Recursive Structure of Noncausal Gauss-Markov Random Fields

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

Abstract—**Causality is a common assumption with one-dimensional (1-D) Markov random processes. It leads to recursive descriptions and recursive filtering algorithms, such as the Kalman-Bucy filter. However, in 2-D, e.g., in physical oceanography or in image processing, noncausality is an important property, leading to Markov random field (MRF) models. To apply recursive techniques to MRF models, many authors work with subclasses that are causal or, more generally, unilateral, thus compromising an important property of the model. An alternate approach for noncausal Gauss-Markov random fields (GMRF) that enables the use of recursive procedures while retaining the noncausality of the field is developed. Recursive representations are established that are equivalent to the original field. This is achieved by first presenting a canonical representation for GMRF's that is based on the inverse of the covariance matrix, which we call the potential matrix. It is this matrix rather than the field covariance that reflects in a natural way the MRF structure. From its properties, two equivalent one-sided representations are derived, each of which is obtained as the successive iterates of a Riccati type equation. For homogeneous fields, these unilateral descriptions are symmetrized versions of each other, the study of only one Riccati equation being required. It is proven that this Riccati equation converges at a geometric rate, therefore the one-sided representations are asymptotically invariant. These unilateral representations enable us to process the fields with well-known recursive techniques such as Kalman-Bucy filters and two-point smoothers.**

Index Terms—**Noncausal, random fields, Gauss-Markov random fields, potential matrix, equivalent one-sided representations, recursive 2-D processing, image processing.**

I. Introduction

In one-dimensional (1-D) signal processing, Markov random processes are commonly assumed models. They capture the natural assumption of causality in time dependent signals and lead to recursive processing algorithms of the Kalman-Bucy type. When describing spatially dependent phenomena, as for example in mapping temperature or salinity fields in physical oceanography or in 2-D image processing, noncausality is a more appropriate characteristic than causality. The noncausal Markov property applied to 2-D fields produces the noncausal Markov random field (MRF) model, see, e.g., [11]. However, noncausality is conducive to iterative processing algorithms and not to recursive algorithms. To recover recursive algorithms for 2-D fields, authors frequently enforce a causal type 2-D Markov property. Examples of these models include Markov mesh fields [1], Pickard random fields [25], mutually compatible Gibbs random fields [12], or unilateral recursive autoregressive fields, e.g., [14], [33].

In this paper, we readdress the issue of recursiveness for 2-D signal processing, but do so while retaining the noncausality property for the 2-D fields. We consider noncausal Gauss-Markov random fields (GMRF's) defined on finite lattices. The key result is the establishment of two recursive spatially varying field formulations that are equivalent to the original noncausal Markov field, and, furthermore, in the case of homogeneous fields, are asymptotically spatially invariant.

The success of the present approach is a consequence of a shift of focus. We characterize the field, not by its covariance, but rather by the inverse of the covariance matrix, which we term the potential matrix. Except in very special cases, like for example fields on a torus (e.g., [5], [17], [9]), separable fields, or binary valued Ising fields (see the work of Kaufman and Onsager [18]), the covariance of noncausal MRF's is difficult if not impossible to parametrize directly.

By taking the alternate point of view of focusing on the inverse of the covariance matrix, we show how GMRF's are easily described by the potential matrix. We present the canonical form of the potential matrix of any GMRF. This matrix is highly sparse, its structure reflecting the order of the field and the type of boundary conditions assumed.

The recursive formulation intrinsic to noncausal GMRF's (which of course include all unilateral classes) is obtained by the Cholesky factorization of the general potential matrix associated with the field. It comprises of two one-sided representations, each of which is statistically equivalent to the original noncausal one. These are derived from the noncausal autoregressive field representation of Woods [32]. For homogeneous fields, these two representations are mirror images of one another, and hence, only one of them needs to be computed explicitly. These representations are computed by the successive iterations of a Riccati equation. We show that under certain conditions the iteration procedure converges at a geometric rate.

Conceptually, this means that every noncausal field defined on a finite lattice is equivalently represented by one-sided fields. These fields are described by finite-order spatially varying unilateral regressors that are asymptotically invariant

Manuscript received June 26, 1990. This work was supported in part by ONR Grant #N00014-91-J-1001 and in part by NSF Equipment Grant CDA-8820575. A draft of this work was partially written while J.M.F. Moura was visiting the Institute for Mathematics and Its Applications under NSF sponsorship. This work was presented in part at the 24th Annual Conference on Information Science and Systems, Princeton, NJ, March 22, 1990.

The authors are with the Department of Electrical and Computer Engineering, Carnegie-Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213-3890.

IEEE Log Number 9104375.

for homogeneous fields. From a practical point of view, it enables the optimal processing of noncausal 2-D GMRF's by recursive techniques such as Kalman–Bucy filtering and two-point smoothers. In addition, it provides the means for fast recursive synthesis of noncausal field samples.

The organization of the paper is as follows. To facilitate the discussion, we defer all but the most elementary proofs to the appendix. In Section II, we define terminology and notation that will be used throughout. The canonical representation of GMRF's in terms of the potential matrix is studied in Section III. The recursive structure of these fields is derived in Section IV. It is shown here that for homogeneous fields the Riccati equation converges and the spatially varying regressors are asymptotically invariant. Application of these results to first- and second-order fields is carried out in Section V. In Section VI, the two equivalent one-sided representations are used to formulate "backward" and "forward" state space descriptions. Section VII briefly indicates how these can be utilized to do recursive smoothing of noisy fields. Finally, in Section VIII, conclusions are presented.

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 a Euclidean distance based measure, as is done, for example, in [11], [15].

Definition 1: A Pth order neighborhood is defined on an $N \times M$ lattice, L, by $S_P^{i,j}$, the set of neighbors of site (i, j), where

$$S_P^{i,j} = \left\{(k, l): 0 < (i-k)^2 + (j-l)^2 \le D_P\right\}, \quad (1)$$

and D_P is an increasing function of P that represents the square of the Euclidean distance between a site and its furthest neighbor.

Usually (see, for example, [11], [15]), 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 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 $P-1$.

Defining N_v, and H_h, respectively, as the number of vertical and horizontal neighbors for a given neighborhood set, we have

$$N_v = N_h = 2\left\lfloor D_P^{1/2} \right\rfloor, \quad (2)$$

where $\lfloor \cdot \rfloor$ is the floor function. This relationship comes from the Euclidean distance based definition of the neighborhood set (see (1)). In addition, because of the symmetry of this definition, we can define

$$k_v = \frac{N_v}{2} = \left\lfloor D_P^{1/2} \right\rfloor, \quad (3)$$

as the number of neighbors south (or north) of any site in the interior of the lattice, and similarly for k_h, the number of east (or west) neighbors.

For a finite lattice, special rules have to be specified to extend the definition of neighborhood order in Definition 1 to the "boundary sites." Here, the so-called boundary sites are

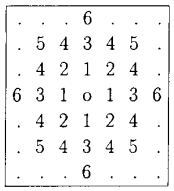

Fig. 1. Hierarchical sequence of neighborhoods.

those that would have one or more neighbors from outside the lattice, if the same neighborhood rule defined for the MRF is applied. These special rules are called *boundary conditions*. In the rest of this paper, we assume the boundary conditions satisfy the following definition.

Definition 2: The boundary conditions are defined to be consistent with the order of the field, if for a Pth order field the rule used to assign the off-lattice neighbors of a boundary site does not assign it sites that lie outside a circle of radius $D_P^{1/2}$.

For any matrix $R = [r_{i,j}]$, define the left bandwidth as $b_l(R) = \max_i \{|i-j|: j \le i, \text{ and } r_{i,j} \ne 0\}$, the right bandwidth as $b_r(R) = \max_i \{|i-j|: j \ge i, \text{ and } r_{i,j} \ne 0\}$. If the matrix is symmetric, we define its bandwidth as $b(R) = b_l(R) = b_r(R)$.

Symbol $\otimes$ represents the Kronecker product (see, for example, [13]), while $\otimes_j$ represents a modified Kronecker product that handles varying matrices, e.g., $A \otimes_j B_i$ means that matrix B_i multiples the entries of row j of matrix A. The vector $\vec{e}_i$ is the ith unit vector,

$$K_N = \begin{bmatrix} 0 & 0 & \cdot & \cdot & \cdot \\ 1 & 0 & 0 & \cdot & \cdot \\ 0 & 1 & 0 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & 1 & 0 \end{bmatrix}, \quad H_N = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot \\ 1 & 0 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & 1 & 0 & 1 \\ \cdot & \cdot & 0 & 1 & 0 \end{bmatrix},$$

$$I_{1,N} = \begin{bmatrix} 1 & 0 & 0 & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & 0 & 0 & 1 \end{bmatrix}, \quad J_N = \begin{bmatrix} 0 & 0 & \cdot & 0 & 1 \\ 0 & 0 & \cdot & 1 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 1 & 0 & \cdot & \cdot \\ 1 & 0 & \cdot & \cdot & \cdot \end{bmatrix},$$

i.e., $K_N = [\vec{e}_2 \cdots \vec{e}_N \vec{0}]$, $H_N = K_N + K_N^T$, $I_{1,N} = \vec{e}_1\vec{e}_1^T + \vec{e}_N\vec{e}_N^T$, and $J_N = [\vec{e}_N \cdots \vec{e}_1]$, is the reflection matrix. Define $E_{i,j} = \vec{e}_i\vec{e}_j^T$, i.e., it has element (i, j) equal to 1 while all the rest are zeros, and let $F_{i,j} = E_{i,j} + E_{i,j}^T$. Let $\mathbf{0}_N$ be an $N \times N$ matrix of zeros, and I_N the N-dimensional identity matrix. When the dimensions are obvious from the context, we will use J, $\mathbf{0}$, and I, without any subscripts, for the reflection, zero, and identity matrices respectively.

The backward difference is $\Delta^- x_i = x_i - x_{i-1}$, the forward difference $\Delta^+ x_i = x_{i+1} - x_i$, and the symmetric difference $\Delta_{\text{sym}} x_i = (\Delta^- x_i + \Delta^+ x_i)/2$.

III. Canonical Representation of Noncausal Gauss–Markov Random Fields

Consider a random field X, defined on an $N \times M$ lattice, L. The lattice sites can be ordered so as to form an $NM \times 1$ vector of random variables. For convenience, we adopt lexicographic ordering and will refer to this as the canonical

ordering. Any other ordering is the result of premultiplication by a permutation matrix. Lexicographic ordering maps each row of the lattice into an $M \times 1$ random vector and these are stacked one on top of the other to form the $NM \times 1$ vector, $\vec{X} = [\vec{x}_1^T, \vec{x}_2^T, \cdots, \vec{x}_N^T]^T$, where $\vec{x}_i = [x_{i1}, x_{i2}, \cdots, x_{iM}]^T$. Boundary conditions are given with respect to bordering sites, lying outside the lattice, collected in row vectors $\vec{x}_0$, $\vec{x}_{N+1}$, and in column vectors $\vec{y}_0$ and $\vec{y}_{M+1}$.

A. Canonical Representation

Hammersley Clifford Theorem: The Markov–Gibbs equivalence provides the joint distribution of a Markov random field as a Gibbs distribution:

$$P(X) = \frac{1}{Z} \exp\left(-U(X)\right), \tag{4}$$

where Z is a normalization constant, called the partition function, and $U(X)$ the sum of the field interactions, is known as the energy function. This theorem is attributed to Hammersley and Clifford by Besag [4] (see also Spitzer [28], Averintsev [2]), and proved in [4] for discrete valued fields. We refer to it here as the Hammersley–Clifford theorem (HCT). Using Brook's factorization theorem [6] and the fact that Bayes' law has exactly the same functional form for distributions and probability density functions (pdf's), it is straightforward to show that the HCT holds for continuously valued random fields, where, in (4), $P(X)$ is now interpreted as a pdf and the right hand side is assumed to be integrable.

Quadratic Fields: Consider zero-mean fields that have only pairwise neighbor interactions (e.g., $x_{i,j}x_{k,l}$) and second-order self interactions (e.g., $x_{i,j}^2$). We call such fields *quadratic* fields. For these fields, the exponent of the Gibbs distribution can be written in a compact form:

$$2\sigma^2 U(X) = \vec{X}^T A \vec{X}, \tag{5}$$

where σ is a positive constant, and A is an $NM \times NM$ matrix that contains all the field *potentials* (i.e., the field interaction parameters); hence, we call it the *potential* matrix. In the particular case of Gauss–Markov random fields (GMRF), the potential matrix, A, is a scaled version of the inverse covariance matrix. Here, we consider nondegenerate Gauss–Markov fields, i.e., those with positive definite (pd, or >0) covariance matrices. With the exception of special cases, such as fields on a torus [24], [17], specification of the covariance structure is difficult, and correspondingly it is difficult to characterize a field of given order through its covariance matrix. However, by looking alternatively to the potential matrix, it becomes easy to characterize a GMRF of any order, by using the structure and properties of its potential matrix. Since a nonzero-mean does not affect the field covariance, without loss of generality, we consider here only zero-mean GMRF's.

Canonical Structure: The *potential* matrix is characterized by highly structured sparseness. The properties of this matrix are defined in a theorem stated next. For clarity, the theorem and the discussion that follows it are stated in terms of the simplest case, the first order model (with neighbors to the north, south, west, and east, see Fig. 1), but is generalized later to higher order models. The theorem assumes the canonical ordering and consistent boundary conditions.

Theorem 1: The matrix

$$A = A^T > 0 \tag{6}$$

is the potential matrix of a zero-mean (nondegenerate) first-order Gauss–Markov random field, if and only if it is decomposable as

$$A = A_c + A_{\text{b.c.}}, \tag{7}$$

where, up to reordering,

1) A_c, the canonical potential matrix, is independent of the boundary conditions, and has the following structure:

$$A_c = \begin{bmatrix} \tilde{B}_1 & \tilde{C}_1 & \mathbf{0} & \cdot & \cdot & \cdot \\ \tilde{C}_1 & \tilde{B}_2 & \tilde{C}_2 & \mathbf{0} & \cdot & \cdot \\ \mathbf{0} & \tilde{C}_2 & \tilde{B}_3 & \tilde{C}_3 & \mathbf{0} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \mathbf{0} & \tilde{C}_{N-2} & \tilde{B}_{N-1} & \tilde{C}_{N-1} \\ \cdot & \cdot & \cdot & \mathbf{0} & \tilde{C}_{N-1} & \tilde{B}_N \end{bmatrix}$$

$$= I_N \otimes_i \tilde{B}_i + K_N \otimes_{i+1} \tilde{C}_i + K_N^T \otimes_i \tilde{C}_i, \tag{8}$$

where

$$\tilde{B}_i = \begin{bmatrix} \alpha^{i,1} & -\beta_h^{i,1} & 0 & \cdot & \cdot \\ -\beta_h^{i,1} & \alpha^{i,2} & -\beta_h^{i,2} & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & -\beta_h^{i,M-2} & \alpha^{i,M-1} & -\beta_h^{i,M-1} \\ \cdot & \cdot & 0 & -\beta_h^{i,M-1} & \alpha^{i,M} \end{bmatrix}$$

$$= I_M \otimes_j \alpha^{i,j} - K_M \otimes_{j+1} \beta_h^{i,j} - K_M^T \otimes_j \beta_h^{i,j} \tag{9}$$

$$\tilde{C}_i = -\operatorname{diag}\left(\beta_v^{i,1}, \beta_v^{i,2}, \cdots, \beta_v^{i,M}\right). \tag{10}$$

2) $A_{\text{b.c.}}$, the boundary potential matrix, is specified by the choice of boundary conditions and has the following structure:

$$A_{\text{b.c.}} = \begin{bmatrix} B_{\text{b.c.}}^1 & C_{\text{b.c.}}^1 & \mathbf{0} & \cdot & \mathbf{0} & C_{\text{b.c.}}^N \\ C_{\text{b.c.}}^1 & B_{\text{b.c.}}^2 & \mathbf{0} & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & B_{\text{b.c.}}^3 & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{0} & \cdot & \cdot & \mathbf{0} & B_{\text{b.c.}}^{N-1} & C_{\text{b.c.}}^{N-1} \\ C_{\text{b.c.}}^N & \mathbf{0} & \cdot & \mathbf{0} & C_{\text{b.c.}}^{N-1} & B_{\text{b.c.}}^N \end{bmatrix}$$

$$= I_N \otimes_i B_{\text{b.c.}}^i + F_{2,1} \otimes C_{\text{b.c.}}^1 + F_{N,N-1} \otimes C_{\text{b.c.}}^{N-1} + F_{N,1} \otimes C_{\text{b.c.}}^N, \tag{11}$$

where, for $i = 1, N$,

$$B^i_{\text{b.c.}} = \begin{bmatrix} \alpha^{i,1}_{\text{b.c.}} & -\beta^{i,1}_{h,\text{b.c.}} & 0 & \cdot & 0 & -\beta^{i,M}_{h,\text{b.c.}} \\ -\beta^{i,1}_{h,\text{b.c.}} & \alpha^{i,2}_{\text{b.c.}} & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \alpha^{i,3}_{\text{b.c.}} & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & \alpha^{i,M-1}_{\text{b.c.}} & -\beta^{i,M-1}_{h,\text{b.c.}} \\ -\beta^{i,M}_{h,\text{b.c.}} & 0 & \cdot & \cdot & -\beta^{i,M-1}_{h,\text{b.c.}} & \alpha^{i,M}_{\text{b.c.}} \end{bmatrix}$$

$$= I_M \otimes_j \alpha^{i,j}_{\text{b.c.}} - \beta^{i,1}_{h,\text{b.c.}} F_{2,1} - \beta^{i,M-1}_{h,\text{b.c.}} F_{M,M-1} - \beta^{i,M}_{h,\text{b.c.}} F_{1,M}, \tag{12}$$

for $i \neq 1, N$,

$$B^i_{\text{b.c.}} = \begin{bmatrix} \alpha^{i,1}_{\text{b.c.}} & -\beta^{i,1}_{h,\text{b.c.}} & 0 & \cdot & 0 & -\beta^{i,M}_{h,\text{b.c.}} \\ -\beta^{i,1}_{h,\text{b.c.}} & 0 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 0 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & 0 & 0 & -\beta^{i,M-1}_{h,\text{b.c.}} \\ -\beta^{i,M}_{h,\text{b.c.}} & 0 & \cdot & 0 & -\beta^{i,M-1}_{h,\text{b.c.}} & \alpha^{i,M}_{\text{b.c.}} \end{bmatrix}$$

$$= I_{1,M} \otimes_j \alpha^{i,j}_{\text{b.c.}} - \beta^{i,1}_{h,\text{b.c.}} F_{2,1} - \beta^{i,M-1}_{h,\text{b.c.}} F_{M,M-1} - \beta^{i,M}_{h,\text{b.c.}} F_{1,M} \tag{13}$$

and

$$C^i_{\text{b.c.}} = -\operatorname{diag}\left(\beta^{i,1}_{v,\text{b.c.}}, \beta^{i,2}_{v,\text{b.c.}}, \cdots, \beta^{i,M}_{v,\text{b.c.}}\right), \quad i = 1, N-1, N. \tag{14}$$

Remarks: Theorem 1 shows that the canonical potential matrix, A_c, and the boundary potential matrix, $A_{\text{b.c.}}$, have the following properties.

1) They are symmetric and highly sparse, with A_c being block tridiagonal, while $A_{\text{b.c.}}$ has a maximum of five nonzero block diagonals.
2) The nonzero blocks are symmetric, banded, and highly sparse. In particular, the diagonal blocks replicate the structure of the corresponding matrix itself, being symmetric, tridiagonal, for A_c, and symmetric, and sparse with a maximum of five nonzero diagonals, for $A_{\text{b.c.}}$. For either matrix, the nonzero off-diagonal blocks are diagonal.
3) For A_c, the number of nonzero diagonals $N_d = N_P + 1$, where N_P is the number of sites in the neighborhood set.

Homogeneous Quadratic Fields: An MRF with potentials that are independent of the site locations is 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. For a homogeneous first-order GMRF, let β_v and β_h, represent the vertical and horizontal field interactions, respectively. Then the following corollary to Theorem 1 holds.

Corollary 1: The matrix

$$A = A^T > 0 \tag{15}$$

is the potential matrix of a zero mean (nondegenerate) homogeneous first-order Gauss–Markov random field, if and only if its canonical potential matrix, A_c, has the following structure:

$$A_c = \begin{bmatrix} \tilde{B} & \tilde{C} & \mathbf{0} & \cdot & \cdot & \cdot \\ \tilde{C} & \tilde{B} & \tilde{C} & \mathbf{0} & \cdot & \cdot \\ \mathbf{0} & \tilde{C} & \tilde{B} & \tilde{C} & \mathbf{0} & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \mathbf{0} & \tilde{C} & \tilde{B} & \tilde{C} \\ \cdot & \cdot & \cdot & \mathbf{0} & \tilde{C} & \tilde{B} \end{bmatrix} = I_N \otimes \tilde{B} + H_N \otimes \tilde{C}, \tag{16}$$

where

$$\tilde{B} = \begin{bmatrix} 1 & -\beta_h & 0 & \cdot & \cdot & \cdot \\ -\beta_h & 1 & -\beta_h & 0 & \cdot & \cdot \\ 0 & -\beta_h & 1 & -\beta_h & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 0 & -\beta_h & 1 & -\beta_h \\ \cdot & \cdot & \cdot & 0 & -\beta_h & 1 \end{bmatrix} = I_M - \beta_h H_M, \tag{17}$$

$$\tilde{C} = -\beta_v I_M, \tag{18}$$

and the boundary potential matrix, $A_{\text{b.c.}}$, is given by (11)–(14) with

$$C^i_{\text{b.c.}} = \begin{cases} -\beta^1_{v,\text{b.c.}} I_M, & i = 1, N-1, \\ -\beta^2_{v,\text{b.c.}} I_M, & i = N, \end{cases} \tag{19}$$

$$B^i_{\text{b.c.}} = \begin{cases} B^1_{\text{b.c.}}, & \text{for } i = 1, N, \\ B^2_{\text{b.c.}}, & \text{for } i \neq 1, N, \end{cases} \tag{20}$$

where $B^1_{\text{b.c.}}$ and $B^2_{\text{b.c.}}$ have the structure given in (12), and (13), respectively, with

$$\alpha^{1,j}_{\text{b.c.}} = \begin{cases} \alpha^{1,1}_{\text{b.c.}}, & \text{for } j = 1, M, \\ \alpha^{1,2}_{\text{b.c.}}, & \text{for } j \neq 1, M, \end{cases} \tag{21}$$

$$\alpha^{2,j}_{\text{b.c.}} = \alpha^2_{\text{b.c.}}, \quad \text{for } j = 1, M, \tag{22}$$

$$\beta^{i,j}_{h,\text{b.c.}} = \begin{cases} \beta^1_{h,\text{b.c.}}, & \text{for } i = 1, j = 1, M-1, \\ \beta^2_{h,\text{b.c.}}, & \text{for } i = 2, j = 1, M-1, \\ \beta^3_{h,\text{b.c.}}, & \text{for } i = 1, 2, j = M. \end{cases} \tag{23}$$

Remarks: For a homogeneous, first order, nondegenerate GMRF, A_c and $A_{\text{b.c.}}$ have the following properties.

1) They are centrosymmetric. In addition, A_c is block tridiagonal, and block Toeplitz.
2) The blocks along the main diagonal replicate the structure of the corresponding matrix, being centrosymmetric in both cases, and in the case of A_c, tridiagonal and Toeplitz, as well. The nonzero off-diagonal blocks are diagonal Toeplitz matrices.

B. Boundary Conditions

In this subsection, we look at examples of first-order GMRF's considered in the literature, and show how they are captured by Theorem 1 through a suitable choice of the boundary potential matrix $A_{\text{b.c.}}$. The boundary conditions are usually imports from the PDE literature. Examples include the following.

1) *Free or Dirichlet b.c.:* Here, the field values of the off-lattice neighbors of a boundary pixel are set to zero (or to some deterministic constant for a GMRF that is not zero-mean). The boundary pixels simply have fewer neighbors. This is the simplest boundary condition because it results in

$$A_{\text{b.c.}} = 0, \tag{24}$$

$$A = A_c, \tag{25}$$

which shows that A_c is a potential matrix in its own right. These boundary conditions lead to the *autonormal* models considered in [4]. Since this term was later applied to more general GMRF's, henceforth, to avoid confusion, we will refer to fields with Dirichlet boundary conditions as *Dirichlet* fields.

2) *Neumann b.c.:* Here, we derive the boundary condition by assuming zero Neumann b.c.'s. This means that the field gradient normal to the boundary is constant at the boundary. Depending on the type of difference operator used to define the gradient, different cases are obtained. Two of these are the following.

   a) *Asymmetric Neumann b.c.:* Here, the gradient is defined by a backward difference operator. For a zero-mean field, the constant gradient value at the boundary is set to zero, so that the off-lattice neighbors of a boundary pixel are assumed to have the same intensity as the pixel. As an illustration, for example, at the top row of the image, it follows that

$$\Delta^- x_1 = x_1 - x_0 = 0. \tag{26}$$

   Collecting (26) and the corresponding relation at the bottom, left and right edges of the lattice, the first-order, homogeneous case leads to

$$A_{\text{b.c.}} = I_N \otimes (-\beta_h I_{1,M}) + I_{1,N} \otimes (-\beta_v I_M). \tag{27}$$

   We will call this the *variational* field. A field with similar potentials has been used by Marroquin, Mitter, and Poggio [22] to represent quadratic intensities.

   b) *Symmtric Neumann b.c.:* Here, the symmetric difference operator is used when defining the gradient. For a zero-mean field, the constant gradient value at the boundary is set to zero, this means that 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). As an illustration, for example, at the top row of the image:

$$\Delta_{\text{sym}} x_1 = (\Delta^- x_1 + \Delta^+ x_1)/2 = (x_2 - x_0)/2 = 0. \tag{28}$$

   In the first-order, homogeneous case, this leads to

$$A_{\text{b.c.}} = I_N \otimes B^1_{\text{b.c.}} + (F_{1,2} + F_{N-1,N}) \otimes (-\beta_v I_M), \tag{29}$$

   where

$$B^1_{\text{b.c.}} = I_M - \beta_h (F_{1,2} + F_{M-1,M}). \tag{30}$$

   We will call this the *symmetric* field. These boundary conditions were mentioned by Kashyap [17], in the context of white noise driven fields.

3) *Periodic or Cyclic b.c.:* Here, the lattice is assumed to be a torus. Thus, the off-lattice neighbors of the boundary pixels on one side of the image are simply the appropriate boundary pixels from the other side of the image, and vice-versa. In the first-order, homogeneous case this leads to

$$A_{\text{b.c.}} = I_N \otimes (-\beta_h F_{1,M}) + F_{1,N} \otimes (-\beta_v I_M). \tag{31}$$

   These fields have been used by many authors, e.g., Moran [24], Besag and Moran [5], Kashyap [17], and Chellappa and Kashyap [9].

C. Generalization to Fields of Any Order

Theorem 1 and Corollary 1 are easily extended to fields of any order. Since in the rest of the paper we are concerned primarily with homogeneous fields, the discussion that follows is centered around these fields. We decompose the potential matrix A as in equation (7), i.e., as the sum of the canonical potential matrix A_c, and the boundary potential matrix $A_{\text{b.c.}}$. We consider first A_c, then $A_{\text{b.c.}}$, and finally the structure of A.

Structure of A_c: The structure and properties of the canonical potential matrix for a homogeneous first-order field were given in Corollary 1. These are extended to homogeneous fields of arbitrary order P in Theorem 2. The theorem assumes the canonical ordering.

Theorem 2: The canonical potential matrix of a zero mean (nondegenerate) homogeneous GMRF of order P is independent of the boundary conditions and has the following structure:

$$A_c = \begin{bmatrix} A_0 & A_1 & A_2 & \cdots & A_{k_v} & \mathbf{0} & \cdots & & \\ A_1^T & A_0 & A_1 & A_2 & \cdots & A_{k_v} & \mathbf{0} & \cdots & \\ A_{k_v}^T & A_{k_v+1}^T & \cdots & A_0 & A_1 & \cdots & A_{k_v} & \mathbf{0} & \cdots \\ \mathbf{0} & A_{k_v}^T & A_{k_v-1}^T & \cdots & A_0 & A_1 & \cdots & A_{k_v} & \cdots \\ & \cdots & \mathbf{0} & A_{k_v}^T & \cdots & A_{k_v-1}^T & \cdots & A_2^T & A_1^T & A_0 \end{bmatrix}, \tag{32}$$

where k_v is given by (3), and the $M \times M$ blocks A_{i-j} have the following properties.

1) $A_{i-j}^T = JA_{i-j}J$, i.e., A_{i-j} are persymmetric.
2) $A_0 = A_0^T$, and $A_0^T = JA_0J$, i.e., A_0 is centrosymmetric.
3) $A_{i-j} = [a_{k,l}^{(i-j)}] = [a_{k-l}^{(i-j)}]$, with $a_{k-l}^{(i-j)} = 0$ for $(i-j)^2 + (k-l)^2 > D_P$, i.e., A_{i-j} are Toeplitz and banded.

The proof is straightforward and follows from the fact that the canonical potential matrix A_c is the potential matrix of a field for which all the off-lattice neighbors of the boundary pixels have been set to zero, i.e., a Dirichlet field. With this in mind, A_c can be partitioned into N rows of $M \times M$ blocks, where the (i, j)th block, $A_{i,j}$ represents the neighbor interactions between the pixels in row i and those in row j. The symmetry of the neighborhood definition means that $A_{j,i}^T = A_{i,j}$. Furthermore, pixels in row i do not have any neighbors in any row j that is more than k_v rows away, therefore,

$$A_{i,j} = \mathbf{0}, \qquad \text{for } |i-j| > k_v.$$

For homogeneous fields it is immediately apparent that $A_{i,j} = A_{i-j}$. The symmetry properties of the blocks A_{i-j} follow immediately as a consequence of homogeneity, as does the fact that these blocks are Toeplitz. Lastly, the banded nature of the blocks and the extent of their nonzero band follows from the definition of neighborhood order P (see Definition 1).

Remarks: From Theorem 2, A_c has the following properties.

1) It is centrosymmetric, block Toeplitz, and block banded, with block bandwidth k_v.
2) The blocks along the main diagonal replicate the structure of A_c, being centrosymmetric, Toeplitz, and banded. The off-diagonal blocks are persymmetric, Toeplitz, and banded.
3) The number of nonzero block diagonals of A_c is given by $N_{bd} = N_v + 1$.
4) The number of nonzero diagonals of A_c is given by $N_d = N_P + 1$, where N_P is the total number of neighbors in the neighborhood set S_P.
5) The bandwidth of A_c is given by

$$b(A_c) = k_v M + b_r(A_{k_v}), \tag{33}$$

where

$$b_r(A_{k_v}) = \left(D_P - \left\lfloor D_P^{1/2} \right\rfloor^2\right)^{1/2}. \tag{34}$$

As an example, note that for homogeneous second order fields, from (3) and Fig. 1, $k_v = 1$; therefore the canonical potential matrix is block tridiagonal as for first-order fields,

$$A_c = I_M \otimes A_0 + K_N \otimes A_1^T + K_N^T \otimes A_1, \tag{35}$$

where A_0 has the same structure as the corresponding matrix, given by (17), for first-order fields, but the off-diagonal blocks are now tridiagonal,

$$A_1 = -\left(\beta_v I_M + \beta_{d1} K_M^T + \beta_{d2} K_M\right). \tag{36}$$

Here, β_v and β_h represent the vertical and horizontal interactions, respectively (as was the case for the first-order model), while β_{d1}, and β_{d2} represent the NW-SE, and the NE-SW, diagonal interactions, respectively.

The canonical potential matrix given by (32) can be repartitioned to cast it into the form of a block tridiagonal matrix, as in (16), with N_1 rows of $M_1 \times M_1$ blocks, where

$$M_1 = k_v M, \qquad N_1 = N/k_v, \tag{37}$$

with k_v being defined in (3). For first- and second-order fields, from (3) and Fig. 1, $k_v = 1$. For higher order fields, to facilitate the discussion, we assume that the factoring in (37) is possible and so ignore possible end effects in the

repartitioning technique when this is not true. Then

$$A_c = \begin{bmatrix} \tilde{B} & \tilde{C} & \mathbf{0} & \cdot & \cdot & \cdot & \cdot \\ \tilde{C}^T & \tilde{B} & \tilde{C} & \mathbf{0} & \cdot & \cdot & \cdot \\ \mathbf{0} & \tilde{C}^T & \tilde{B} & \tilde{C} & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \mathbf{0} & \tilde{C}^T & \tilde{B} & \tilde{C} & \mathbf{0} \\ \cdot & \cdot & \cdot & \mathbf{0} & \tilde{C}^T & \tilde{B} & \tilde{C} \\ \cdot & \cdot & \cdot & \cdot & \mathbf{0} & \tilde{C}^T & \tilde{B} \end{bmatrix}, \quad (38)$$

with the $M_1 \times M_1$ blocks, $\tilde{B}$ and $\tilde{C}$, given by

$$\tilde{B} = \begin{bmatrix} A_0 & A_1 & A_2 & \cdot & \cdot & A_{k_v-1} \\ A_1^T & A_0 & A_1 & \cdot & \cdot & A_{k_v-2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ A_{k_v-1}^T & A_{k_v-2}^T & \cdot & \cdot & A_1^T & A_0 \end{bmatrix},$$

$$\tilde{C} = \begin{bmatrix} A_{k_v} & \mathbf{0} & \mathbf{0} & \cdot & \cdot \\ A_{k_v-1} & A_{k_v} & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ A_1 & A_2 & \cdot & \cdot & A_{k_v} \end{bmatrix}. \quad (39)$$

From the properties of the blocks A_{i-j}, it follows that

1) $\tilde{B}$ is centrosymmetric, i.e., $\tilde{B} = \tilde{B}^T$ and $\tilde{B} = J\tilde{B}J$;
2) $\tilde{C}$ is persymmetric, i.e., $\tilde{C}^T = J\tilde{C}J$.

The repartitioning of the canonical potential matrix into block tridiagonal structure has the effect of regrouping the rows of the lattices into larger entities, which we call *pseudo rows*, that have a first-order neighborhood relationship, i.e., all the pixels in a pseudo row depend only on pixels from the pseudo rows that lie immediately above and below it. Then, from (37), the number of lattice rows that form a pseudo row is k_v, and the number of pseudo rows is N_1. Repartitioning is purely conceptual, i.e., no changes are made to A_c; we simply look at the block structure in a different way that, as we will see in the next section, is more convenient.

Structure of $A_{b.c.}$: The boundary conditions are assumed to be *consistent* with the order of the field. As explained in Definition 2, this means that, for a Pth-order field, the rule used to assign the off-lattice neighbors of a boundary pixel cannot assign it pixels that lie outside a circle of radius $D_P^{1/2}$. Examples of such boundary conditions are provided by the direct generalization of the ones discussed earlier in the context of first order fields, e.g., the free and Neumann boundary conditions. If the field is defined on a torus, it also includes the periodic b.c. case. As an illustration, consider the symmetric Neumann condition for a second-order field. For a second-order field, the neighbors of the top left corner pixel (1, 1) are (0, 0), (0, 1), (0, 2), (1, 0), (2, 0), (1, 2), (2, 1), and (2, 2). Of these, the first five lie off the lattice (see Fig. 2). As explained in Section III-B, for symmetric Neumann b.c., the off-lattice neighbors of (1, 1) are assumed to have the same intensity as the on-lattice neighbors that are their corresponding mirror images (when reflected across x and y axes that are defined with (1, 1) as their origin). Thus, pixels (0, 0), (0, 2), and (2, 0), are all defined to have the same intensity as pixel (2, 2), while pixels (0, 1), and (1, 0), are defined to have the same intensity as pixels (2, 1), and (1, 2), respectively.

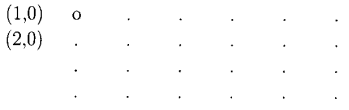

Fig. 2. Second-order fields: Neighbors of corner pixel (1, 1) marked by "o," that lie off the lattice 13.

We pursue the discussion of $A_{b.c.}$ in the context of the repartitioning technique discussed above. Once again assuming the canonical ordering, the structure of the repartitioned $A_{b.c.}$ is given by the following theorem.

Theorem 3: The repartitioned boundary potential matrix $A_{b.c.}$ of a homogeneous GMRF of given order P, with consistent b.c., has the following structure:

$$A_{b.c.} = \begin{bmatrix} B_{b.c.}^1 & C_{b.c.}^1 & \mathbf{0} & \cdot & \cdot & \mathbf{0} & (C_{b.c.}^{N_1})^T \\ (C_{b.c.}^1)^T & B_{b.c.}^2 & C_{b.c.}^2 & \mathbf{0} & \cdot & \cdot & \mathbf{0} \\ \mathbf{0} & (C_{b.c.}^2)^T & B_{b.c.}^2 & C_{b.c.}^2 & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \mathbf{0} & (C_{b.c.}^2)^T & B_{b.c.}^2 & C_{b.c.}^2 & \mathbf{0} \\ \mathbf{0} & \cdot & \cdot & \mathbf{0} & (C_{b.c.}^2)^T & B_{b.c.}^2 & C_{b.c.}^{N_1-1} \\ C_{b.c.}^{N_1} & \mathbf{0} & \cdot & \cdot & \mathbf{0} & (C_{b.c.}^{N_1-1})^T & B_{b.c.}^{N_1} \end{bmatrix}, \quad (40)$$

where the $M_1 \times M_1$ blocks have the following properties:

1) $B_{b.c.}^i = (B_{b.c.}^i)^T$ and $(B_{b.c.}^2)^T = JB_{b.c.}^2 J$,
2) $(C_{b.c.}^i)^T = JC_{b.c.}^i J$,
3) $B_{b.c.}^{N_1} = JB_{b.c.}^1 J$ and $C_{b.c.}^{N_1-1} = C_{b.c.}^1$,
4) For nonperiodic lattices,

$$C_{b.c.}^{N_1} = \mathbf{0}, \quad (41)$$

i.e., $A_{b.c.}$ is block tridiagonal.

The proof is straightforward and uses a line of reasoning similar to the one followed in the corresponding proof for Theorem 2. The matrix $A_{b.c.}$ can be partitioned into N rows of $M \times M$ blocks, where the (i, j)th block, $A_{i,j}^{b.c.}$, represents only those neighbor interactions between the pixels in row i and row j that arise as a consequence of the boundary

conditions. For nonperiodic lattices, the definition of consistent b.c. constrains $A_{\text{b.c.}}$ to a block banded structure with block bandwidth k_v. This matrix can be repartitioned to obtain a block tridiagonal matrix. The symmetry properties of the blocks and the relationships between them, including the perturbed block Toeplitz nature of $A_{\text{b.c.}}$, follow from the symmetry of the neighborhood definition and the homogeneity of the field. In particular, since the same boundary condition rule, along with the same boundary potentials, are applied to every pixel along the lattice boundary, only the interactions that involve the first and last pseudo rows can be different. For periodic lattices, the first and last pseudo rows become neighbors and their interaction is represented by additional blocks, $C_{\text{b.c.}}^{N_1}$ and $(C_{\text{b.c.}}^{N_1})^T$.

Remarks: From Theorem 3, the repartitioned boundary potential matrix $A_{\text{b.c.}}$ has the following properties.

1) It is centrosymmetric and sparse with a maximum of five nonzero block diagonals; for nonperiodic lattices it is block tridiagonal.
2) All the blocks are persymmetric. In addition, the interior blocks $B_{\text{b.c.}}^2$ along the main diagonal are centrosymmetric.
3) For nonperiodic lattices, the bandwidth of $A_{\text{b.c.}}$ is given by

$$b(A_{\text{b.c.}}) = k_v M + \max_i b_r(C_{\text{b.c.}}^i), \tag{42}$$

where

$$\max_i b_r(C_{\text{b.c.}}^i) \le \left(D_P - \lfloor D_P^{1/2} \rfloor^2\right)^{1/2}. \tag{43}$$

The structure and properties of GMRF's on periodic lattices have been thoroughly investigated; see, e.g., Moran [24], Besag and Moran [5], Kashyap [17], and Chellappa and Kashyap [9], and fast transform based processing algorithms have been derived in [9]. Accordingly, we proceed here with GMRF's defined on nonperiodic lattices. Unless otherwise stated, in the sequel all references to GMRF's assume the underlying lattice to be nonperiodic.

Structure of A: Adding A_c, and $A_{\text{b.c.}}$, (from (38) and (40), respectively), and applying (41), the block structure of the potential matrix A for nonperiodic lattices is obtained as

$$A = \begin{bmatrix} B_1 & C_1 & \mathbf{0} & \cdot & \cdot & \cdot & \cdot \\ C_1^T & B & C & \mathbf{0} & \cdot & \cdot & \cdot \\ \mathbf{0} & C^T & B & C & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \mathbf{0} & C^T & B & C & \mathbf{0} \\ \cdot & \cdot & \cdot & \mathbf{0} & C^T & B & C_{N_1-1} \\ \cdot & \cdot & \cdot & \cdot & \mathbf{0} & C_{N_1-1}^T & B_{N_1} \end{bmatrix}, \tag{44}$$

where the $M_1 \times M_1$ blocks B_1, B, B_{N_1}, C_1, C, and C_{N_1-1}, and their properties are obtained from the corresponding blocks of A_c and $A_{\text{b.c.}}$. A valid potential matrix A for a nondegenerate GMRF must be positive define (pd). Consequently, the blocks on the main diagonal, B, B_1, and B_{N_1} are pd.

Thus, a valid potential matrix for a zero-mean nondegenerate homogeneous GMRF of given order, with consistent b.c. defined on a nonperiodic lattice, has the following properties.

1) It is centrosymmetric, pd, and block tridiagonal, with the block structure previously given, in (44).
2) The blocks along the main block diagonal are symmetric and pd, i.e.,

$$B_1 = B_1^T > 0, \qquad B = B^T > 0, \qquad B_{N_1} = B_{N_1}^T > 0, \tag{45}$$

while B is also persymmetric, i.e., $B^T = JBJ$, therefore it is centrosymmetric, and B_{N_1} is related to B_1 through the reflection matrix, J, by

$$B_{N_1} = JB_1J. \tag{46}$$

3) The off-diagonal blocks are persymmetric, i.e., $C_1^T = JC_1J$ and $C^T = JCJ$, while $C_{N_1-1} = C_1$.
4) It is banded, with bandwidth

$$b(A) = k_v M + \left(D_P - \lfloor D_P^{1/2} \rfloor^2\right)^{1/2}, \tag{47}$$

where k_v was defined in (3), and $\left(D_P - \lfloor D_P^{1/2} \rfloor^2\right)^{1/2}$ is the right bandwidth of C.

As we have seen, the potential matrix, in particular its block structure, reflects the structure and properties of the underlying field. The blocks along the main diagonal represent the interactions of pixels within each pseudo row (or lattice row, depending on whether the repartitioned or original potential matrix is considered), while each off-diagonal block represents the interactions between the pixels in two different (pseudo) rows. The symmetry properties of A and its blocks reflect the homogeneity of the field, while the perturbed block Toeplitz nature of A reflects the effect of finite lattice boundaries. The block bandedness of A and the banded nature of each of its blocks is a consequence of the locally limited spatial dependence of the field. In particular, the block tridiagonal nature of the repartitioned potential matrix represents the fact that any pseudo row is only related to the one that immediately precedes it and the one that immediately follows it; equivalently, in terms of rows of the lattice, any row is only dependent on k_v rows on either side of it.

Inhomogeneous Fields: The above argument regarding the effect of the neighborhood order on the structure of A, can be applied to inhomogeneous fields to derive a repartitioned potential matrix with a block tridiagonal structure,

$$A = I_{N_1} \otimes_i B_i + K_{N_1} \otimes_{i+1} C_i^T + K_{N_1}^T \otimes_i C_i, \tag{48}$$

with its bandwidth still given by (47).

IV. Recursive Structure

In this section, we develop a recursive structure for noncausal GMRF's by using the properties of the potential matrix. The recursive structure consists of two equivalent one-sided representations that are obtained by the Cholesky factorization of the potential matrix, using a Riccati equation.

Each of these equivalent representations is defined through a set of spatially varying regressors. The transformation is simple; what makes it interesting and useful is the effect of the special structure of A on the structure and properties of the equivalent one-sided representations. The end result is an alternate representation, for a *noncausal* field of any order, consisting of a set of finite one-sided spatially varying regressors that, in the case of homogeneous fields, are shown to converge asymptotically to a spatially invariant regressor. We emphasize that, unless stated otherwise, the results to be developed in this section are valid for GMRF's of arbitrary order, not necessarily first or second order only. This generality is a result of the repartitioning technique of Section III-C, which in essence makes the block structure of the potential matrix of higher order fields look like that of first or second order.

In Section IV-A, a noncausal AR model for GMRF's is presented in terms of the potential matrix. The equivalent one-sided representations are derived in Section IV-B and some important properties of their spatially varying regressors, including their convergence to a spatially invariant regressor, are presented in Section IV-C.

A. Noncausal Autoregressive Field Representation

Alternatively to the Gibbs representation, a GMRF can be described by a *noncausal* autoregressive (AR) model driven by a correlated input field. This is the minimum mean-square (mmse) representation considered by Woods [32]. A matrix-vector formulation of this representation can be derived in terms of the potential matrix, as follows. Let

$$\vec{v} = A\vec{X}, \tag{49}$$

where A is the potential matrix and $\vec{X}$ is the $NM \times 1$ vector of the field values arranged using the canonical ordering. Since $\vec{X}$ is zero-mean Gaussian with covariance, from (5),

$$E(\vec{X}\vec{X}^T) = \left(\frac{1}{\sigma^2}A\right)^{-1}, \tag{50}$$

it is immediately apparent that $\vec{v}$ is zero-mean Gaussian with

$$E(\vec{v}\vec{v}^T) = \sigma^2 A, \tag{51}$$

and

$$E(\vec{X}\vec{v}^T) = \sigma^2 I. \tag{52}$$

Then,

$$A\vec{X} = \vec{v}, \qquad \vec{v} \sim \eta(\vec{0}, \sigma^2 A), \tag{53}$$

is a noncausal AR representation for a GMRF with potential matrix A.

From (53) and the structure of the potential matrix, the intensity (pseudo) row vectors satisfy a *noncausal* regressor whose structure is determined by the order of the field and the choice of boundary conditions. For example, for a homogeneous Dirichlet field, the noncausal regressor is

$$C^T\vec{x}_{i-1} + B\vec{x}_i + C\vec{x}_{i+1} = \vec{v}_i \tag{54}$$

with boundary conditions

$$\vec{x}_0 = 0, \qquad \vec{x}_{N+1} = 0.$$

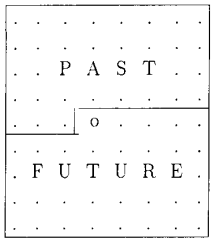

Fig. 3. Partitioning of the lattice produced by the equivalent one-sided regressor model: "o" represents the "present."

Equation (54) is not to be confused with (3.1) in [20] which superficially has the same form. Equation (54), which follows from Wood's mmse representation in (53), represents *2-D* fields. As a result, even though the iteration index i operates in one dimension, the second dimension's interactions are found in the highly structured form of the matrix coefficients B and C. In contrast, (3.1) in [20] is concerned with *1-D* fields. It has matrix coefficients simply because it represents a 1-D *vector* field, i.e., each site in the 1-D lattice has a vector of attributes. For this equation, the one-dimensional interactions are in the direction of the index i and the matrix coefficients, that appear to be analogous to B and C in (54), actually have arbitrary structure since there is no second dimension of interactions to be encoded.

For homogeneous fields, the regressor produced by the mmse representation is spatially invariant, barring the initial and final conditions. It is also noncausal, and therefore nonrecursive. Next, we see that it is possible to derive alternate representations that are one-sided, recursive, and, in the case of homogeneous fields, asymptotically spatially invariant.

B. Spatially Varying One-Sided Regressors

In this subsection, we derive two completely equivalent one-sided representations for a noncausal field, one "backward," the other "forward." Each of these is based on a set of finite, one-sided, spatially varying regressors that turn out to be asymptotically invariant when the field is homogeneous.

Backward Representation: Since A is pd, we can define its lower/upper Cholesky factorization,

$$A = U^T U, \tag{55}$$

where U is upper triangular. Then equation (53) can be rewritten as

$$U\vec{X} = \vec{w}, \tag{56}$$

where

$$\vec{w} = (U^T)^{-1}\vec{v} \tag{57}$$

and

$$\vec{w} \sim \eta(\vec{0}, \sigma^2 I). \tag{58}$$

As a result,

$$E(\vec{X}\vec{w}^T) = \sigma^2 U^{-1}. \tag{59}$$

Equations (56) and (58) define an equivalent one-sided "backward" regressor model, in which any pixel (i, j), depends only on those pixels that lie in its "future" (see Fig. 3).

The importance of this model lies in the fact that the special structure of A is reflected in the structure of U. Since A is banded as well as block banded, so is U. In fact, since A is block tridiagonal, U has only one nonzero off-diagonal block per block row:

$$U = \begin{bmatrix} U_1 & \Theta_1 & \mathbf{0} & \cdot & \cdot & \cdot \\ \mathbf{0} & U_2 & \Theta_2 & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \mathbf{0} & \mathbf{0} & U_{N_1-1} & \Theta_{N_1-1} \\ \cdot & \cdot & \cdot & \mathbf{0} & \mathbf{0} & U_{N_1} \end{bmatrix},$$

$$= I_{N_1} \otimes_i U_i + K_{N_1}^T \otimes_i \Theta_i, \tag{60}$$

where U_i and Θ_i are $M_1 \times M_1$ blocks, with the U_i's being upper triangular.

The matrix vector representation in (56) can be expanded as an $(M_1 \times 1)$ vector AR process:

$$U_i \vec{x}_i + \Theta_i \vec{x}_{i+1} = \vec{w}_i, \qquad 1 \le i \le N_1 - 1, \tag{61}$$

$$U_{N_1} \vec{x}_{N_1} = \vec{w}_{N_1},$$

with $\vec{w}_i \perp \vec{x}_j$ for $i < j$, and

$$E(\vec{w}_i \vec{w}_j^T) = \sigma^2 I_{M_1} \delta_{i,j}, \qquad E(\vec{x}_i \vec{w}_i^T) = \sigma^2 U_i^{-1}, \tag{62}$$

for $1 \le i \le N_1$, where $\delta_{i,j}$ is the Kronecker symbol. Here, the $\vec{x}_i$ are $M_1 \times 1$ (i.e., $k_v M \times 1$) vectors, each of which represents a pseudo row (i.e., k_v rows of the lattice, stacked one on top of the other). The driving noise, $\{w_{i,j}\}$, is white and uncorrelated with all the pixel values, $x_{k,l}$, that lie in its "future."

At the pixel level, for a pixel in pseudo row i, U_i and Θ_i contain the regressor coefficients for the "future" pixels in, respectively, the same pseudo row (i), and the next pseudo row $(i + 1)$. The order of the regressor follows from the right bandwidth of U, which is equal to the bandwidth of A,

$$b_r(U) = b(A) = k_v M + \left(D_P - \left\lfloor D_P^{1/2} \right\rfloor^2\right)^{1/2}. \tag{63}$$

As an illustration, consider first-order fields. For these fields, $k_v = 1$ and $M_1 = M$, i.e., the pseudo rows reduce to being the rows of the lattice. Then

$$b_r(U) = M, \tag{64}$$

and, as a result, the Θ_i's are lower triangular. At the pixel level, this means that the noncausal neighborhood set of four pixels (labeled as pixels "1" in Fig. 4(a)) is replaced by a "anticausal" neighborhood set of M pixels (labeled as pixels "1" in Fig. 4(b)).

More generally, a noncausal neighborhood set of N_P pixels is replaced by a "anticausal" neighborhood set of $k_v M + \left(D_P - \left\lfloor D_P^{1/2} \right\rfloor^2\right)^{1/2}$ pixels. For example, with reference to Fig. 4, the second-order noncausal neighborhood set (labeled as pixels "1" and "2" in Fig. 4(a)), is replaced by an "anticausal" neighborhood set of $M + 1$ pixels (labeled as pixels "1" and "2" in Fig. 4(b)). Similarly, the same applies to higher order noncausal models as illustrated in Fig. 4.

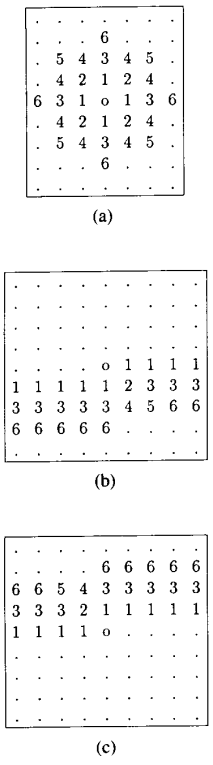

Fig. 4. (a) Neighbors for Pth-order noncausal MRF, $P = 1, 2, 3, 4, 5, 6$. (b) Neighbors for the equivalent backward regressor. (c) Neighbors for the equivalent forward regressor.

Thus, (61) provides an "anticausal" or "backward," or "bottom up" recursive procedure to generate the field using a set of *one-sided* spatially varying regressors that are defined by the matrices $\{U_i\}$ and $\{\Theta_i\}$.

Forward Representation: Alternatively to (55), an upper/lower factorization of A as

$$A = L^T L, \tag{65}$$

where L is lower triangular, produces a "forward" regressor model

$$L\vec{X} = \vec{z}, \tag{66}$$

where

$$\vec{z} = (L^T)^{-1} \vec{v} \tag{67}$$

and

$$\vec{z} \sim \eta(\vec{0}, \sigma^2 I), \tag{68}$$

resulting in

$$E(\vec{X}\vec{z}^T) = \sigma^2 L^{-1}. \tag{69}$$

As a consequence of the fact that A is block tridiagonal, L has only one nonzero off-diagonal block per block row:

$$L = \begin{bmatrix} L_{N_1} & \mathbf{0} & \mathbf{0} & \cdot & \cdot & \cdot \\ P_{N_1-1} & L_{N_1-1} & \mathbf{0} & \mathbf{0} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \mathbf{0} & P_2 & L_2 & \mathbf{0} \\ \cdot & \cdot & \cdot & \mathbf{0} & P_1 & L_1 \end{bmatrix},$$

$$= I_{N_1} \otimes_i L_{N_1-i+1} + K_{N_1}^T \otimes_i P_{N_1-i+1}, \quad (70)$$

where L_i and P_i are $M_1 \times M_1$ blocks, with the L_i's being lower triangular.

Analogous to (61), a "causal" or "forward" procedure, in the form of a set of spatially varying one-sided regressors defined by the matrices $\{L_i\}$ and $\{P_i\}$, can be derived from (66):

$$L_{N_1}\vec{x}_1 = \vec{z}_1,$$

$$L_{N_1-i+1}\vec{x}_i + P_{N_1-i+1}\vec{x}_{i-1} = \vec{z}_i, \qquad 2 \le i \le N_1, \quad (71)$$

with $\vec{z}_i \perp \vec{x}_j$ for $j < i$, and

$$E(\vec{z}_i\vec{z}_j^T) = \sigma^2 I_{M_1}\delta_{i,j}, \qquad E(\vec{x}_i\vec{z}_i^T) = \sigma^2 L_{N_1-i+1}^{-1}, \quad (72)$$

for $1 \le i \le N_1$. Using the same argument as for the "backward" representation, the one-sided neighborhood set for the "forward" regressors is shown to be a reflected version of the corresponding one for the "backward" regressors (see Fig. 4(c)).

For homogeneous fields, since A is centrosymmetric, it follows that

$$L = JUJ, \quad (73)$$

i.e.,

$$L_i = JU_iJ, \qquad 1 \le i \le N_1, \quad (74)$$

$$P_i = J\Theta_iJ, \qquad 1 \le i \le N_1 - 1. \quad (75)$$

In other words, each block in L can be obtained directly from the corresponding block in U, by rotating it through 180°.

The main thrust of this subsection is the fact that for any given noncausal field, not necessarily homogeneous, there exist two equivalent one-sided representations, one of which is "backward" or "anticausal," the other "forward" or "causal." Each of the two representations consists of a set of spatially varying regressors, defined respectively by the two sets of matrices $\{U_i, \Theta_i\}$, and $\{L_i, P_i\}$. The effect of this transformation is to go from a noncausal neighborhood interaction with N_P neighbors to a one-sided spatially varying interaction with $k_v M + \left(D_P - \left\lfloor D_P^{1/2} \right\rfloor^2\right)^{1/2}$ neighbors which lie either entirely in the "future" or entirely in the "past" (see Fig. 4).

For homogeneous fields, the "backward" and "forward" representations are related to each other through a series of reflections (see (74) and (75)); therefore, only one of them, and correspondingly only one set of spatially varying regressors, has to be computed explicitly. Consequently, in the sequel, we concentrate on the properties and structure of the "backward" representation.

C. Asymptotic Convergence of the Spatially Varying Regressors

The one-sided regressors are found by solving a pair of Riccati type equations. We will write them for homogeneous fields only, so as not to clutter the issue with notational complexity. However, we note that these equations follow from the block tridiagonal structure of the potential matrix, so that a similar algorithm is true for inhomogeneous fields.

Let

$$\Sigma_i = U_i^T U_i \quad (76)$$

and

$$\Lambda_i = \Theta_i^T \Theta_i. \quad (77)$$

Then, by multiplying out the block rows in the Cholesky factorization equation (55) and using the block structure of A (from (44)), the following pair of iterative schemes for computing $\{U_i\}_{i=1}^{N_1}$ and $\{\Theta_i\}_{i=1}^{N_1-1}$, respectively, is obtained.

1)

$$\Sigma_1 = B_1, \quad (78)$$

$$\Sigma_2 = B - C_1^T B_1^{-1} C_1, \quad (79)$$

$$\text{for } 3 \le i \le N_1 - 1, \qquad \Sigma_i = B - C^T \Sigma_{i-1}^{-1} C, \quad (80)$$

$$\Sigma_{N_1} = JB_1J - C_1^T \Sigma_{N_1-1}^{-1} C_1. \quad (81)$$

2)

$$\Lambda_1 = C_1^T B_1^{-1} C_1, \quad (82)$$

$$\text{for } 2 \le i \le N_1 - 2, \qquad \Lambda_i = C^T \left[B - \Lambda_{i-1}\right]^{-1} C, \quad (83)$$

$$\Lambda_{N_1-1} = C_1^T \left[B - \Lambda_{N_1-2}\right]^{-1} C_1. \quad (84)$$

Actually only one of these two iterations is needed since

$$U_1^T \Theta_1 = C_1, \quad (85)$$

$$U_i^T \Theta_i = C, \qquad 2 \le i \le N_1 - 2, \quad (86)$$

$$U_{N_1-1}^T \Theta_{N_1-1} = C_1. \quad (87)$$

Equations of the type of (80) and (83) arise in the study of stability and control problems in linear systems theory, being referred to as Riccati equations, see, e.g., [7].

We will see that for homogeneous fields, under minor assumptions, the spatially varying regressors $(\{U_i, \Theta_i\})$ converge asymptotically at a geometric rate to a spatially invariant regressor $(\{U_\infty, \Theta_\infty\})$. This has considerable practical significance as far as numerical implementations are concerned.

We proceed with the iteration in (80) and consider the properties of the sequence of matrices, $\{\Sigma_i\}$, defined by (80), with initial condition (79), under the assumptions given next.

Let $\|\cdot\|_s$ be the matrix spectral norm, and define the quantity

$$\zeta = 2\|B^{-1}\|_s\|C\|_s. \tag{88}$$

Assumed Conditions:

1) $B = B^T > 0$,
2) $0 < \zeta < 1$,
3) the initial condition matrix, Σ_2, has
   a) $B/2 < \Sigma_2 \leq B$,
   b) $\Sigma_2 - B + C^T\Sigma_2^{-1}C$ is not indefinite.

Condition 1) holds for the potential matrix of any nondegenerate GMRF (see (45)). The right inequality in Condition 2) can be interpreted as the assumption that the block tridiagonal, block Toeplitz, matrix,

$$I_{N_1} \otimes B + K_{N_1} \otimes C^T + K_{N_1}^T \otimes C,$$

is strictly *block diagonally dominant* with respect to the spectral norm (i.e., [29]). In order to avoid technical complications, such as division by zero, fields with

$$C = \mathbf{0} \tag{89}$$

are handled separately. Such fields have only horizontal interactions, i.e., they are degenerate 2-D fields that are really only 1-D. For these fields, substituting (89) into (80) produces a constant sequence,

$$\Sigma_i = B, \tag{90}$$

which is trivially convergent. We could also include the case where (89) holds, by requiring $\beta_h \neq 0$ and stacking the field column-wise instead of row-wise. In the sequel, then, we need only to consider explicitly the case where

$$C \neq \mathbf{0}. \tag{91}$$

This restriction is enforced through the left inequality in Condition 2), i.e., $\zeta > 0$. Condition 3) ensures that sequence $\{\Sigma_i\}$ is initialized as a bounded and monotone sequence. Later, when we consider specific fields as examples, we will see that mild restrictions on the parameter space are sufficient to ensure that the above conditions hold.

The conditions assumed are used to derive a series of lemmas that capture the important properties of sequence $\{\Sigma_i\}$ and these in turn lead to the main convergence results of this section, presented in Theorems 4 and 5 and Corollary 2. The proofs are all relegated to the Appendix.

Lemma 1: Given Conditions 1) and 2), $B > 4C^TB^{-1}C$.

Lemma 2: Given Conditions 1), 2), and 3a), $\{\Sigma_i\}$ is a bounded sequence, i.e.,

$$\forall i \geq 2, \qquad B/2 < \Sigma_i \leq B. \tag{92}$$

Lemma 3: Given Conditions 1)–3), $\{\Sigma_i\}$ is a monotone sequence, i.e.,

$$(\forall i \geq 2, \Sigma_i - \Sigma_{i+1} \geq 0) \quad \text{or} \quad (\forall i \geq 2, \Sigma_i - \Sigma_{i+1} \leq 0). \tag{93}$$

Theorem 4: Given Conditions 1)–3), the sequence, $\{\Sigma_i\}$, converges, i.e.,

$$\Sigma_i \to \Sigma_\infty, \tag{94}$$

where Σ_∞ is a pd solution of the matrix Riccati equation

$$\Sigma_\infty = B - C^T\Sigma_\infty^{-1}C. \tag{95}$$

Alternatively, if we proceed with the parallel iteration in (83), it is possible to derive a theorem along the same lines as Theorem 4 to prove that sequence $\{\Lambda_i\}$ converges.

An explicit expression for the steady state solution, Σ_∞, can be obtained, if an additional constraint is imposed on the matrix C:

$$C = C^T\begin{cases} >0, \\ <0, \end{cases} \tag{96}$$

i.e., C is symmetric pd, or symmetric nd (negative definite). In the last section, we saw that for third- or higher order fields repartitioning of A is required, and the resulting C is block lower triangular. Thus, (96) is only plausible for first- or second-order fields.

Then the following lemma can be derived.

Lemma 4: Given Conditions 1) and 2), and the additional condition in (96):

$$B \pm 2C > 0. \tag{97}$$

For any symmetric, pd matrix R, define the square root as the symmetric, pd matrix $R^{1/2}$ such that $R = R^{1/2}R^{1/2}$. The additional condition in (96) and Lemma 4 can be used to derive the following corollary to Theorem 4. Let

$$D = \begin{cases} C, & \text{if } C > 0, \\ -C, & \text{if } C < 0. \end{cases} \tag{98}$$

Corollary 2: Given the condition in (96), in addition to Conditions 1)–3), sequence $\{\Sigma_i\}$ converges to

$$\Sigma_\infty = (B/2) + D^{1/2}\left[\left(D^{-1/2}(B/2)D^{-1/2}\right)^2 - I\right]^{1/2}D^{1/2}, \tag{99}$$

where D is given by (98).

Given the additional condition in (96), the Riccati equation (95) has two pd solutions (see the proof of Corollary 2 in the Appendix), one of which, (99), is the steady state solution of the iteration in (80). It is straightforward to show that the other one,

$$\Lambda_\infty = (B/2) - D^{1/2}\left[\left(D^{-1/2}(B/2)D^{-1/2}\right)^2 - I\right]^{1/2}D^{1/2}, \tag{100}$$

is the steady state solution of the parallel iteration in (83).

Theorem 4 provides the convergence of sequence $\{\Sigma_i\}$. Under the same conditions that are sufficient for the convergence, it is possible to derive a bound on the convergence rate, that is an explicit function of blocks B and C.

Theorem 5: Given Conditions 1)–3), the rate of convergence of sequence $\{\Sigma_i\}$ is geometric, i.e.,

$$\|\Sigma_i - \Sigma_\infty\|_s \leq K_0 a^i, \qquad \forall i \geq 2, \tag{101}$$

where

$$a = \zeta^2, \tag{102}$$

$$K_0 = \frac{\|\Sigma_2 - \Sigma_\infty\|_s}{\zeta^4}, \tag{103}$$

ζ is defined by (88), and

$$0 < a < 1, \tag{104}$$

$$0 < K_0 \le \frac{2\|B\|_s}{\zeta^4}. \tag{105}$$

Most applications involve numerical implementation on a digital computer with finite precision. For such applications the geometric rate of convergence of the iteration is highly significant because it means that after a small number of steps

$$n \ll N_1, \tag{106}$$

it can be safely assumed that

$$\Sigma_i \approx \Sigma_\infty, \qquad n \le i \le N_1 - 1, \tag{107}$$

and so

$$U_i \approx U_\infty, \qquad n \le i \le N_1 - 1, \tag{108}$$

$$\Theta_i \approx \Theta_\infty, \qquad n \le i \le N_1 - 1, \tag{109}$$

where U_∞ is the Cholesky factor of Σ_∞ i.e.,

$$\Sigma_\infty = U_\infty^T U_\infty, \tag{110}$$

and

$$\Theta_\infty = U_\infty^{-T} C. \tag{111}$$

To satisfy the end condition, (81), one more iteration is needed at the end:

$$\Sigma_{N_1} \approx JB_1 J - C_1^T \Sigma_\infty^{-1} C_1. \tag{112}$$

These results are applicable to fields of any order, for the parameter space (i.e., a range of parameter values) in which Conditions 1)–3) hold. When considering a specific field, it is often the case that tighter bounds can be derived for the convergence rate, than the one in Theorem 5, by using additional information regarding the structure of matrices B and C. This will be demonstrated in the next section where we consider some examples.

V. First- and Second-Order Fields

In the previous section, we proved that for a homogeneous field of any order, under certain conditions (Conditions 1)–3)), the spatially varying one-sided regressors converge asymptotically, with a geometric rate, to a spatially invariant regressor. Here, we illustrate the fact that for a given field these conditions are easily verified if mild restrictions are imposed on the parameter space. We also illustrate how the general results can be tightened when we consider explicitly the structure of first- and second-order homogeneous fields. Finally, we show the practical significance of the geometric rate of convergence by running a simple computational experiment.

For first and second order fields, $k_v = 1$, therefore, $M_1 = M$ and $N_1 = N$, i.e., the pseudo rows reduce to being the rows of the lattice.

A. First-Order Fields

For homogeneous first-order fields, the potential matrix, A, has the block structure given in (44), with β_v and β_h being the vertical and horizontal potentials,

$$C = -\beta_v I_M, \tag{113}$$

and B, B_1, C_1 depending on the b.c. chosen (see Corollary 1).

Substituting (113) into (88), we find that, for first-order fields, ζ simplifies to the following form:

$$\zeta = 2\,|\,B_v\,|\,\|B^{-1}\|_s. \tag{114}$$

We consider the following three fields that were introduced in Section III-B.

1) *Dirichlet Fields:* From (16) and (25), it follows that B is given by (17), and

$$B_1 = B, \qquad C_1 = C. \tag{115}$$

2) *Variational Fields:* From (7), (16), (17), and (27), it follows that

$$B = \begin{bmatrix} 1-\beta_h & -\beta_h & 0 & \cdot & \cdot & \cdot \\ -\beta_h & 1 & -\beta_h & 0 & \cdot & \cdot \\ 0 & -\beta_h & 1 & -\beta_h & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & -\beta_h & 1 & -\beta_h & 0 \\ \cdot & \cdot & 0 & -\beta_h & 1 & -\beta_h \\ \cdot & \cdot & \cdot & 0 & -\beta_h & 1-\beta_h \end{bmatrix},$$

$$= I_M - \beta_h(H_M + I_{1,M}), \tag{116}$$

$$B_1 = B + C, \tag{117}$$

$$C_1 = C. \tag{118}$$

3) *Symmetric Fields:* From (7), (16), (17), (29), and (30), it follows that

$$B = \begin{bmatrix} 1 & -2\beta_h & 0 & \cdot & \cdot & \cdot \\ -2\beta_h & 1 & -\beta_h & 0 & \cdot & \cdot \\ 0 & -\beta_h & 1 & -\beta_h & 0 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 0 & -\beta_h & 1 & -\beta_h & 0 \\ \cdot & \cdot & 0 & -\beta_h & 1 & -2\beta_h \\ \cdot & \cdot & \cdot & 0 & -2\beta_h & 1 \end{bmatrix},$$

$$= I_M - \beta_h(H_M + F_{1,2} + F_{M-1,M}), \tag{119}$$

$$B_1 = B, \qquad C_1 = 2C. \tag{120}$$

Note that for the variational and symmetric fields, B fails to be Toeplitz because of the effects of the boundary conditions. In the sequel, if not otherwise specified, when simply referring to first order fields we mean one of these fields.

Verification of Conditions 1)–3): Sufficient conditions for a symmetric matrix to be pd are that all its diagonal

elements are positive and that it is strictly diagonally dominant (see the corollary to Theorem 1.8 in [31]). For Dirichlet and variational fields, applying these conditions to the potential matrix results in the constraint

$$|\beta_v| + |\beta_h| < 1/2, \tag{121}$$

while for symmetric fields the corresponding constraint is

$$|\beta_v| + |\beta_h| < 1/3. \tag{122}$$

To avoid the trivial case of a null C (see discussion following (89)) we also assume that

$$\beta_v \neq 0. \tag{123}$$

Result 1: For the parameter space defined by (121) (or (122)) and (123), Conditions 1), 2), and 3) are satisfied for Dirichlet and variational (symmetric) fields.

The details of the proof are given in the Appendix

Steady State Solution: From (113) and (123), the additional constraint in (96) is satisfied for all three of the first-order fields, and Corollary 2 can be applied. We can obtain a simpler form for the steady state solution of the Riccati equation. The steady state solution of the iteration in (80) is obtained from (98), (99), (113), and (123). After a little algebraic manipulation, it leads to the following expression:

Result 2: For first-order fields:

$$\Sigma_\infty = B/2 + \sqrt{(B/2)^2 - \beta_v^2 I}, \tag{124}$$

where B depends on the particular field.

Monotonic Behavior of $\{\Sigma_i\}$*:* Lemma 3 has shown that $\{\Sigma_i\}$ is monotonic. For first-order fields more can be said. From (167), (217), (220), (222), and (223), we have the following.

Result 3: The sequence $\{\Sigma_i\}$ is:

1) monotonically decreasing for first-order Dirichlet fields, i.e.,

$$\Sigma_i - \Sigma_{i+1} > 0, \qquad \forall i \geq 2; \tag{125}$$

2) monotonically increasing or decreasing with $\beta_v < 0$ or $\beta_v > 0$, respectively, for first-order variational fields, i.e.,

$$\beta_v > 0 \Rightarrow \Sigma_i - \Sigma_{i+1} < 0, \qquad \forall i \geq 2, \tag{126}$$

$$\beta_v < 0 \Rightarrow \Sigma_i - \Sigma_{i+1} > 0, \qquad \forall i \geq 2; \tag{127}$$

3) monotonic increasing for first-order symmetric fields, i.e.,

$$\Sigma_i - \Sigma_{i+1} < 0, \qquad \forall i \geq 2. \tag{128}$$

Tighter Bounds on the Rate of Convergence: For first-order fields, it is possible to derive for the rate of convergence in (101) bounds tighter than the one given by (102).

Lemma 5: Given Σ_∞, defined by (124), and ζ defined by (114), it follows that

$$\|\Sigma_\infty^{-1}\|_s = \frac{2\|B^{-1}\|_s}{1 + \sqrt{1 - \zeta^2}}, \tag{129}$$

and

$$a = \frac{\zeta^2}{1 + \sqrt{1 - \zeta^2}}. \tag{130}$$

The bound in (130) is tighter than the one in (102), because Condition 2) implies that

$$\frac{1}{2} < \frac{1}{1 + \sqrt{1 - \zeta^2}} < 1. \tag{131}$$

Lemma 6: When $\{\Sigma_i\}$ is monotonically decreasing, then

$$a = \frac{\zeta^2}{\left(1 + \sqrt{1 - \zeta^2}\right)^2}. \tag{132}$$

The proof is simple. Since $\{\Sigma_i\}$ is monotonically decreasing,

$$\forall i \geq 2, \quad \Sigma_i > \Sigma_\infty \Rightarrow \Sigma_i^{-1} < \Sigma_\infty^{-1}$$
$$\Rightarrow \|\Sigma_i^{-1}\|_s < \|\Sigma_\infty^{-1}\|_s. \tag{133}$$

Substituting (133) into (190), in place of (195), and using (129) instead of (196), in the proof for Theorem 4, we obtain (132).

The bound in (132) is tighter than either (102) or (130), because of (131). From (127) and (125), respectively, variational fields with negative β_v and Dirichlet fields have monotonic decreasing sequences. Consequently, the bound in (132) can be applied to them.

Effect of the Field Interactions on the Convergence Rate: From (102), (130), or (132), it is clear that the convergence rate factor, a, increases or decreases as ζ increases or decreases. When the eigenvalues of B are known we can get exact expressions for $\|B^{-1}\|_s$, and hence, from (114), for ζ. For example, for the first-order Dirichlet field, B has the structure given in (17), which corresponds to the sine transform matrix in [16], whose eigenvalues are given by

$$\lambda_i(B) = 1 - 2\beta_h \cos\frac{i\pi}{M+1}, \qquad 1 \leq i \leq M. \tag{134}$$

Similarly, for the first-order variational model, B has the structure given in (116), which corresponds to the cosine transform matrix in [16], whose eigenvalues are given by

$$\lambda_i(B) = 1 - 2\beta_h \cos\frac{(i-1)\pi}{M}, \qquad 1 \leq i \leq M. \tag{135}$$

Substituting (134) and (135), into (114), we get the following result.

Results 4: For first-order fields,

$$\zeta = \begin{cases} \dfrac{2\,|\beta_v|}{\left(1 - 2\,|\beta_h| \cos\dfrac{\pi}{M+1}\right)}, \\ \qquad \text{for Dirichlet fields,} \\ \dfrac{2\,|\beta_v|}{(1 - 2\,|\beta_h|)}, \\ \qquad \text{for variational fields, with } \beta_h > 0, \\ \dfrac{2\,|\beta_v|}{\left(1 - 2\,|\beta_h| \cos\dfrac{\pi}{M}\right)}, \\ \qquad \text{for variational fields, with } \beta_h < 0. \end{cases} \tag{136}$$

From (136), it is clear that for first-order Dirichlet and variational fields, weak field interactions, as represented by small values of the interaction parameters, β_v and β_h, lead to faster rates of convergence, while strong interactions imply a slower convergence rate. A similar result, although not as clean, can be obtained for symmetric fields and is presented elsewhere.

Practical Significance of Convergence: A simple computational experiment illustrates the practical significance of the geometric rate of convergence. For a first-order Dirichlet field with parameter space defined by (123) and (121), the iteration in equation (80) was used to generate the Cholesky factor U, up to a specified precision, ϵ, i.e., the iteration was stopped at $i = n$ if

$$\| \Sigma_i - \Sigma_{i-1} \|_s \leq \epsilon \tag{137}$$

and the approximation

$$U_j = U_n, \qquad \text{for } j > n \tag{138}$$

was used, or continued to $i = N$ if the tolerance level was not achieved. This experiment was carried out with $N = M = 8$, and tolerance level ϵ arbitrarily set to 10^{-6}, for several different pairs of parameter values. In each of these cases, $n_{N,M}$, the number of iterations needed for convergence (up to the specified tolerance level) is recorded in Table I. The experiments were repeated for $N = M = 16$, 32, 64, and 128. The columns for $N = M = 32$, 64, and 128 are omitted because the values of $n_{N,M}$ are the same as obtained for $N = M = 16$, given β_h, β_v, and ϵ. The table shows that even for values of β_v and β_h near the upper bound in (121), the iteration converges very fast. The worst case recorded is $n_{N,M} = 9$, which for values of $N > 8$ represents considerable savings. The results also indicate that the lattice size (M) has an insignificant effect on the number of iterations required which is consistent with the fact that the upper bound on the convergence rate (see (132) and (136)) only depends on M through the cosine term ($\cos(\pi/(M+1))$) which stays practically constant as M increases past 8.

TABLE I
FACTORIZATION OF POTENTIAL MATRIX FOR FIRST-ORDER DIRICHLET FIELD

ϵ	β_v	β_h	$n_{8,8}$	$n_{16,16}$
10^{-6}	0.10	0.15	4	4
10^{-6}	0.10	0.30	5	5
10^{-6}	0.15	0.10	5	5
10^{-6}	0.15	0.15	5	5
10^{-6}	0.15	0.22	6	6
10^{-6}	0.22	0.15	7	7
10^{-6}	0.22	0.22	8	9
10^{-6}	0.30	0.10	8	9

B. Second-Order Fields

For the purpose of illustration, we consider only Dirichlet fields. Other second-order fields follow with slightly more complicated constraints on the parameter space. For homogeneous Dirichlet fields, B is given by (17), C by (36), and B_1 and C_1 are still given by (115) (recall that free b.c. imply $A = A_c$).

A sufficient condition for A to be pd is strict diagonal dominance. This translates to the constraint equation

$$|\beta_v| + |\beta_h| + |\beta_{d1}| + |\beta_{d2}| < 1/2. \tag{139}$$

To avoid $C = \mathbf{0}$, the parameters are restricted to

$$\max(|\beta_v|, |\beta_{d1}|, |\beta_{d2}|) \neq 0, \tag{140}$$

i.e., the field must have at least one nonzero interaction between pixels in two adjoining rows. Alternatively, as previously mentioned, we could assume $\beta_h \neq 0$ and stack the field column-wise instead of row-wise.

Verification of Conditions 1)–3): It is straightforward to verify Condition 1). To verify Condition 2), let $\| \cdot \|_\infty$ be the matrix infinity norm. Using (203), (204), and (17), in (88), we get

$$\begin{aligned} \zeta &\leq \frac{2\|C\|_s}{(1 - 2\,|\beta_h|)} \\ &\leq \frac{2\|C\|_\infty}{(1 - 2\,|\beta_h|)} = \frac{2(|\beta_v| + |\beta_{d1}| + |\beta_{d2}|)}{(1 - 2\,|\beta_h|)} \\ &< 1 \qquad (\text{using (139)}), \end{aligned} \tag{141}$$

i.e., Condition 2) holds. Finally, the procedure to prove Conditions 3a) and 3b) follows along the same lines as the corresponding procedure for first-order Dirichlet fields.

Thus, for second-order Dirichlet fields the convergence results derived earlier (in Theorems 4 and 5) are valid over the parameter space defined by (14) and (139). Note that the parameter space previously defined is only a *sufficient* space, and does not preclude the possibility of these results being applicable for parameter values that lie outside it.

VI. STATE SPACE REPRESENTATION

The "backward" and "forward" regressor models derived in the last section can be put in the form of "backward" and "forward" state space representations, which are convenient for field synthesis and applications such as filtering or smoothing of noisy fields.

Inverting U_i in (61), the following "backward" state space row model results:

$$\vec{x}_i = F_i^b \vec{x}_{i+1} + G_i^b \vec{w}_i, \qquad 1 \le i \le N_1 - 1, \quad (142)$$
$$\vec{x}_{N_1} = G_{N_1}^b \vec{w}_{N_1},$$

where

$$F_i^b = -U_i^{-1}\Theta_i = \begin{cases} -\Sigma_1^{-1}C_1, & \text{for } i = 1, \\ -\Sigma_i^{-1}C, & \text{for } 2 \le i \le N_1 - 2, \\ -\Sigma_{N_1-1}^{-1}C_1, & \text{for } i = N_1 - 1, \end{cases} \quad (143)$$

$$G_i^b = U_i^{-1}, \qquad 1 \le i \le N_1. \quad (144)$$

The second equality in (143) follows from (85)–(87) and (76).

For homogeneous fields, the results of Theorem 5 say that this backward model is asymptotically invariant, so that in actual computations only a few distinct models need to be taken into consideration, i.e., for $i \ge n$

$$F_i^b \approx F^b = -\Sigma_\infty^{-1}C \quad (145)$$

and

$$G_i^b \approx G^b = U_\infty^{-1}. \quad (146)$$

The corresponding "forward" state space model is obtained from (71),

$$\vec{x}_1 = G_1^f \vec{z}_1,$$
$$\vec{x}_i = F_i^f \vec{x}_{i-1} + G_i^f \vec{z}_i, \qquad 2 \le i \le N_1, \quad (147)$$

where

$$F_i^f = -L_{N_1-i+1}^{-1} P_{N_1-i+1}$$
$$= \begin{cases} -J\Sigma_{N_1-1}^{-1}C_1 J, & \text{for } i = 2, \\ -J\Sigma_{N_1-i+1}^{-1}CJ, & \text{for } 3 \le i \le N_1 - 1, \\ -J\Sigma_1^{-1}C_1 J, & \text{for } i = N_1, \end{cases} \quad (148)$$

$$G_i^f = L_{N_1-i+1}^{-1}, \qquad 1 \le i \le N_1. \quad (149)$$

The second equality in (148) is obtained from (74), (75), (85)–(87), and (76). Substituting (143) into (148), and (144) into (149), we get

$$F_i^f = JF_{N_1-i+1}^b J, \qquad 2 \le i \le N_1, \quad (150)$$

$$G_i^f = JG_{N_1-i+1}^b J, \qquad 1 \le i \le N_1, \quad (151)$$

and for the invariant model

$$F^f = JF^b J, \quad (152)$$

$$G^f = JG^b J. \quad (153)$$

Either of the state space representations may be applied to do fast recursive synthesis of samples of noncausal GMRF's, see, e.g., [3].

VII. Recursive Smoothing of Noisy Fields

The "backward" or "forward" state space representations previously given by (142) and (147), respectively, can be used as the basis for recursive estimation from noisy observations. Let the noisy field be represented by

$$y_{i,j} = x_{i,j} + n_{i,j} \qquad 1 \le i \le N, \qquad 1 \le j \le M, \quad (154)$$

where $n_{i,j}$ is additive white Gaussian noise independent of $x_{i,j}$. Equation (154) can be applied in conjunction with either (142) or (147) to produce the optimal fixed interval smoother estimate using, for example, the recursive double sweep Rauch–Tung–Striebel (RTS) implementation [26]. For other formulations of recursive smoothers, see, for example, Mayne [23], Fraser and Potter [10], and Bucy [8]. See [3] for results on the enhancement of noisy images using the recursive framework and the RTS smoother.

VIII. Conclusion

The main points presented are the following.

- Rather than describing GMRF's by their covariance structure, this paper proposes that the inverse of the covariance structure, herein called the potential matrix, be used.
- The paper introduces a canonical representation for GMRF's in terms of the structure of the potential matrix that reflects in a natural way the modeling assumptions on the fields.
- Noncausal GMRF's have a recursive structure that is intrinsic to the field. This structure is derived by utilizing the properties of the canonical representation. As a result of this recursive structure, a noncausal field of any given order can alternatively be represented by a set of finite-order one-sided spatially varying field regressors. Fig. 4 captures the essence of this equivalence.
- The noncausal mmse field regressor (see (54)) is driven by correlated noise. In contrast, the corresponding one-sided regressors (see (61) or (71)) are driven by white noise.
- The spatially varying regressors correspond to the successive iterates of a Riccati equation.
- For homogeneous fields, the one-sided spatially varying regressors converge at a geometric rate to an invariant regressor. This is of considerable practical significance as far as numerical implementations on finite precision machines are concerned since it reduces the computation required to derive the alternate representations.
- The equivalent unilateral descriptions enable samples of noncausal fields to be synthesized recursively.
- The recursive equivalent representations are convenient for applications such as field enhancement from noisy observations, allowing the use of techniques such as Kalman–Bucy filtering and two-point recursive smoothers.

Elsewhere, we apply the results on the structure of the potential matrix to the problem of estimating the field parameters.

APPENDIX

Proof of Theorem 1: From (6), the inverse of A is a nondegenerate covariance matrix. Then (6) is equivalent to the existence of a Gauss random field, X, whose covariance is a scaled version of A^{-1}. It remains to show that the structure of A, as defined by (7)–(14), is equivalent to X being a first-order Gauss MRF.

Assume the canonical ordering. Any other ordering is just a permutation of this one.

If: Given A, satisfying the properties listed in the theorem, it is straightforward to verify that $\vec{X}^T A \vec{X}$ only involves quadratic terms, $x_{i,j}^2$, and $x_{i,j}x_{k,l}$, such that (i, j) and (k, l) are neighbors according to the first-order neighborhood rule. Therefore, X is a zero-mean, nondegenerate, first-order GMRF.

Only if: Given a zero-mean, nondegenerate, first-order GMRF with boundary conditions that are consistent with the first-order neighborhood rule (but allowing for the possibility that the lattice is periodic) and assuming the canonical ordering, the structure of A in the statement of the theorem, follows immediately from the HCT. □

Proof of Corollary 1: The Corollary follows from Theorem 1, by applying the definition of homogeneity. □

Proof of Lemma 1: Using Condition 1) and properties of pd matrices, it follows that

$$B^{-1} = B^{-1/2}B^{-1/2} \tag{155}$$

and

$$C^T B^{-1} C \geq 0. \tag{156}$$

Then

$$\begin{aligned} B > 4C^T B^{-1} X \geq 0 \\ \Leftrightarrow I > 4B^{-1/2}C^T B^{-1/2}B^{-1/2}CB^{-1/2} \geq 0 \\ \Leftrightarrow I > D^T D \geq 0, \end{aligned} \tag{157}$$

where

$$D = 2B^{-1/2}CB^{-1/2}. \tag{158}$$

The second inequality in (157) follows immediately. Let $\lambda_i(D^T D)$ represent the ith eigenvalue of the matrix $D^T D$. Using Theorem 1.3 and Definition 1.2 from [31, pp. 9, 11],

$$\|D\|_s = \max_i \left(\lambda_i(D^T D)\right)^{1/2}. \tag{159}$$

From (158) and matrix norm properties (Theorem 1.2 in [31, p. 9]), we get

$$\begin{aligned} \|D\|_s &\leq 2\|B^{-1/2}\|_s^2 \|C\|_s = 2\|B^{-1}\|_s \|C\|_s \\ &< 1 \quad \left(\text{using (88) and Condition 2)}\right). \end{aligned} \tag{160}$$

From (159) and (160),

$$\max_i \lambda_i(D^T D) < 1 \tag{161}$$

$$\Rightarrow D^T D < I. \qquad \square \tag{162}$$

Proof of Lemma 2: Prove by induction.

Basis: Follows immediately from Condition 3a).

Induction: Assume for arbitrary $i > 2$

$$B/2 < \Sigma_{i-1} \leq B. \tag{163}$$

This implies that

$$C^T B^{-1} C \leq C^T \Sigma_{i-1}^{-1} C \leq 2C^T B^{-1} C. \tag{164}$$

Applying (164) to (80), we get

$$\Sigma_i \leq B, \qquad \left(\text{because Condition 1)} \Rightarrow C^T B^{-1} C \geq 0\right), \tag{165}$$

and

$$\begin{aligned} \Sigma_i &\geq B - 2C^T B^{-1} C \\ &> B - B/2 \qquad \left(\text{applying Lemma 1}\right). \end{aligned}$$ □

Proof of Lemma 3: Prove by induction.

Basis:

$$\Sigma_2 - \Sigma_3 = \Sigma_2 - B + C^T \Sigma_2^{-1} C \tag{166}$$

which, by Condition 3b), is either nonnegative definite, or nonpositive definite.

Induction:

$$\Sigma_i - \Sigma_{i+1} = C^T\left(\Sigma_i^{-1} - \Sigma_{i-1}^{-1}\right) \cdot C \begin{cases} \geq 0, & \text{if } \Sigma_{i-1} - \Sigma_i \geq 0, \\ \leq 0, & \text{if } \Sigma_{i-1} - \Sigma_i \leq 0. \end{cases} \tag{167}$$

Depending on the induction hypothesis, $(\Sigma_i - \Sigma_{i+1})$ is either monotone nonincreasing or monotone nondecreasing. Conditions 1), 2), and 3a) are used implicitly because (167) assumes Σ_{i-1} and Σ_i are pd. □

Proof of Theorem 4: From lemmas 2 and 3, $\{\Sigma_i\}$ is a bounded monotone sequence. Hence, from [27, p. 263],

$$\Sigma_i \to \Sigma_\infty.$$

Taking the limit on both sides of equation (80), we get the Riccati equation (95). □

Proof of Lemma 4: There are two cases to consider.

1) $C = C^T > 0$.

Then

$$B + 2C > 0$$

follows immediately.

From the corollary to Theorem 1.3 in [31, p. 11],

$$\|B^{-1}\|_s = \frac{1}{\lambda_{\min}(B)}, \tag{168}$$

$$\|C\|_s = \lambda_{\max}(C), \tag{169}$$

where

$$\lambda_{\min}(B) = \min_i \lambda_i(B), \tag{170}$$

$$\lambda_{\max}(C) = \max_i \lambda_i(C), \tag{171}$$

with $\lambda_i(B)$ and $\lambda_i(C)$ being the eigenvalues of B and C, respectively. Substituting (168) and (169) into (88) and applying Condition 2), we get

$$\lambda_{\min}(B) > 2\lambda_{\max}(C). \tag{172}$$

Applying Theorem 4 from [21, p. 203], we get

$$2\frac{Y^T C Y}{Y^T Y} \leq 2\lambda_{\max}(C) < \lambda_{\min}(B) \leq \frac{Y^T B Y}{Y^T Y}, \quad \text{for any vector } Y, \tag{173}$$

which implies that

$$B - 2C > 0. \tag{174}$$

2) $C = C^T < 0$.

The proof for this case is identical to the first one, except for a change in sign ($-C$ is used everywhere in place of C). □

Proof of Corollary 2: This requires that we solve Riccati equation (95) given the additional condition in (96). From (98) and (96),

$$D = D^T > 0. \tag{175}$$

Substituting (98) into (95), the Riccati equation to be solved becomes

$$\Sigma_\infty = B - D\Sigma_\infty^{-1}D. \tag{176}$$

Premultiplying and postmultiplying (176) by $D^{-1/2}$, and substituting

$$\tilde{\Sigma}_\infty = D^{-1/2}\Sigma_\infty D^{-1/2}, \tag{177}$$

$$\tilde{B} = D^{-1/2}BD^{-1/2}, \tag{178}$$

we get a simpler Riccati equation,

$$\tilde{\Sigma}_\infty = \tilde{B} - \tilde{\Sigma}_\infty^{-1}. \tag{179}$$

Premultiplying (179) by $\tilde{\Sigma}_\infty$ and rearranging, we get

$$\tilde{\Sigma}_\infty\tilde{B} = \tilde{\Sigma}_\infty^2 + I, \tag{180}$$

while postmultiplying (179) by $\tilde{\Sigma}_\infty$ and rearranging gives us

$$\tilde{B}\Sigma_\infty = \tilde{\Sigma}_\infty^2 + I. \tag{181}$$

From (180) and (181), we have

$$\tilde{\Sigma}_\infty\tilde{B} = \tilde{B}\tilde{\Sigma}_\infty. \tag{182}$$

Then, from Theorem 7.2 [31, p. 220], it follows that $\tilde{\Sigma}_\infty$ and $\tilde{B}$ have the sane set of orthonormal eigenvectors. Consequently, the matrix

$$\left[(\tilde{B}/2)^2 - I\right]$$

has the same set of orthonormal eigenvectors as $\tilde{\Sigma}_\infty$.
From Lemma 4,

$$B > \pm 2C \Leftrightarrow B > \pm 2D \quad (\text{using } (98))$$

$$\Leftrightarrow \frac{\tilde{B}}{2} > \pm I \quad (\text{using } (178)). \tag{183}$$

The product of two symmetric pd matrices, $((\tilde{B}/2) + I)$ and $((\tilde{B}/2) - I)$, has all its eigenvalues positive (see, for example, [19, p. 218]). Furthermore, since the product

$$((\tilde{B}/2) + I)((\tilde{B}/2) - I) = (\tilde{B}/2)^2 - I \tag{184}$$

is symmetric, it follows that

$$(\tilde{B}/2)^2 - I > 0. \tag{185}$$

Then we can define its unique symmetric, pd, square root,

$$\left[(\tilde{B}/2)^2 - I\right]^{1/2},$$

which has the same set of orthonormal eigenvectors as

$$\left[(\tilde{B}/2)^2 - I\right],$$

and $\tilde{\Sigma}_\infty$. As a result, the aforementioned matrices commute. This allows us to express (180) (or (181)) in factorized form

$$\left(\tilde{\Sigma}_\infty - (\tilde{B}/2) - \left[(\tilde{B}/2)^2 - I\right]^{1/2}\right)$$
$$\cdot\left(\tilde{\Sigma}_\infty - (\tilde{B}/2) + \left[(\tilde{B}/2)^2 - I\right]^{1/2}\right) = 0. \tag{186}$$

From the factored form, the two pd solutions of the Riccati equation (179) are immediately available. Substituting (177) and (178) into these two solutions, we obtain the 2 pd solutions of (95)

$$\Sigma_\infty^{(1)} = (B/2) + D^{1/2}\left[\left(D^{-1/2}(B/2)D^{-1/2}\right)^2 - I\right]^{1/2}D^{1/2} > B/2, \tag{187}$$

$$\Sigma_\infty^{(2)} = (B/2) - D^{1/2}\left[\left(D^{-1/2}(B/2)D^{-1/2}\right)^2 - I\right]^{1/2}D^{1/2} < B/2, \tag{188}$$

From Lemma 2, it follows that

$$\forall i \geq 2, \qquad \Sigma_i > B/2.$$

Then

$$\Sigma_i \to \Sigma_\infty \Rightarrow \Sigma_\infty > B/2,$$

which implies that (187) is the only valid steady state solution for (80). □

Proof of Theorem 5: Substituting from (80) and (95),

$$\begin{aligned}\Sigma_i - \Sigma_\infty &= C^T\left(\Sigma_\infty^{-1} - \Sigma_{i-1}^{-1}\right)C,\\ &= C^T\Sigma_{i-1}^{-1}\left(\Sigma_{i-1} - \Sigma_\infty\right)\Sigma_\infty^{-1}C,\\ &= C^T\Sigma_{i-1}^{-1}C^T\left(\Sigma_\infty^{-1} - \Sigma_{i-2}^{-1}\right)C\Sigma_\infty^{-1}C,\\ &\vdots\\ &= C^T\Sigma_{i-1}^{-1}C^T\Sigma_{i-2}^{-1}\cdots C^T\\ &\quad\cdot\Sigma_2^{-1}\left(\Sigma_2 - \Sigma_\infty\right)\left(\Sigma_\infty^{-1}C\right)^{i-2}.\end{aligned} \tag{189}$$

Taking the spectral norm of both sides of (189) and applying properties of matrix norms (see Theorem 1.2 [31, p. 9]) and the fact that for any C

$$\|C^T\|_s = \|C\|_s,$$

we get

$$\|\Sigma_i - \Sigma_\infty\|_s \leq \|\Sigma_2 - \Sigma_\infty\|_s\|C\|_s^{2(i-2)}\|\Sigma_\infty^{-1}\|_s^{i-2} \cdot \prod_{j=2}^{i-1}\|\Sigma_j^{-1}\|_s. \tag{190}$$

From Lemma 2,

$$\forall j \geq 2, \qquad \Sigma_j > B/2, \tag{191}$$

$$\Rightarrow \Sigma_\infty > B/2. \tag{192}$$

From (191), we get

$$\forall j \geq 2, \qquad \Sigma_j^{-1} < 2B^{-1}, \tag{193}$$

and from (192),

$$\Sigma_\infty^{-1} < 2B^{-1}. \tag{194}$$

By applying Theorem 9 in [21, p. 208], it is straightforward to show that for any two symmetric, pd matrices R and T,

$$R > T \Rightarrow \|R\|_s > \|T\|_s.$$

Applying this result to (193) and (194), we get

$$\forall j \geq 2, \qquad \|\Sigma_j^{-1}\|_s < 2\|B^{-1}\|_s \tag{195}$$

and

$$\|\Sigma_\infty^{-1}\|_s < 2\|B^{-1}\|_s. \tag{196}$$

Substituting (195), (196), and (88) into (190) we get

$$\begin{aligned}\|\Sigma_i - \Sigma_\infty\|_s &\leq \|\Sigma_2 - \Sigma_\infty\|_s \zeta^{2(i-2)}, \\ &= K_0 a^i,\end{aligned} \tag{197}$$

where

$$K_0 = \frac{\|\Sigma_2 - \Sigma_\infty\|_s}{\zeta^4} > 0, \tag{198}$$

$$a = \zeta^2. \tag{199}$$

The inequality in (104) follows from (88), Condition 2), and (199). From Condition 3a),

$$\Sigma_2 \leq B \tag{200}$$

$$\Rightarrow \|\Sigma_2\|_s \leq \|B\|_s, \tag{201}$$

while from Lemma 1

$$\begin{aligned}\forall j \geq 2, \qquad \Sigma_j \leq B &\Rightarrow \Sigma_\infty \leq B \\ &\Rightarrow \|\Sigma_\infty\|_s \leq \|B\|_s.\end{aligned} \tag{202}$$

Using the properties of matrix norms (from Theorem 1.2 in [31, p. 9]) and (201) and (202) in (198), we get (105).

Proof of Result 1—Verification of Conditions 1)-3) for First-Order Fields:

Condition 1): This follows immediately for all three fields because the potential matrix is symmetric pd.

Condition 2): From Corollary 2 in [30], it follows that for a symmetric, diagonally dominant matrix, $B = [b_{i,j}]$,

$$\|B^{-1}\|_s \leq \frac{1}{\alpha}, \tag{203}$$

where

$$\alpha = \min_i \left(|b_{i,i}| - \sum_{j \neq i} |b_{i,j}| \right). \tag{204}$$

Here

$$\alpha = \begin{cases} 1 - 2|\beta_h|, & \text{for Dirichlet or variational fields,} \\ 1 - 3|\beta_h|, & \text{for symmetric fields.} \end{cases} \tag{205}$$

Using (203) and (205) in (114), we get

$$\begin{aligned}\zeta &\leq \begin{cases} \dfrac{2|\beta_v|}{(1 - 2|\beta_h|)}, & \text{for Dirichlet or variational fields,} \\ \dfrac{2|\beta_v|}{(1 - 3|\beta_h|)}, & \text{for symmetric fields,} \end{cases} \\ &< 1, \qquad \text{(using (121), and (122))},\end{aligned} \tag{206}$$

i.e., Condition 2) holds.

Condition 3a): For *Dirichlet* fields, from (79) and (115),

$$\Sigma_2 = B - C^T B^{-1} C. \tag{207}$$

Then,

$$\Sigma_2 < B \qquad (\text{because } B > 0 \Rightarrow B^{-1} > 0 \Rightarrow C^T B^{-1} C > 0).$$

Since Lemma 1 only requires Conditions 1) and 2), which have already been shown to hold for the fields we are considering in this subsection, we can apply it to (207), thus getting

$$\Sigma_2 > B/2.$$

For *variational* fields, from (79) and (118),

$$\Sigma_2 = B - C^T B_1^{-1} C. \tag{208}$$

From (116), (113), and (121), and the sufficient conditions given above for a symmetric matrix to be pd, we get

$$B \pm 2C > 0. \tag{209}$$

Then, from (117) and (209),

$$\begin{aligned}B_1 - B/2 > 0 &\Rightarrow B_1^{-1} < 2B^{-1} \\ &\Rightarrow C^T B_1^{-1} C < 2C^T B^{-1} C \\ &\Rightarrow \Sigma_2 > B - 2C^T B^{-1} C \qquad \text{(using (208))} \\ &\quad > B/2 \qquad \text{(using lemma 1)}.\end{aligned}$$

In addition,

$$\Sigma_2 < B \qquad (\text{because } B_1 > 0 \Rightarrow C^T B_1^{-1} C > 0). \tag{210}$$

For *symmetric* fields, from (79) and (120),

$$\Sigma_2 = B - 4C^T B^{-1} C. \tag{211}$$

Then,

$$\Sigma_2 < B. \tag{212}$$

Using (119) and (113),

$$B \pm 3C = (1 \mp 3\beta_v) I_M - \beta_h (H_M + F_{1,2} + F_{M-1,M}). \tag{213}$$

Applying (122) to (213), we get

$$B \pm 3C > 0, \tag{214}$$

because the matrix in (213) has positive diagonal elements, and is strictly diagonally dominant. Using (214), (113), and (123), and the properties of pd matrices (see Theorem 24 in [21, p. 22]), it is straightforward to show that

$$B > 9C^T B^{-1} C. \tag{215}$$

Then, from (211) and (215), it follows that

$$\Sigma_2 > \tfrac{5}{9} B, \tag{216}$$

i.e., Condition 3a) holds.

Condition 3b): For *Dirichlet* fields,

$$\begin{aligned}\Sigma_2 - B + C^T \Sigma_2^{-1} C &= C^T (\Sigma_2^{-1} - B^{-1}) C \qquad \text{(using (207))} \\ &> 0 \ (\text{because } \Sigma_2 < B \Rightarrow \Sigma_2^{-1} > B^{-1}).\end{aligned} \tag{217}$$

For *variational* fields,

$$\Sigma_2 - B + C^T \Sigma_2^{-1} C = C^T (\Sigma_2^{-1} - B_1^{-1}) C$$

$$\text{(using (208))}. \quad (218)$$

There are two possible cases.

Case 1) $\beta_v < 0$ (i.e., $C > 0$):

$$\begin{aligned} \Sigma_2 &< B \quad \text{(from (210))} \\ &< B + C = B_1 \quad \text{(from (117))} \\ &\Rightarrow \Sigma_2^{-1} > B_1^{-1} \quad (219) \\ &\Rightarrow \Sigma_2 - B + C^T \Sigma_2^{-1} C > 0 \\ &\cdot \text{(using (218) and (219))}. \quad (220) \end{aligned}$$

Case 2) $\beta_v > 0$ (i.e., $-C > 0$):

$$\begin{aligned} B_1 &> -2C + C = -C \quad \text{(using (117) and (209))} \\ &\Rightarrow B_1^{-1} < -C^{-1}. \\ &\Rightarrow -C^T B_1^{-1} C > C \\ &\Rightarrow \Sigma_2 > B_1 \quad \text{(adding } B \text{ to both sides and using} \\ &\text{(117) and (208))}. \\ &\Rightarrow \Sigma_2^{-1} - B_1^{-1} < 0. \quad (221) \end{aligned}$$

From (218) and (221),

$$\Sigma_2 - B + C^T \Sigma_2^{-1} C < 0. \quad (222)$$

For *symmetric* fields:

$$\begin{aligned} \Sigma_2 - B + C^T \Sigma_2^{-1} C &= C^T \left(\Sigma_2^{-1} - 4B^{-1}\right) C \ \text{(using (211))} \\ &< 0 \ \left(\text{because (216)} \Rightarrow \Sigma_2^{-1} < 4B^{-1}\right). \end{aligned}$$

□ (223)

Proof of Lemma 5: Since B is symmetric (by Condition 1)), it can be represented as

$$B = Q \Lambda_B Q^{-1}, \quad (224)$$

where

$$Q = Q^T = Q^{-1}$$

is the orthogonal matrix that contains the orthonormal eigenvectors of B, and

$$\Lambda_B = \text{diag}\left(\lambda_i(B)\right)$$

is the diagonal matrix of eigenvalues of B.

Using (224) in (124),

$$\begin{aligned} \Sigma_\infty &= Q\left[\Lambda_B/2 + \left((\Lambda_B/2)^2 - \beta_v^2 I_M\right)^{1/2}\right] Q^{-1} \\ &= Q \Lambda_\infty Q^{-1}, \quad (225) \end{aligned}$$

with

$$\Lambda_\infty = \text{diag}\left(\lambda_i(\infty)\right), \quad (226)$$

where $\{\lambda_i(\infty)\}$, the eigenvalues of Σ_∞, are given by

$$\lambda_i(\infty) = \frac{\lambda_i(B)}{2} + \sqrt{\left(\frac{\lambda_i(B)}{2}\right)^2 - \beta_v^2}. \quad (227)$$

Defining

$$\lambda_{\min}(\infty) = \min_{1 \le i \le M} \lambda_i(\infty), \quad (228)$$

we get

$$\|\Sigma_\infty^{-1}\|_s = \frac{1}{\lambda_{\min}(\infty)}. \quad (229)$$

Substituting (227) into (228), applying (170), and then substituting into (229), we get

$$\|\Sigma_\infty^{-1}\| = \frac{1}{\frac{\lambda_{\min}(B)}{2} + \sqrt{\left(\frac{\lambda_{\min}(B)}{2}\right)^2 - \beta_v^2}}. \quad (230)$$

Substituting (168) into (230), and using (114), (129) follows after some algebraic manipulation. Using (129) in place of (196), when substituting into (190) in the proof for Theorem 4, we obtain (130). □

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] M. B. Averintsev, "The description of Markov random fields by Gibbs distribution," *Teor. Verojatrost. i Primen.*, vol. 17, pp. 21-35, 1972.

[3] N. Balram and J. M. F. Moura, "Recursive enhancement of noncausal images," in *Proc. IEEE ICASSP'91*, May 1991, pp. 2997-3000.

[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 and P. A. P. Moran, "On the estimation and testing of spatial interaction in Gaussian lattice processes," *Biometrika*, vol. 62, pp. 555-562, 1975.

[6] D. Brook, "On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbor systems," *Biometrika*, vol. 51, pp. 481-483, 1964.

[7] R. S. Bucy, "*A priori* bounds for the Riccati equation," in *Proc. Sixth Berkeley Symp. Math. Statist. Probab.*, Univ. California Press, 1970, pp. 645-656.

[8] R. S. Bucy, "Interpolation," *Int. J. Contr.*, vol. 39, no. 4, pp. 767-772, 1984.

[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] D. C. Fraser and J. E. Potter, "The optimum linear smoother as a combination of two optimum linear filters," *IEEE Trans. Automat. Contr.*, vol. AC-14, pp. 387-390, Aug. 1969.

[11] 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.

[12] J. K. Goutsias, "Mutually compatible Gibbs random fields," *IEEE Trans. Inform. Theory*, vol. 35, pp. 538-544, Nov. 1989.

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

[14] A. Habibi, "Two-dimensional Bayesian estimate of images," in *Proc. IEEE*, vol. 60, pp. 878-883, 1972.

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

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

[17] R. L. Kashyap, "Finite lattice random field models for finite images," in *Conf. Inform. Sci. Syst.*, Baltimore, MD, Mar. 1981.

[18] B. Kaufman and Lars Onsager, "Crystal statistics, III, Short-range order in binary Ising lattice," *Phys. Rev.*, vol. 76, no. 4, pp. 1244-1252, Oct. 1949.

[19] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, second ed. Orlando, FL: Academic Press, 1985.

[20] B. C. Levy, R. Frezza, and A. J. Krener, "Modeling and estimation of discrete-time Gaussian reciprocal processes," *IEEE Trans. Automat. Cont.*, vol. 35, pp. 1013-1023, Sept. 1990.

[21] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with*

ill-posed problems in computational vision," *J. Am. Statist. Soc.*, vol. 82, pp. 76–89, 1987.

[23] D. Q. Mayne, "A solution of the smoothing problem for linear dynamic systems," *Automatica*, vol. 4, pp. 73–92, 1966.

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

[25] D. K. Pickard, "A curious binary lattice," *J. Appl. Probab.*, vol. 14, pp. 717–731, 1977.

[26] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *J. AIAA*, vol. 3, no. 8, pp. 1445–1450, Aug. 1965.

[27] F. Riesz and B. SZ. Nagy, *Functional Analysis*, second ed. New York: Frederick Ungar Publishing Co., 1955. (Translated from the 2nd French edition by L. F. Boron.)

[28] Frank Spitzer, "Markov random fields and Gibbs ensembles," *Am. Math. Monthly*, vol. 78, pp. 142–154, Feb. 1971.

[29] J. M. Varah, "On the solution of block-tridiagonal systems arising from certain finite-difference equations," *Math. Comput.*, vol. 26, no. 120, pp. 859–868, Oct. 1972.

[30] J. M. Varah, "A lower bound for the smallest singular value of a matrix," *Linear Algebra and its Applicat.*, vol. 11, pp. 3–5, 1975.

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

[32] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 232–240, Mar. 1972.

[33] J. W. Woods and C. H. Radewan, "Kalman filtering in two dimensions," *IEEE Trans. Inform. Theory*, vol. IT-23, pp. 473–481, July 1977.