirichlet, variational, and symmetric fields, and (63)–(66) for the second-order Dirichlet field with symmetric diagonal interactions, in the bracketing procedure to provide bracketing points that lie on the portion of the line enclosed by the parameter space boundary. As mentioned earlier, the line minimization algorithm only searches the portion of the line between the bracketing points, hence the above modification to the bracketing procedure enables this to be solved as an unconstrained minimization problem. Experimental results are provided in Section IX.

VIII. Parameter Estimation in the Presence of Noise

In many applications, the field sample is corrupted by noise. The corresponding ML procedure for estimation of the parameters from noisy data is complicated by the fact that the inverse covariance of the noisy field no longer has the sparse structure of the noiseless field. In the case of toroidal fields, authors have exploited the special structure of the spectral density function to construct ML procedures for estimation of the parameters in the presence of additive white Gaussian noise, see [6], [19]. For nonperiodic fields, we now present a practical method that is similar in philosophy to the approach used in [21] for least squares estimation of the parameters of a causal MRF in the presence of noise. The method assumes the noise variance σ_n^2 is known. In most applications, σ_n^2 would have to be estimated, which is in general a nontrivial problem. An exception is in image processing in which it is very often straightforward to obtain an accurate estimate of σ_n^2 by preprocessing a flat portion of the observed image, see [2], [9].

Consider the case where the field sample X is corrupted by an independent, additive, white Gaussian field W, i.e., the observations are:

$$y_{i,j} = x_{i,j} + w_{i,j}, \quad w_{i,j} \sim \mathcal{N}(0, \sigma_n^2), \quad 1 \le i \le N, \ 1 \le j \le M, \tag{114}$$

where the noise $w_{i,j}$ is white, i.e., $w_{i,j} \perp w_{l,k}$, $\forall (i,j) \neq (l,k)$, and uncorrelated with the field, i.e., $E\{x_{i,j} w_{l,k}\} = 0, \forall i,j,l,k$. We consider first the case for first-order fields.

First-Order Fields: For simplicity, we provide details of the approach in terms of the Dirichlet field which is the simplest to examine. The data dependent terms in (111) are S_x, X_{h_1}, X_{v_1}, i.e., the sample correlations defined in (90), (102)–(105). The equivalent quantities for the observed field Y are S_y, Y_{h_1}, Y_{v_1}, which are defined in analogous fashion to the corresponding terms for X. Substituting (114) into the definitions for S_y, Y_{h_1}, Y_{v_1}, we get, after minor algebraic simplification,

$$S_y = S_x + \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} w_{i,j}^2 + \frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} x_{i,j} w_{i,j}, \tag{115}$$

$$Y_{h_1} = X_{h_1} + \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M-1} (x_{i,j} w_{i,j+1} + w_{i,j} x_{i,j+1}) + \frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M-1} w_{i,j} w_{i,j+1}, \tag{116}$$

$$Y_{v_1} = X_{v_1} + \frac{1}{NM} \sum_{i=1}^{N-1} \sum_{j=1}^{M} (x_{i,j} w_{i+1,j} + w_{i,j} x_{i+1,j}) + \frac{1}{NM} \sum_{i=1}^{N-1} \sum_{j=1}^{M} w_{i,j} w_{i+1,j}. \tag{117}$$

Considering the summations in (115)–(117) to be approximations of the corresponding expectations and using the statistical properties of the noise (i.e., the fact that it is white), and its independence with respect to the original field X, we get

$$S_y \approx S_x + \sigma_n^2, \tag{118}$$

$$Y_{h_1} \approx X_{h_1}, \tag{119}$$

$$Y_{v_1} \approx X_{v_1}. \tag{120}$$

Applying the same approximations to first-order variational and symmetric fields, we get the same relations as previously shown for the symmetric field and an additional term in (119) and (120) for the variational field. The additional term arises because of the squared terms in the boundary condition corrections for X_{h_1} and X_{v_1}, see (104), (105). Consequently,

TABLE I
RESULTS OF ML ESTIMATION FOR FIRST-ORDER DIRICHLET FIELD $N = M = 32$, $\beta_{h_1} = 0.20$, $\beta_{v_1} = 0.29$, $\sigma^2 = 1.0$. NOISE VARIANCE: FOR 10db $\sigma_n^2 = 0.18$, for 6db $\sigma_n^2 = 0.45$. NUMBER OF SAMPLES = 30

Data	$< \widehat{\beta_{h_1}} >$	$< \widehat{\beta_{v_1}} >$	$< \widehat{\sigma^2} >$	var($\widehat{\beta_{h_1}}$)	var($\widehat{\beta_{v_1}}$)	var($\widehat{\sigma^2}$)
noiseless	0.198505	0.289605	0.995211	0.000306	0.000336	0.002109
noisy($\sigma_n^2 = 0.18$)	0.194630	0.293368	0.993994	0.000504	0.000598	0.002068
noisy($\sigma_n^2 = 0.45$)	0.192400	0.295691	0.991375	0.000642	0.000764	0.003380

TABLE II
RESULTS OF ML ESTIMATION FOR FIRST-ORDER VARIATIONAL FIELD $N = M = 32$, $\beta_{h_1} = 0.20$, $\beta_{v_1} = 0.29$, $\sigma^2 = 1.0$. NOISE VARIANCE: FOR 10 db $\sigma_n^2 = 0.21$, for 6 db $\sigma_n^2 = 0.54$. NUMBER OF SAMPLES = 30

Data	$< \widehat{\beta_{h_1}} >$	$< \widehat{\beta_{v_1}} >$	$< \widehat{\sigma^2} >$	var($\widehat{\beta_{h_1}}$)	var($\widehat{\beta_{v_1}}$)	var($\widehat{\sigma^2}$)
noiseless	0.198492	0.290208	0.994926	0.000340	0.000331	0.002066
noisy($\sigma_n^2 = 0.21$)	0.194581	0.294033	0.994132	0.000538	0.000569	0.002261
noisy($\sigma_n^2 = 0.54$)	0.192522	0.296050	0.991253	0.000711	0.000757	0.004022

for first-order fields the correlations of the noiseless sample can be approximated from the noisy sample using the following relationships:

$$S_x \approx S_y - \sigma_n^2, \tag{121}$$

$$X_{h_1} \approx Y_{h_1} - \begin{cases} 0, & \text{for Dirichlet or symmetric fields,} \\ \frac{\sigma_n^2}{M}, & \text{for variational fields,} \end{cases} \tag{122}$$

$$X_{v_1} \approx Y_{v_1} - \begin{cases} 0, & \text{for Dirichlet or symmetric fields,} \\ \frac{\sigma_n^2}{N}, & \text{for variational fields.} \end{cases} \tag{123}$$

The modified estimation procedure uses (121)–(123) to obtain S_x, X_{h_1}, X_{v_1} from the sample correlations, S_y, Y_{h_1}, Y_{v_1}, of the observed data, and then proceeds exactly as outlined in Section VII.

Higher Order Fields: Generalizing the previous procedure to any order Dirichlet field is straightforward because there are no boundary correction terms to be concerned with, unlike the case of variational or symmetric fields. The same approach as above leads to

$$S_x \approx S_y - \sigma_n^2, \tag{124}$$

$$X_\tau \approx Y_\tau, \qquad \forall \tau \in \eta_p. \tag{125}$$

The estimation procedure outlined in Section VI may be used after the sample correlations have been approximated from the noisy data correlations by using the relationship in (124) and (125). Generalizing to other b.c. is simply a matter of obtaining the corresponding definitions for the correlations X_τ and following the same approach, i.e., approximating the summations as expectations and applying the statistical properties. Experimental results for the modified procedure are presented in Section IX.

IX. EXPERIMENTAL RESULTS

In this section, we present results obtained when the ML estimation procedures discussed in Sections VI and VII are applied to noiseless fields and also to noisy fields, with the modifications outlined in Section VIII being used in the latter case. The experiments were conducted using synthetic fields generated recursively by means of the equivalent one–sided representation for the noncausal field [30]. We also present the results of an experiment with real data and a comparison with toroidal boundary conditions.

First-Order Fields: Thirty samples of a 32×32 first-order Dirichlet field with parameters

$$\beta_{h_1} = 0.2, \qquad \beta_{v_1} = 0.29, \qquad \sigma^2 = 1.0, \tag{126}$$

were generated. The ML estimation procedure outlined in Section VII was applied to each sample and the corresponding ML estimates were obtained. The mean and variance of the estimates for each parameter, defined respectively as $< . >$ and var(.), were computed for the 30 samples and the results recorded in Table I. The results demonstrate that even for a small number of samples (30), the mean value of the estimates for each of the parameters is close to the actual value, and the variance of the estimates is small. The same procedure was repeated with measurement noise being simulated by adding independent Gaussian random numbers with variance $\sigma_n^2 = 0.18$, chosen to get a signal-to-noise ratio (SNR) of approximately 10db (computed using the average power of the 30 samples), to each sample field prior to the estimation. The procedure outlined in Section VIII for the case of noisy data was used. The mean and variance of the estimates for each parameter are presented in Table I. The same procedure was repeated with noise variance $\sigma_n^2 = 0.45$, chosen to get an SNR of approximately 6db, and the corresponding results also recorded in Table I. These results demonstrate that, even for high noise levels, the modifications in Section VIII enable the estimation procedure to produce estimates of almost the same quality as the noiseless estimates.

The entire experiment was repeated for a 32×32 variational field with the same parameters. As before, the measurement noise variances were chosen to correspond to SNR's of approximately 10 dB and 6 dB, respectively. The results are in Table II. They show the same properties as the corresponding results for the Dirichlet field.

Second-Order Fields: The next experiment uses the ML estimation procedure outlined in Section VI to estimate the

TABLE III
RESULTS OF ML ESTIMATION FOR SECOND-ORDER DIRICHLET (MEANS OF THE ESTIMATES) $N = M = 32$, $\beta_{h_1} = 0.0500$, $\beta_{v_1} = 0.1950$, $\beta_{ld_{11}} = -0.1350$, $\beta_{rd_{11}} = 0.1010$, $\sigma^2 = 1.0$. Noise variance (for SNR = 6 db) $\sigma_n^2 = 0.29$. NUMBER OF SAMPLES = 30

Data	$<\widehat{\beta_{h_1}}>$	$<\widehat{\beta_{v_1}}>$	$<\widehat{\beta_{ld_{11}}}>$	$<\widehat{\beta_{rd_{11}}}>$	$<\widehat{\sigma^2}>$
noiseless	0.043060	0.194016	-0.132248	0.110510	0.988772
noisy($\sigma_n^2 = 0.29$)	0.041380	0.203856	-0.123848	0.101423	0.963911

TABLE IV
RESULTS OF ML ESTIMATION FOR SECOND-ORDER DIRICHLET (VARIANCES OF THE ESTIMATES) $N = M = 32$, $\beta_{h_1} = 0.0500$, $\beta_{v_1} = 0.1950$, $\beta_{ld_{11}} = -0.1350$, $\beta_{rd_{11}} = 0.1010$,$\sigma^2 = 1.0$. NOISE VARIANCE (FOR SNR = 6 db) $\sigma_n^2 = 0.29$. NUMBER OF SAMPLES = 30

Data	$\text{var}(\widehat{\beta_{h_1}})$	$\text{var}(\widehat{\beta_{v_1}})$	$\text{var}(\widehat{\beta_{ld_{11}}})$	$\text{var}(\widehat{\beta_{rd_{11}}})$	$\text{var}(\widehat{\sigma^2})$
noiseless	0.000952	0.000921	0.000586	0.000767	0.001827
noisy($\sigma_n^2 = 0.29$)	0.001632	0.000932	0.000781	0.001869	0.002355

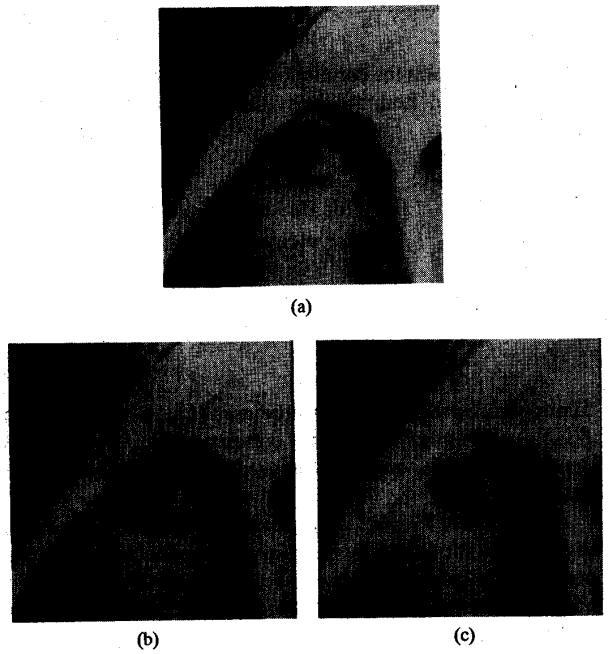
parameters of a 32×32 second-order Dirichlet field with parameters,

$$\beta_{h_1} = 0.0500, \quad \beta_{v_1} = 0.1950, \quad \beta_{ld_{11}} = -0.1350, \quad \beta_{rd_{11}} = 0.1010, \quad \sigma^2 = 1.0. \tag{127}$$

Thirty samples of this field were generated and the ML estimation procedure was applied to estimate the parameters from each sample. With reference to the discussion in Section VI regarding the computation of the gradient using the Monte Carlo sampling algorithm, see (97) and (98), we found that a small number of samples, $K = 25$ for this experiment, was sufficient to provide high quality estimates. The mean and variance of each of the parameter estimates for the 30 samples are presented in Tables III and IV, respectively. The results are promising, with the estimate means being close to the actual parameter values while the variances are small, even though the estimates for only 30 samples were averaged. The fact that a small value for K, in (98), was adequate to provide good results enables fast computation of the estimates. The experiment was repeated with the samples corrupted by the addition of white Gaussian noise with variance $\sigma_n^2 = 0.29$ (chosen to produce an SNR of approximately 6 db), with the modified procedure being used to compute the estimates. These results are also recorded in Tables III and IV.

Application to Image Coding:

An important application of the recursive framework for noncausal GMRF's [30] is the coding of images for transmission or storage. The results presented in this paper are an essential part of the framework since they enable the identification of the underlying noncausal field. The importance of this framework, and correspondingly of the present work, is illustrated by a simple experiment in which we code a 128×128 portion of the Lenna image, see Fig. 5(a). The image is coded using a noncausal first-order Dirichlet field, and also, for the purposes of comparison, using a causal field model with three neighbors, the third-order Markov Mesh [1], see Fig. 6. The ML estimation procedure from Section VII was used to compute the parameters for the noncausal field model, while the parameters of the causal field were obtained using least squares estimation, e.g., [21]. The noncausal field parameters were obtained as $\widehat{\beta_{h_1}} = 0.103827$, $\widehat{\beta_{v_1}} = 0.394309$, $\widehat{\sigma^2} = 130.448801$, while the causal field



(a) (b) (c)

Fig. 5. (a) 128×128 Lenna, (b) Lenna coded to 2 bits/pixel using noncausal first-order Dirichlet field, (c) Lenna coded to 2 bits/pixel using causal (third-order Markov Mesh) field model.

(a) (b)

Fig. 6. Causal field model (third-order Markov Mesh): (a) Neighbors marked by "x," (b) Coefficients for the 3 pairwise interactions.

parameters were $\widehat{\beta_h} = 0.666224$, $\widehat{\beta_v} = 0.863320$, $\widehat{\beta_d} = -0.556740$, and $\widehat{\sigma^2} = 118.731850$. Real images, such as the Lenna, are often highly correlated, leading to parameter estimates that are close to the boundary of the parameter space. Furthermore, the (negative log) likelihood function $L(.)$ varies very slowly in the direction parallel to the boundary. As a result, there is a local region of parameter values for which

TABLE V
RESULTS OF ML ESTIMATION FOR FIRST-ORDER DIRICHLET FIELD $N = 8$, $M = 32$, $\beta_{h_1} = 0.10$, $\beta_{v_1} = 0.425$, $\sigma^2 = 1.0$. NUMBER OF SAMPLES = 30

Assumed model	$< \widehat{\beta_{h_1}} >$	$< \widehat{\beta_{v_1}} >$	$< \widehat{\sigma^2} >$	var($\widehat{\beta_{h_1}}$)	var($\widehat{\beta_{v_1}}$)	var($\widehat{\sigma^2}$)
Dirichlet	0.100758	0.423846	1.006015	0.000238	0.000261	0.004898
Toroidal	0.140650	0.355827	1.179194	0.001673	0.001458	0.021216

TABLE VI
RESULTS OF ML ESTIMATION FOR FIRST-ORDER DIRICHLET FIELD $N = 8$, $M = 64$, $\beta_{h_1} = 0.10$, $\beta_{v_1} = 0.425$, $\sigma^2 = 1.0$. NUMBER OF SAMPLES = 30

Assumed model	$< \widehat{\beta_{h_1}} >$	$< \widehat{\beta_{v_1}} >$	$< \widehat{\sigma^2} >$	var($\widehat{\beta_{h_1}}$)	var($\widehat{\beta_{v_1}}$)	var($\widehat{\sigma^2}$)
Dirichlet	0.104525	0.418060	1.000391	0.000669	0.000657	0.007466
Toroidal	0.150306	0.348399	1.232698	0.001160	0.001115	0.015102

$L(.)$ is within 1% of its minimum value. Exactly which point is selected from this region depends on the tolerances chosen in the estimation procedure. See [3] for a discussion of this issue.

The noncausal field representation was transformed into its equivalent one–sided formulation which is driven by an uncorrelated error field [30]. The error field for the given image was computed from this formulation and is quantized at a rate of 2 bits/pixel using the optimal scalar Lloyd–Max quantizer, e.g., [17]. The corresponding error field for the causal representation wass computed and quantized at the same bit rate using the same quantizer.

The reconstructed images for the noncausal and causal field models are provided in Fig. 5(b) and (c), respectively. A comparison of the two images shows the dramatic difference between the results of noncausal versus causal modeling. Noncausal modeling suppresses the directional streaking and other artifacts of causality, while we retain the computational benefits of recursiveness because of the recursive structure [30]. This simple example is presented here as part of the motivation for this work, in particular to highlight the importance of noncausal representations for 2-D phenomena. We have used vector quantization to achieve much higher compression ratios. A detailed discussion of applications of this work to image coding and reconstruction of 2-D oceanographic data will be presented elsewhere.

Comparison with Toroidal Boundary Conditions: In many applications where physical phenomena are modeled using elliptic PDE's, Dirichlet or Neumann boundary conditions are used rather than periodic boundary conditions. Correspondingly, the stochastic autoregressive (AR) models obtained from the discretization of the PDE model are given the same nonperiodic b.c. A question that may arise is whether there is a significant difference when toroidal boundary conditions are used to approximate nonperiodic fields. It is not our goal to provide a complete or definitive answer to this question. We address it from the point of view of parameter estimation.

Asymptotically, the effect of boundary conditions is negligible, see Moran, [28], in which it is shown that infinite lattice GMRF's may be constructed as a limiting case of fields defined on a toroidal lattice. However, as illustrated below through some experiments, for fields on finite lattices, particularly in applications where one or more of the lattice dimensions is not large, the use of toroidal boundary conditions may produce a noticeable difference in the parameter estimates. This is because the parameter space constraints for a toroidal field differ from those of other finite lattice fields, for example, Dirichlet fields. How significant the impact of these discrepancies is depends on the actual application.

In Table V, we present the results of a first experiment where 30 samples of a first-order Dirichlet field with lattice dimensions 8×32 were synthesized. The parameters for a first-order Dirichlet field model were estimated from each sample using our algorithm. In addition, each sample was also used to estimate the parameters of a first-order toroidal field model. The object of the experiment was to see how well the toroidal boundary approximated a nonperiodic (in this case, Dirichlet) field. The mean and variance, defined respectively as $< . >$ and var(.), of the estimates for each parameter were computed for the 30 samples and are given in Table V.

The same experiment was repeated for 8×64, 16×32, 16×64, 16×128, and 64×64 with the results being recorded respectively, in Tables VI–X. These block sizes may well arise in segmenting more complex fields. The toroidal boundary approximation consistently produces biased estimates.

From Tables V–X, we conclude that the errors of the toroidal field estimates are 1–2 orders of magnitude larger than the corresponding errors of the Dirichlet field estimates, see Table XI where r_h, r_v, r_{σ^2}, are the ratios of the bias of the toroidal estimates versus the Dirichlet for respectively, β_{h_1}, β_{v_1}, and σ^2. The bias is due to the fact that the parameter space for first–order toroidal fields is always subsumed by the one for Dirichlet fields, compare Result 1 for Dirichlet fields with the corresponding result for toroidal fields in [20].

The results just mentioned show that the periodic boundary condition does not necessarily provide a good approximation for a Dirichlet field. In fact, there can be significant differences in the estimates. One can expect to obtain the converse result, i.e., Dirichlet fields are not necessarily good approximations of periodic boundary conditions. Ideally, therefore, the boundary conditions used in the model should be matched to the characteristics of the phenomenon being modeled.

X. CONCLUSION

In this paper, we have studied the parameter structure of noncausal finite lattice Gauss Markov Random Fields. The results include the following.

- The eigenstructure of the inverse covariance matrix for first-order fields with boundary conditions drawn from

TABLE VII
RESULTS OF ML ESTIMATION FOR FIRST-ORDER DIRICHLET FIELD $N = 16$, $M = 32$, $\beta_{h_1} = 0.10$, $\beta_{v_1} = 0.405$, $\sigma^2 = 1.0$. NUMBER OF SAMPLES = 30

Assumed model	$<\widehat{\beta_{h_1}}>$	$<\widehat{\beta_{v_1}}>$	$<\widehat{\sigma^2}>$	var($\widehat{\beta_{h_1}}$)	var($\widehat{\beta_{v_1}}$)	var($\widehat{\sigma^2}$)
Dirichlet	0.100223	0.402252	1.008461	0.000340	0.000330	0.005091
Toroidal	0.109232	0.384998	1.052985	0.000447	0.000416	0.005811

TABLE VIII
RESULTS OF ML ESTIMATION FOR FIRST-ORDER DIRICHLET FIELD $N = 16$, $M = 64$, $\beta_{h_1} = 0.10$, $\beta_{v_1} = 0.405$, $\sigma^2 = 1.0$. NUMBER OF SAMPLES = 30

Assumed model	$<\widehat{\beta_{h_1}}>$	$<\widehat{\beta_{v_1}}>$	$<\widehat{\sigma^2}>$	var($\widehat{\beta_{h_1}}$)	var($\widehat{\beta_{v_1}}$)	var($\widehat{\sigma^2}$)
Dirichlet	0.097117	0.405307	0.993270	0.000138	0.000142	0.002459
Toroidal	0.106049	0.390370	1.029178	0.000151	0.000143	0.002524

TABLE IX
RESULTS OF ML ESTIMATION FOR FIRST-ORDER DIRICHLET FIELD $N = 16$, $M = 128$, $\beta_{h_1} = 0.10$, $\beta_{v_1} = 0.40$, $\sigma^2 = 1.0$. NUMBER OF SAMPLES = 30

Assumed model	$<\widehat{\beta_{h_1}}>$	$<\widehat{\beta_{v_1}}>$	$<\widehat{\sigma^2}>$	var($\widehat{\beta_{h_1}}$)	var($\widehat{\beta_{v_1}}$)	var($\widehat{\sigma^2}$)
Dirichlet	0.100312	0.399454	0.998255	0.000092	0.000099	0.001491
Toroidal	0.108931	0.383039	1.028518	0.000102	0.000106	0.001634

TABLE X
RESULTS OF ML ESTIMATION FOR FIRST-ORDER DIRICHLET FIELD $N = 64$, $M = 64$, $\beta_{h_1} = 0.10$, $\beta_{v_1} = 0.400$, $\sigma^2 = 1.0$. NUMBER OF SAMPLES = 30

Assumed model	$<\widehat{\beta_{h_1}}>$	$<\widehat{\beta_{v_1}}>$	$<\widehat{\sigma^2}>$	var($\widehat{\beta_{h_1}}$)	var($\widehat{\beta_{v_1}}$)	var($\widehat{\sigma^2}$)
Dirichlet	0.099788	0.399854	0.998306	0.000042	0.000040	0.000648
Toroidal	0.101355	0.397359	1.003812	0.000045	0.000042	0.000693

TABLE XI
RATIO OF BIAS ERRORS (TOROIDAL/DIRICHLET) r_h, r_v, r_{σ^2}, ARE THE RATIOS OF THE BIAS OF THE TOROIDAL ESTIMATES VERSUS THE DIRICHLET FOR, RESPECTIVELY, $\widehat{\beta_{h_1}}, \widehat{\beta_{v_1}}$, AND $\widehat{\sigma^2}$

Relative Error (Toroidal/Dirichlet)	Table V	Table VI	Table VII	Table VIII	Table IX	Table X
r_{h_1}	8.98	66.37	41.40	2.10	28.625	6.39
r_{v_1}	9.97	66.38	7.28	47.65	31.06	18.09
r_{σ^2}	458.30	38.69	6.26	34.04	16.34	2.25

the PDE literature, as well as for a special case of a second-order field.

- A complete specification of the parameter space for the first and second-order fields mentioned above.
- Sufficient conditions on the parameter space of Dirichlet fields of arbitrary order, that are always tighter than the corresponding strict diagonal dominance sufficient conditions.
- Explicit expressions for the likelihood function for the first and second-order fields mentioned above. These expressions facilitate fast computation of the ML estimates.
- A new parameter estimation procedure for fields of arbitrary order. This procedure exploits the recursive structure of noncausal finite lattice GMRF's to produce a computationally practical implementation. Work is continuing on studying the properties of the algorithm.
- Practical modifications that enable the ML estimators to produce good results even when the field sample is corrupted by additive white Gaussian noise.

The present work complements earlier work deriving a framework for recursive processing of noncausal fields in that it enables the estimation of the underlying noncausal field model for a given 2-D field. Taken as a whole, the work herein and in [30] facilitates the practical application of noncausal fields in 2-D signal processing areas. This is important because, as illustrated by the image coding example in Section IX, noncausal field representations provide better results for 2-D spatial phenomena than artificially formulated causal ones.

APPENDIX
PROOFS AND DERIVATIONS

Proof of Lemma 2: Let

$$D(\alpha) = \begin{bmatrix} 1-\alpha & -\alpha & 0 & . & . & . \\ -\alpha & 1 & -\alpha & 0 & . & . \\ 0 & -\alpha & 1 & -\alpha & 0 & . \\ . & . & . & . & . & . \\ . & . & 0 & -\alpha & 1 & -\alpha \\ . & . & . & 0 & -\alpha & 1-\alpha \end{bmatrix}. \quad (128)$$

This is the transform matrix for the discrete cosine transform (DCT) [18]. Its eigenvalues are given by [18]:

$$\lambda_k(D(\alpha)) = 1 - 2\alpha\cos\frac{(k-1)\pi}{K}, \quad 1 \le k \le K. \quad (129)$$

For the first-order variational field, the matrix S_K, given by (20), can be expressed as

$$S_K = I_K - D(1), \quad (130)$$

where $D(1)$ is the DCT matrix evaluated at $\alpha = 1$. Then, the eigenvalues of S_K are given by

$$\lambda_k(S_K) = 1 - \lambda_k(D(1)), \quad 1 \le k \le K, \quad (131)$$

$$= 2\cos\frac{(k-1)\pi}{K}, \quad 1 \le k \le K. \quad (132)$$

□

Proof for Lemma 3: The eigenvalues of S_K are the roots of the characteristic polynomial,

$$D_K(\lambda) = |\lambda I_K - S_K|, \quad (133)$$

$$= \begin{vmatrix} \lambda & -2 & 0 & . & . & . & . \\ -2 & \lambda & -1 & 0 & . & . & . \\ 0 & -1 & \lambda & -1 & 0 & . & . \\ 0 & 0 & -1 & \lambda & -1 & 0 & . \\ . & . & . & . & . & . & . \\ . & . & 0 & -1 & \lambda & -1 & 0 \\ . & . & . & 0 & -1 & \lambda & -2 \\ . & . & . & . & 0 & -2 & \lambda \end{vmatrix}, \quad (134)$$

where we have assumed $K > 5$. A recurrence relation for $D_K(\lambda)$ is obtained by expanding the above determinant through elements chosen to eliminate the four 2's, stopping only when this is achieved. The resulting recurrence relation is,

$$D_K(\lambda) = \lambda^2 T_{K-2}(\lambda) - 8\lambda T_{K-3}(\lambda) + 16 T_{K-4}(\lambda), K > 5, \quad (135)$$

where $T_K(\lambda)$ is the characteristic polynomial for the symmetric, Toeplitz, tridiagonal $K \times K$ matrix given in (18), i.e.,

$$T_K(\lambda) = \begin{vmatrix} \lambda & -1 & 0 & . & . & . \\ -1 & \lambda & -1 & 0 & . & . \\ 0 & -1 & \lambda & -1 & 0 & . \\ . & . & . & . & . & . \\ . & . & 0 & -1 & \lambda & -1 \\ . & . & . & 0 & -1 & \lambda \end{vmatrix}. \quad (136)$$

The recurrence relation for $T_K(\lambda)$ is given by, e.g., [12],

$$T_K(\lambda) = \lambda T_{K-1}(\lambda) - T_{K-2}(\lambda), \quad T_0 = 1, T_1 = \lambda. \quad (137)$$

The closed form solution for (137) for $\lambda > 0$ is provided, for example, in [12], and is easily extended to $\lambda \le 0$ to get,

$$T_K(\lambda) = \begin{cases} \frac{\sinh(K+1)\theta}{\sinh\theta} & \text{if } \lambda > 2, \ 2\cosh\theta = \lambda, \\ (-1)^K\left(\frac{\sinh(K+1)\theta}{\sinh\theta}\right) & \text{if } \lambda < -2, \ -2\cosh\theta = \lambda, \\ (K+1) & \text{if } \lambda = 2, \\ (-1)^K(K+1) & \text{if } \lambda = -2, \\ \frac{\sin(K+1)\theta}{\sin\theta} & \text{if } |\lambda| < 2, \ 2\cos\theta = \lambda, \end{cases} \quad (138)$$

We will consider each of the cases in (138).

Case 1 ($\lambda > 2$): We can represent λ as

$$\lambda = 2\cosh\theta, \quad (139)$$

and solve for the corresponding θ values. Substituting (139) and the closed form solution of $T_K(\lambda)$ for $\lambda > 2$ from (138) into (135), we get, after some simplification using standard hyperbolic identities,

$$D_K(\theta) = \sinh(K+1)\theta - 6\sinh(K-1)\theta + 9\sinh(K-3)\theta. \quad (140)$$

A plot of this function for $K = 8$ is provided in Fig. 7 to make the discussion that follows more concrete. The hyperbolic equation (25) is obtained by setting (140) equal to zero. Since the sinh function is odd, the roots of (25) are symmetric about the origin, i.e., they come in $\pm$ pairs. Note that $D_K(\theta)$ is not defined by the right hand side of (140) for $\theta = 0$ because that corresponds to $\lambda = 2$, which is handled separately. The fact that (25) has only four roots is proved by examining the relative effects of each of the three terms in the function. Consider the form of (140) for $\theta \ge 0$, see, for example, Fig. 7(b). At $\theta = 0$, this function evaluates to zero, while its first derivative is positive. Therefore, for small θ, or more formally, for the range,

$$0 < \epsilon \le \theta \le \theta_1,$$

with ϵ arbitrarily small, and θ_1 depending on K, the positive term $(9\sinh(K-3)\theta)$ dominates and the function remains positive. As θ increases, the relative size of the arguments of the sinh functions becomes more important than their coefficients. Consequently, at some θ, which we call θ_1 above, the negative term $(-6\sinh(K-1)\theta)$ begins to dominate, and we have the first zero crossing as the function becomes negative. Eventually, as θ continues to increase, $(K+1)\theta$ becomes large enough that the positive term $(\sinh(K+1)\theta)$ dominates from then onwards, leading to the second zero crossing as the function rises monotonically. There are no other terms to offset this tendency, hence the monotonic increase continues indefinitely and there are no more zero crossings. As a result of the odd symmetry of the function, the two positive zero crossings, θ_{k_1}, θ_{k_2}, have mirror images, $-\theta_{k_1}$, $-\theta_{k_2}$. Since the cosh function is symmetric, the four roots of (25) substituted into (139) provide only two eigenvalues of S_K.

Case 2 ($\lambda < -2$): Let

$$\lambda = -2\cosh\theta, \quad (141)$$

and solve for the corresponding θ values exactly as above, except that we consider separately odd and even values of K. For even values of K, substituting from (138) into (135), we get the polynomial given in (140). For odd values of K, we get the negative of the same polynomial. Therefore, for any value of K we get the hyperbolic equation (25), which was shown above to have four roots, $\theta_{k_1}, -\theta_{k_1}, \theta_{k_2}, -\theta_{k_2}$. As before, because of the symmetry of the cosh function, substituting the four roots into (141) provides only two eigenvalues of S_K.

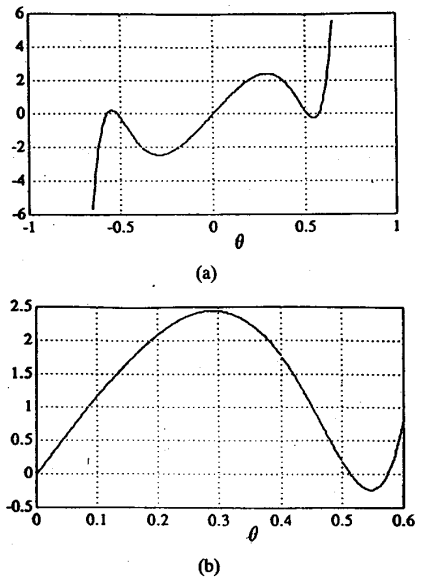


Fig. 7. Plot of $D_8(\theta)$ for $\lambda > 2$ as a function of θ (in radians): (a) Plot for $\theta \in [-0.65, 0.65]$, (b) Close up showing the zero crossings on the positive θ axis.

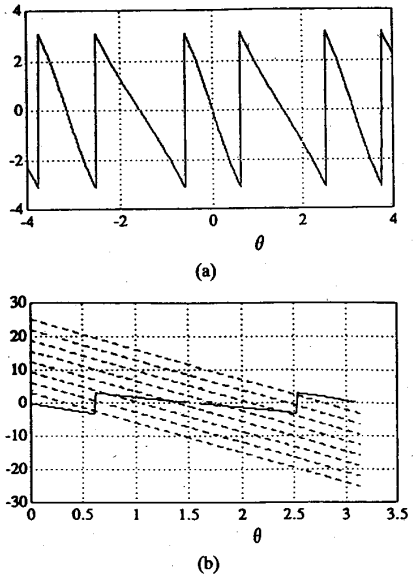


Fig. 8. (a) Plot of $2\tan^{-1}\frac{3\sin 2\theta}{1-3\cos 2\theta}$ as a function of θ (in radians), (b) Same function plotted for $\theta \in [0, \pi]$, using solid lines, superimposed with the family of lines indexed by k, $[k\pi - (K+1)\theta]$, for $1 \le k \le K = 8$, plotted using dashed lines.

Case 3 ($|\lambda| = 2$): Substituting

$$\lambda = 2, \tag{142}$$

along with the corresponding solution for $T_K(\lambda)$ from (138), into (135), we get,

$$D_K(2) = 4(K-5), \tag{143}$$

consequently, (142) is not an eigenvalue of S_K for $K > 5$. Using

$$\lambda = -2, \tag{144}$$

instead of (142), and repeating the above procedure, we get,

$$D_K(-2) = \begin{cases} 4(K-5) & \text{for } K \text{ even} \\ 4(5-K) & \text{for } K \text{ odd} \end{cases}, \tag{145}$$

in either case, (144) is not an eigenvalue of S_K for $K > 5$.

Case 4 ($|\lambda| < 2$): We can represent λ as

$$\lambda = 2\cos\theta, \tag{146}$$

and solve for the corresponding θ values. Substituting (146) and the closed form solution of $T_K(\lambda)$ for $|\lambda| < 2$, from (138), into (135), after some simplification using standard trigonometric identities, we get,

$$D_K(\theta) = \sin(K+1)\theta - 6\sin(K-1)\theta + 9\sin(K-3)\theta. \tag{147}$$

Using Euler's identity to replace each of the sine terms in (147), and setting the expression equal to zero, we obtain the characteristic equation in the following form,

$$\exp\{j2(K+1)\theta\}\frac{(1-3\exp\{-j2\theta\})^2}{(1-3\exp\{j2\theta\})^2} = 1, \quad \text{(where } j = \sqrt{-1}\text{)}, \tag{148}$$

$$\Longrightarrow \exp\{j[2(K+1)\theta + 4\tan^{-1}\frac{3\sin 2\theta}{1-3\cos 2\theta}]\} = 1, \tag{149}$$

$$\Longrightarrow [(K+1)\theta - k\pi] + 2\tan^{-1}\frac{3\sin 2\theta}{1-3\cos 2\theta}\} = 0, \quad k = 0, \overset{+}{-}1, \overset{+}{-}2, \ldots, \tag{150}$$

i.e., k is any integer. The roots of (150) can be interpreted as the points of intersection between the function,

$$2\tan^{-1}\frac{3\sin 2\theta}{1-3\cos 2\theta}, \tag{151}$$

and a family of lines indexed by k,

$$k\pi - (K+1)\theta, \quad k = 0, \overset{+}{-}1, \overset{+}{-}2, \ldots, \tag{152}$$

with constant slope $-(K+1)$, and θ axis intercept $\pi/(K+1)$. It is straightforward to show that the function in (151) is periodic in θ with period π, and has odd symmetry about the θ axis. Both these properties are verified by the plot in Fig. 8(a). Since the slopes of the lines in (152) are all always (i.e., for all K) larger in magnitude than the slope of the continuous portions of (151), each of these lines may have no more than one point of intersection with (151), see, for example, Fig. 8(b).

From the symmetry of (147) it is clear that the roots come in pairs $(\theta, \pi - \theta)$, i.e., if θ_k is a root of (147), then so is $(\pi - \theta_k)$. From the periodicity of (151) it follows that any point of intersection, θ_{k_2}, between (151) and a line, from (152), indexed by $k_2 > K$ or $k_2 < 0$, can be expressed in the form $(\theta_{k_1} + l\pi)$ where θ_{k_1} is an intersection point for a line indexed by $k_1 \in [0, K]$, and l is some integer. Thus, θ_{k_2} does not

provide any additional eigenvalues beyond those provided by the lines indexed by $k \in [0, K]$, because

$$\cos(\theta_{k_1} + l\pi) = \begin{cases} \cos\theta_{k_1} & \text{if } l \text{ is even} \\ \cos(\pi - \theta_{k_1}) & \text{if } l \text{ is odd} \end{cases}. \quad (153)$$

In addition, we discard the line indexed by $k = 0$ because it intersects (151) at $\theta = 0$, which is excluded from this case (see (138)). We are left with the family of lines indexed by $1 \le k \le K$.

As an illustration, in Fig. 8(b) we plot (151) for $\theta \in [0, \pi]$ as well as the family of lines from (152) for $1 \le k \le K = 8$. The points of intersection of the two functions correspond to the roots of (147) which provide those eigenvalues of S_K that take the form of (146). From Fig. 8(b), we see that only four valid points of intersections exist for $1 \le k \le 8$, because four of the eight lines from (152) pass through the discontinuities in (151). Below, using simple trigonometry we show that this is always the case, i.e., for $1 \le k \le K$, there are always only $K - 4$ roots of (147).

Let k_1 index the line from (152) that is the last to intersect (151) to the left of the first discontinuity, see Fig. 8(a), and let k_2 index the line that is the first to intersect after the first discontinuity. We assume that these lines intersect (151) at function value $-\pi$ and π respectively, see Fig. 9, and derive the horizontal distance, (marked as "x" in the figure) that separates them. Using the right angle triangle in Fig. 9, we derive this distance as

$$\frac{2\pi}{K+1}.$$

Since lines in (152) have intercepts that are spaced $\pi/(K+1)$ apart, it follows that $k_2 = k_1 + 1$, i.e., for this configuration one line is lost in the first discontinuity. In actuality, for integer k, K, it is not possible to have a line intersect (151) exactly at $-\pi$ (or π) which corresponds to $\theta = 0.5\cos^{-1}(1/3)$; as a result two lines are lost in the gap. The same analysis can be carried out to show that two more lines are lost in the second discontinuity in $[0, \pi]$, see Fig. 8. Therefore, only $K-4$ of the K lines indexed by $1 \le k \le K$ have valid points of intersection with (151), providing us with $K - 4$ distinct eigenvalues of S_K that take the form of (146). This concludes the proof. □

Proof of Lemma 4: From (15)–(17),

$$A = I_N \otimes (I_M - \beta_{h_1} S_M) + S_N \otimes (-\beta_{v_1} I_M), \quad (154)$$

Using the distributive property of Kronecker products (property II, in [14]), in (154), we get

$$A = I_N \otimes I_M - \beta_{h_1}(I_N \otimes S_M) - \beta_{v_1}(S_N \otimes I_M). \quad (155)$$

Let the $K \times 1$ column vector, ϕ_k, represent the kth orthonormal eigenvector of the matrix S_K. Using property IX, from [14, p. 27], the NM eigenvalues of $I_N \otimes I_M$, $I_N \otimes S_M$, and $S_N \otimes I_M$, are given by $\{\lambda_i(I_N)\lambda_j(I_M)\}_{i=1,j=1}^{N,M}$, $\{\lambda_i(I_N)\lambda_j(S_M)\}_{i=1,j=1}^{N,M}$, and $\{\lambda_i(S_N)\lambda_j(I_M)\}_{i=1,j=1}^{N,M}$, respectively, with the corresponding eigenvectors given by $\{\phi_i \otimes \phi_j\}$, where ϕ_i and ϕ_j are $N \times 1$, and $M \times 1$, respectively. Since all three terms in (155) have the same set of eigenvectors, the eigenvalues of A are obtained as the sum of the corresponding eigenvalues of the terms:

$$\lambda_{i,j}(A) = 1 - \beta_{h_1}\lambda_j(S_M) - \beta_{v_1}\lambda_i(S_N), \quad 1 \le i \le N, \ 1 \le j \le M. \quad (156)$$

□

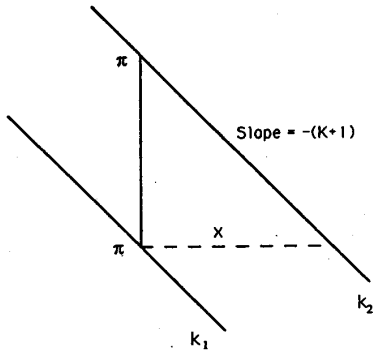


Fig. 9. Triangle formed by the intersection of linesindexed by k_1 and k_2 with the function2 $\tan^{-1}\frac{3\sin 2\theta}{1-3\cos 2\theta}$ at function values $-\pi$ and π respectively. The base of the triangle is the horizontal distance between these two lines, marked by a dashed line and labeled as"x."

Proof of Lemma 5: Using (16), (27), (28), and the distributive property of Kronecker products, we get

$$A = I_N \otimes I_M - \beta_{h_1}(I_N \otimes S_M) - \beta_{v_1}(S_N \otimes I_M) - \beta_{d_{11}}(S_N \otimes S_M), \quad (157)$$

with $S_K (K = N, M)$ being given by (18). The rest of the proof follows along the same lines as the proof for Lemma 4. Even though (157) has an additional term, $(-\beta_{d_{11}}(S_N \otimes S_M))$, as compared to (155), this term has the same set of eigenvectors as the others and is handled in exactly the same manner. □

Proof for Lemma 7: *a)* $\|I_H^j\|_s$: From property IX, [14, p. 27], the NM eigenvalues of the Kronecker product $I_H^j = I_N \otimes H_M^j$ are given by,

$$\lambda_{n,m}(I_N \otimes H_M^j) = \lambda_n(I_N)\lambda_m(H_M^j), \quad 1 \le n \le N, \ 1 \le m \le M. \quad (158)$$

Therefore,

$$\|I_H^j\|_s = \|H_M^j\|_s \quad (159)$$

The characteristic polynomial of H_M^j is provided in [15, Thm. 3.5] as

$$|\lambda I - H_M^j| = [\prod_{l=1}^{p+1}(\lambda - 2\cos\frac{l\pi}{p+2})]^q \, [\prod_{l=1}^{p}(\lambda - 2\cos\frac{l\pi}{p+1})]^{j-q}, \quad (160)$$

where $M = pj + q$, $0 \le q \le j - 1$. From this we get,

$$\|H_M^j\|_s = 2\cos\frac{\pi}{M_j + 1}, \quad (161)$$

with M_j defined in (45), and (47) follows immediately by substituting (161) into (159).

b) $\|H_i^I\|_s$: The same property of Kronecker products applied to $H_i^I = H_N^i \otimes I_M$ provides the NM eigenvalues as

$$\lambda_{n,m}(H_i^I \otimes I_M) = \lambda_n(H_i^I)\lambda_m(I_M), \quad 1 \le n \le N, \ 1 \le m \le M. \tag{162}$$

Correspondingly,

$$\|H_i^I\|_s = \|H_N^i\|_s, \tag{163}$$

and (48) follows by using (161) with j, M and M_j replaced by i, N, and N_i, respectively.

c) $\|P_{i,j}\|_s$: The desired result is obtained using graph theory. The directed graph of an $NM \times NM$ matrix $P_{i,j} = [p_{l,k}]$ is constructed by considering each nonzero element $p_{l,k}$ to represent a connection (arc) from node l to node k, in other words the matrix elements are considered to represent the interactions between NM nodes, see, for example, [37]. It is well known, e.g., [15], that if the graph of a symmetric matrix has D disjoint components, the rows and columns of the matrix can be permuted to obtain a block diagonal matrix with D blocks, each corresponding to one of the components of the graph. Since $P_{i,j}$ has no more than 2 nonzero elements in any row (or column), see (10), (2), the components of the graph are all linear trees, hence the corresponding blocks are tridiagonal. In particular, since all the nonzero elements are 1's, the tridiagonal blocks are of the form H_K^1, with K being the length of the corresponding component. Defining Q to represent the number of nodes in the largest disjoint component of the graph, which also makes it the dimension of the largest nonzero block along the diagonal of the permuted matrix, we get,

$$\|P_{i,j}\|_s = \|H_Q^1\|_s = 2\cos\frac{\pi}{Q+1} \quad \text{(using (19))}. \tag{164}$$

It remains to show that Q has the form given in (46). This is done by deriving the component that originates from node 1 and showing that no other component can be longer. The details of this derivation are very lengthy and may be found in [3, Appendix C].

d) $\|P_{i,j,T}\|_s$: $P_{i,j,T}$ has the same block structure as $P_{i,j}$ except that the K_M^j and $(K_M^j)^T$ blocks are swapped. Consequently, the proof for (50) follows the same lines as above except that the largest component in the graph starts at node M instead of node 1. Alternatively, it is straightforward to permute the rows and columns of $P_{i,j,T}$ to obtain $P_{i,j}$ which means the 2 matrices have the same eigenvalues and (50) follows immediately from (49). □

Proof of Result 5: From (38), we get

$$\lambda_{min}(A_p) = \lambda_{min}(A_{p-1} + \Delta_p), \tag{165}$$
$$\ge \lambda_{min}(A_{p-1}) + \lambda_{min}(\Delta_p) \quad \text{(applying Thm. 5 from [25, pg. 205])}, \tag{166}$$
$$= \lambda_{min}(A_{p-1}) + \lambda_{min}\Big(\sum_{\tau\in\Gamma_p}(-\beta_\tau\Delta_p^\tau)\Big) \quad \text{(substituting (39))}, \tag{167}$$
$$\ge L_{p-1} + \lambda_{min}\Big(\sum_{\tau\in\Gamma_p}(-\beta_\tau\Delta_p^\tau)\Big) \quad \text{(using definition of } L_{p-1}\text{)}, \tag{168}$$
$$\ge L_{p-1} + \sum_{\tau\in\Gamma_p}\lambda_{min}(-\beta_\tau\Delta_p^\tau) \quad \text{(applying Thm. 5 from [25, pg. 205])}, \tag{169}$$
$$= L_{p-1} - \sum_{\tau\in\Gamma_p}|\beta_\tau|\|\Delta_p^\tau\|_s. \tag{170}$$

The final equality in (170) is derived as follows. From Lemma 6, we know that Δ_p^τ is one of the four matrices defined in (6)–(13). Each of these matrices has symmetric pairs of eigenvalues, i.e., for any i, $\exists j$ s.t. $\lambda_i(\Delta_p^\tau) = -\lambda_j(\Delta_p^\tau)$. Hence,

$$\lambda_{min}(-\beta_\tau\Delta_p^\tau) = -|\beta_\tau|\lambda_{max}(\Delta_p^\tau), \tag{171}$$
$$= -|\beta_\tau|\|\Delta_p^\tau\|_s. \tag{172}$$

□

Proof of Result 7: Prove by induction.

Basis: Comparing (75) and (80), we see that (85) holds for $p = 1$.

Induction: Assume $L_{p-1} > S_{p-1}$. From (74),

$$L_p = L_{p-1} - \sum_{\tau\in\Gamma_p}|\beta_\tau|\|\Delta_p^\tau\|_s \tag{173}$$
$$\ge L_{p-1} - \sum_{\tau\in\Gamma_p}|\beta_\tau|\|\Delta_p^\tau\|_\infty \quad \text{(because } \|\Delta_p^\tau\|_s \le \|\Delta_p^\tau\|_\infty\text{)}, \tag{174}$$
$$> S_{p-1} - \sum_{\tau\in\Gamma_p}|\beta_\tau|\|\Delta_p^\tau\|_\infty \quad \text{(using the induction hypothesis)} \tag{175}$$
$$= S_p \quad \text{(from (71))}. \tag{176}$$
$$\tag{177}$$

□

Proof of Lemma 9: Recall, from (86), that A_p is a scaled version of the inverse covariance matrix, i.e.,

$$A_p^{-1} = \frac{1}{\sigma^2}E\{\vec{X}\vec{X}^T\}. \tag{178}$$

Consider any τ in the interaction set η_p. As explained in Section IV-C, the matrix Δ_p^τ places β_τ, the coefficient for interaction τ, at the appropriate locations in A_p. Consequently, Δ_p^τ has 1's at these locations and zeroes everywhere else. With this in mind it is straightforward to verify that

$$\frac{1}{2}\text{trace}(A_p^{-1}\Delta_p^\tau) = \begin{cases} \frac{1}{\sigma^2}\sum_{r=1}^{N-i}\sum_{s=1}^{M}E\{x_{r,s}x_{r+i,s}\} & \text{for } \tau = v_i, \\ \frac{1}{\sigma^2}\sum_{r=1}^{N}\sum_{s=1}^{M-j}E\{x_{r,s}x_{r,s+j}\} & \text{for } \tau = h_j, \\ \frac{1}{\sigma^2}\sum_{r=1}^{N-i}\sum_{s=1}^{M-j}E\{x_{r,s}x_{r+i,s+j}\} & \text{for } \tau = ld_{ij}, \\ \frac{1}{\sigma^2}\sum_{r=1}^{N-i}\sum_{s=j+1}^{M}E\{x_{r,s}x_{r+i,s-j}\} & \text{for } \tau = rd_{ij}, \end{cases} \tag{179}$$

Interchanging the order of expectation and summation in (179), applying the definition of X_τ from (91), and substituting into (93), we get (97). □

REFERENCES

[1] K. Abend, T. J. Harley, and L. N. Kanal "Classification of binary random patterns," *IEEE Trans. Inform. Theory*, vol. IT-11, pp. 538–544, Oct. 1965.

[2] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice Hall, 1977.

[3] N. Balram, "A recursive framework for noncausal Gauss Markov random fields," Ph.D. thesis, Carnegie Mellon Univ., Pittsburgh, PA, Jan. 1992.

[4] J. E. Besag, "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *J. Roy. Statist. Soc., B*, vol. 36, no. 2, pp. 192–236, 1974.

[5] J. E. Besag, "Statistical analysis of non-lattice data, *The Statistician*, vol. 24, pp. 179–195, 1975.

[6] ——, "Errors–in–variable estimation for Gaussian lattice schemes," *J. R. Statist. Soc., B*, pp. 73–78, 1977.

[7] J. E. Besag and P. A. P. Moran, "On the estimation and testing of spatial interaction in Gaussian lattice processes," *Biometrika*, vol. 62, pp. 555–562, 1975.

[8] R. Chellappa, "Two-dimensional discrete Gaussian Markov random field models for image processing, in *Progress in Pattern Recognition 2*. Amsterdam: North-Holland, 1985, pp. 79–112.

[9] R. Chellappa and R. L. Kashyap, "Digital image restoration using spatial interaction models," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 30, pp. 461–472, June 1982.

[10] P. B. Chou and R. Raman, "On relaxation algorithms based on Markov random fields," Tech. Rep. 212, Dept. of Comput. Sci., Univ. of Rochester, July 1987.

[11] H. Derin and C. Won, "A parallel image segmentation algorithm using relaxation with varying neighbourhoods and its mapping to array processors," *Comput. Vision, Graphics, Image Process.*, vol. 40, pp. 54–78, 1987.

[12] C. F. Fischer and R. A. Usmani, "Properties of some tridiagonal matrices and their application to boundary value problems," *Siam J. Numer. Anal.*, vol. 6, no. 1, pp. 127–142, Mar. 1969.

[13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 721–741, Nov. 1984.

[14] A. Graham, *Kronecker Products and Matrix Calculus With Applications*. Chichester, England: Ellis Horwood, Ltd., 1981.

[15] S. L. Handy and J. L. Barlow, "Eigenvalue bounds for banded, symmetric Toeplitz Matrices," Tech. Rep. CS-88-19, Dept. Comput. Sci., Pennsylvania State Univ., May 1988.

[16] M. Hassner and J. Sklansky, "The use of Markov random fields as models of texture," *Computer Graphics and Image Processing*, vol. 12, pp. 357–370, 1980.

[17] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice Hall, 1989.

[18] ——, "A sinusoidal family of unitary transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-1, pp. 356–365, Oct. 1979.

[19] H. Jinchi and R. Chellappa, "Restoration of blurred and noisy images using Gaussian Markov random field models," in *20th Ann. Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1986, pp. 34–39.

[20] R. L. Kashyap, "Finite lattice random field models for finite images," in *15th Ann. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 1981, pp. 215–220.

[21] H. Kaufman, J. W. Woods, S. Dravida, and A. M. Tekalp, "Estimation and identification of two-dimensional images," *IEEE Trans. Automat. Cont.*, vol. AC-28, no. 7, pp. 745–756, July 1983.

[22] S. Kirkpatrick, C. D. Gellat, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, pp. 671–680, 1983.

[23] S. Lakshmanan and H. Derin, "Parameter space for Gaussian Markov random fields with separable autocorrelation," in *24th Ann. Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1990, pp. 226–231.

[24] ——, "Necessary and sufficient conditions for the parameters of Gaussian Markov randon fields," presented at the *25th Ann. Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 1991.

[25] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: John Wiley, 1988.

[26] J. Marroquin, S. K. Mitter, and T. Poggi, "Probabilistic solution of ill-posed problems in computational vision," *J. Amer. Statist. Soc.*, vol. 82, pp. 76–89, 1987.

[27] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Phys. Chem.*, vol. 21, no. 6, pp. 1087–1091, 1953.

[28] P. A. P. Moran, "A Gaussian Markovian process on a square lattice," *J. Appl. Probab.*, vol. 10, pp. 54–62, 1973.

[29] J. M. F. Moura and N. Balram, "A recursive framework for noncausal Gauss Markov random fields," in *24th Ann. Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1990, pp. 608–613.

[30] ——, "Recursive structure of noncausal Gauss–Markov random fields," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 334–354, Mar. 1992.

[31] J. O'Brien, Ed., *Advanced Physical Oceanographic Numerical Modeling*. New York: D. Reidel, 1986.

[32] K. Ord, "Estimation methods for models of spatial interaction," *J. Amer. Statist. Assoc.*, vol. 70, no. 349, pp. 120–126, Mar. 1975.

[33] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge, MA: Cambridge Univ. Press, 1989.

[34] T. J. Rivlin, *The Chebyshev Polynomials*. New York: Wiley-Interscience, 1974.

[35] G. Sharma and R. Chellappa, "A model-based approach for estimation of two-dimensional maximum entropy power spectra," *IEEE Trans. Inform. Theory*, vol. IT-31, no. 1, pp. 90–99, Jan. 1985.

[36] T. Simchony, R. Chellappa, and Z. Lichtenstein, "Relaxation algorithms for MAP estimation of gray–level images with multiplicative noise," *IEEE Trans. Inform. Theory*, vol. 36, no. 3, pp. 608–613, May 1990.

[37] R. S. Varga, *Matrix Iterative Analysis*. Englewood Cliffs, NJ: Prentice Hall, 1962.

[38] J. W. Woods and C. H. Radewan, "Kalman filtering in two dimensions," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 4, pp. 473–481, July 1977.